TO: Consensus Standards Approval Committee (CSAC)

FR: Karen Johnson, Elisa Munthali, and Wunmi Isijola

RE: Vetting Requirement for "Endorsement Plus" designation

DA: March 16, 2016

NQF's Board of Directors recently approved a set of recommendations put forward by the Intended Use Advisory Panel. One of these recommendations includes creating a new designation that will identify measures that have exceeded NQF's endorsement criteria in several key areas. Measures achieving this designation (for now, "endorsement plus", although this label may change) will have the following characteristics:

- Meets evidence for measure focus without an exception
- Measure is reliable as demonstrated by reliability testing of the measure score
- Measure is valid as demonstrated by empirical validity testing of the measure score (i.e., not via face validity only)
- The candidate measure is *well-vetted in real world settings by those being measured and other users*.

The first three characteristics are fully defined in our endorsement process through our evaluation criteria. However, NQF must develop an operational definition of "real-world vetting" of measures and determine how to integrate this characteristic into the criteria for measure endorsement.

CSAC ACTION REQUIRED
The CSAC will review, discuss, and offer feedback on a proposed definition of real-world vetting of measures. The CSAC also will discuss and make a preliminary recommendation on how this new requirement should be incorporated into NQF's evaluation criteria for endorsement.

PROPOSED OPERATIONAL DEFINITION OF REAL-WORLD VETTING OF MEASURES
Based on the discussions of the Intended Use Advisory Panel, as well as input from prior CSAC discussion, NQF staff has developed a <u>draft</u> four-part operational definition of real-world vetting of measures, as follows.

Vetting of the measure by those being measured and others is demonstrated when:
1) those being measured have been given performance results and data, as well as assistance with interpreting the measure results and data
2) those being measured and other users have been given an opportunity to provide feedback on the measure performance and implementation
3) this feedback has been incorporated into the measure

4) there is general agreement that the results of the measure, as constructed, can be used to distinguish good from poor quality

This proposed definition reflects a "continuum" of effort that would be required of measure developers. Staff particularly acknowledges the aspirational nature part #4 of the proposed definition. As mentioned in earlier discussions, we do not expect most developers to be able to meet this requirement for their measures at this time.

OPTIONS FOR INCLUSION INTO NQF MEASURE EVALUATION CRITERIA
How the vetting requirement is incorporated into NQF's measure evaluation criteria depends to a large extent on how vetting is defined. Accordingly, staff offers two options for CSAC consideration regarding potential changes to the evaluation criteria:
- Option #1: Add an additional subcriterion under Validity
- Option #2: Add an additional subcriterion under Usability and Use

Option #1: Add an additional subcriterion under Validity
The validity of a measure is impugned when those being measured (or others) do not agree with how a measure is constructed or agree that the results of the measure accurately reflect the quality of care that has been provided. Thus, if vetting is defined as proposed above, it is reasonable to incorporate the vetting of a measure by those being measured and by other users into the Validity subcriterion. This would equate, essentially, to a requirement for the systematic assessment of the face validity of the measure by those being measured and by other users. Staff propose including the vetting requirement as another assessment of potential threats to validity (as with 2b3-2b7, below):

2b. Validity (must-pass)
    2b1. Specifications consistent with evidence
    2b2. Validity testing
    2b3. Justification of exclusions—relates to evidence
    2b4. Risk adjustment—typically for outcome/cost/resource use
    2b5. Identification of differences in performance
    2b6. Comparability of data sources/methods
    2b7. Missing data
    2b8. Vetting of the measure by those being measured and others

In the current measure development environment, the complete vetting of a measure, as described in the proposed definition above, typically is not part of the standard development process. Thus, its inclusion under the Validity subcriterion—which is a must-pass subcriterion— would be a strong signal to the field as to its importance.

Initially, however, we would include this requirement as an ***optional*** part of the Validity subcriterion, to be used solely for evaluating whether a measure should receive the "endorsement plus" designation. At least initially, failure to address the vetting requirement

would **NOT** be grounds for refusing to endorse a measure. If desired, requiring vetting to be a must-pass component of the Validity subcriterion could be considered in the future.

Consistent with current NQF policy, threshold values regarding the number of users included or the level of agreement achieved would NOT be specified, nor would the methods for generating and incorporating feedback. Instead, Standing Committees would consider the scope, method, and results and decide if they believe the vetting requirement has been sufficiently addressed.

Option #2: Add an additional subcriterion under Usability and Use

Usability and Use currently is defined as the "*extent to which potential audiences 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*." Inclusion of the vetting requirement under the Usability and Use criterion is a reasonable option particularly if there is a perception that measure implementers are disinclined to use measures when those being measured have not had a chance to offer feedback on the measure. Even though the Usability and Use criterion is not must-pass, inclusion of the vetting requirement under this criterion would be a compelling signal to the field about its importance.

Staff propose including the vetting requirement as a fourth subcriterion under Usability and Use, as follows:

4. Usability and Use (not a must-pass criterion)
    4a. Accountability and Transparency
    4b. Improvement
    4c. Benefits outweigh unintended negative consequences
    4d. Vetting of the measure by those being measured and others

Note, however, that a *less expansive operational definition* than what is proposed above would be more appropriate if included as part of the Usability and Use criterion. Specifically, staff would propose wording the subcriterion as follows:

   4d. Vetting of the measure by those being measured and others is demonstrated when:
        1) those being measured have been given performance results and data, as well as assistance with interpreting the measure results and data
        2) those being measured and other users have been given an opportunity to provide feedback on the measure performance and implementation
        3) this feedback has been incorporated into the measure

As noted with the Validity subcriterion above, Standing Committees would consider the scope, method, and results and decide if they believe the vetting requirement has been sufficiently addressed.

Defining and Implementing the Vetting Requirement for the "Endorsement Plus" designation

NATIONAL QUALITY FORUM

CSAC Meeting
March 24, 2016

Criteria for achieving the "Endorsement Plus" designation

- Meets evidence for measure focus without an exception

- Measure is reliable as demonstrated by reliability testing of the measure score

- Measure is valid as demonstrated by empirical validity testing of the measure score (i.e., not via face validity only)

- The candidate measure is *well-vetted in real world settings by those being measured and other users*

Proposed operational definition of real-world vetting of measures

Vetting of the measure by those being measured and others is demonstrated when:

1. those being measured have been given performance results and data, as well as assistance with interpreting the measure results and data

2. those being measured and other users have been given an opportunity to provide feedback on the measure performance and implementation

3. this feedback has been incorporated into the measure

4. there is general agreement that the results of the measure, as constructed, can be used to distinguish good from poor quality

Two options for inclusion into NQF measure evaluation criteria

- Option #1: Add an additional subcriterion under Validity

- Option #2: Add an additional subcriterion under Usability and Use

Option #1: Add an additional subcriterion under Validity

2b. Validity (must-pass)

    2b1. Specifications consistent with evidence

    2b2. Validity testing

    2b3. Justification of exclusions—relates to evidence

    2b4. Risk adjustment—typically for outcome/cost/resource use

    2b5. Identification of differences in performance

    2b6. Comparability of data sources/methods

    2b7. Missing data

    2b8. Vetting of the measure by those being measured and others

If included under Validity…

- It would be **optional** part of the Validity subcriterion,
    - Used solely for evaluating whether a measure should receive the "endorsement plus" designation
    - Failure to address the vetting requirement would **NOT** be grounds for refusing to endorse a measure.
    - Making it a must-pass component could be considered in the future

Option #2: Add an additional subcriterion under Usability and Use

4. Usability and Use (not a must-pass criterion)

    4a. Accountability and Transparency

    4b. Improvement

    4c. Benefits outweigh unintended negative consequences

    4d. Vetting of the measure by those being measured and others

- A *less expansive operational definition* would be more appropriate (e.g., dropping the 4th component)