NATIONAL QUALITY FORUM

06/14/11

GUIDANCE ON COMPETING MEASURES AND SELECTION OF THE BEST MEASURE

NQF is increasingly faced with the submission of multiple measures with the same measure focus and same target population. The NQF Board recently reiterated the policy to endorse the best measure (often referred to as best-in-class) and asked the Consensus Standards Approval Committee (CSAC) to draft guidance to assist steering committees in applying NQF's policy and criteria to identify the best measure for endorsement from among competing measures. This guidance document addresses the evaluation of competing measures and should be useful both to project steering committees and measure developers. Guidance on evaluating related measures for harmonization was the subject of a prior project and approved by the NQF Board in 2010.

Definition of Competing Measures

Competing measures are those that essentially address the same concepts for the target process, condition, event or outcome and the same target patient population. Competing measures are the same at the conceptual level, but differ in technical specifications. The goal is to endorse the best measure and minimize confusing or conflicting information.

Table 1. Related versus Competing Measures

	Same concepts for measure focus—target process, condition, event, outcome	**Different concepts for measure focus—target process, condition, event, outcome**
Same target patient population	**Competing measures—Select best measure** from competing measures or justify endorsement of additional measure(s).	**Related measures—Harmonize** on target patient population or justify differences.
Different target patient population	**Related measures—Combine** into one measure with expanded target patient population or justify why different harmonized measures are needed.	Neither harmonization nor competing measure issue

Although not the subject of this guidance, it is helpful to distinguish competing measures from related measures, which are the primary focus of measure harmonization and addressed in a prior report. Related measures fall into one of two categories: 1) those that address the same concepts for measure focus but different patient populations; and 2) those that address different concepts for measure focus for the same patient population. For the first category, the developers should be encouraged to combine the two measures into a single measure with an expanded target patient population. For the second category, two measures may be appropriate but efforts should be made to harmonize definitions of the target patient population.

Principles for Selection of the Best from among Competing Measures

1. The endorsement of multiple competing measures should be by exception with adequate justification.
2. NQF prefers endorsement of measures that include the broadest possible target patient population for whom the measure is appropriate as indicated by the evidence.
3. NQF prefers endorsement of measures that assess performance for the broadest possible application (e.g., for as many possible individuals, entities, settings, and levels of analysis) for which the measure is appropriate.
4. If a single measure cannot accommodate the inclusion of all relevant patient populations or entities for performance measurement, a second measure could be considered for endorsement. The two measures should be harmonized to the extent possible.
5. When best in class is not clear, it may be appropriate to endorse more than one competing measure. At the time of initial endorsement, NQF should identify analyses needed to conduct a rigorous evaluation of the use and usefulness of the measures. This information should be provided by the developers to support a "best in class" determination at the time of 3-year maintenance.
6. Until clinical data from electronic health records (EHRs) are widely available for performance measurement and reporting, endorsement of competing measures based on different data types (e.g., claims and EHRs) may be justified. Two measures may be needed to achieve the dual goals of 1) advocating widespread access to performance results and 2) migrating to performance measures based on clinical data from EHRs.

Guidance for Evaluating Competing Measures

All measures must first be evaluated individually and judged to adequately meet all four evaluation criteria to be suitable for a steering committee to recommend endorsement before comparing to competing measures. This is intended to give each measure a thorough evaluation and also to prevent expending time and effort on comparing measures if some competing measures are not evaluated favorably.

If a new measure competes with an NQF-endorsed measure, the developer should be expected to address how the proposed measure is superior to competing measures, or the added value of endorsing multiple measures. Ideally, the developer will be able to present analyses demonstrating how the submitted measure is superior; however in some situations that will not be feasible (e.g., no access to an alternative data source) and then they should be able to present a rationale for superiority that is based on the NQF evaluation criteria. If the competing measure also is a new submission, the developers will be asked to address that question after the committee determines that both meet the evaluation criteria.

The quality enterprise is beginning a transition to performance measures derived from clinical data in EHRs. Oftentimes the same (or similar) concepts can be measured using different data types, specifically data from claims and clinical records (e.g., EHRs, paper charts, or data submitted to registries). Given the greater specificity of data derived from clinical records, particularly EHRs, performance measures based on such data may be more reliable and valid indicators of quality than similarly focused measures using only claims data. Nonetheless, claims data are often more available for measurement; and historically, measures based on claims are more likely to be publicly reported. Until such time as EHR-based measures are the primary source of publicly available performance measures of quality, it may be reasonable to allow endorsement of measures of similar concepts derived from both clinical and claims-based data sources.

The algorithm developed for harmonization provided a useful starting point for depicting the steps in identifying and evaluating competing measures (Figure 1). The first part of the algorithm applies to both competing and related measures. The left side applies to competing measures.

Figure 1. Addressing Competing Measures in the NQF Evaluation Process

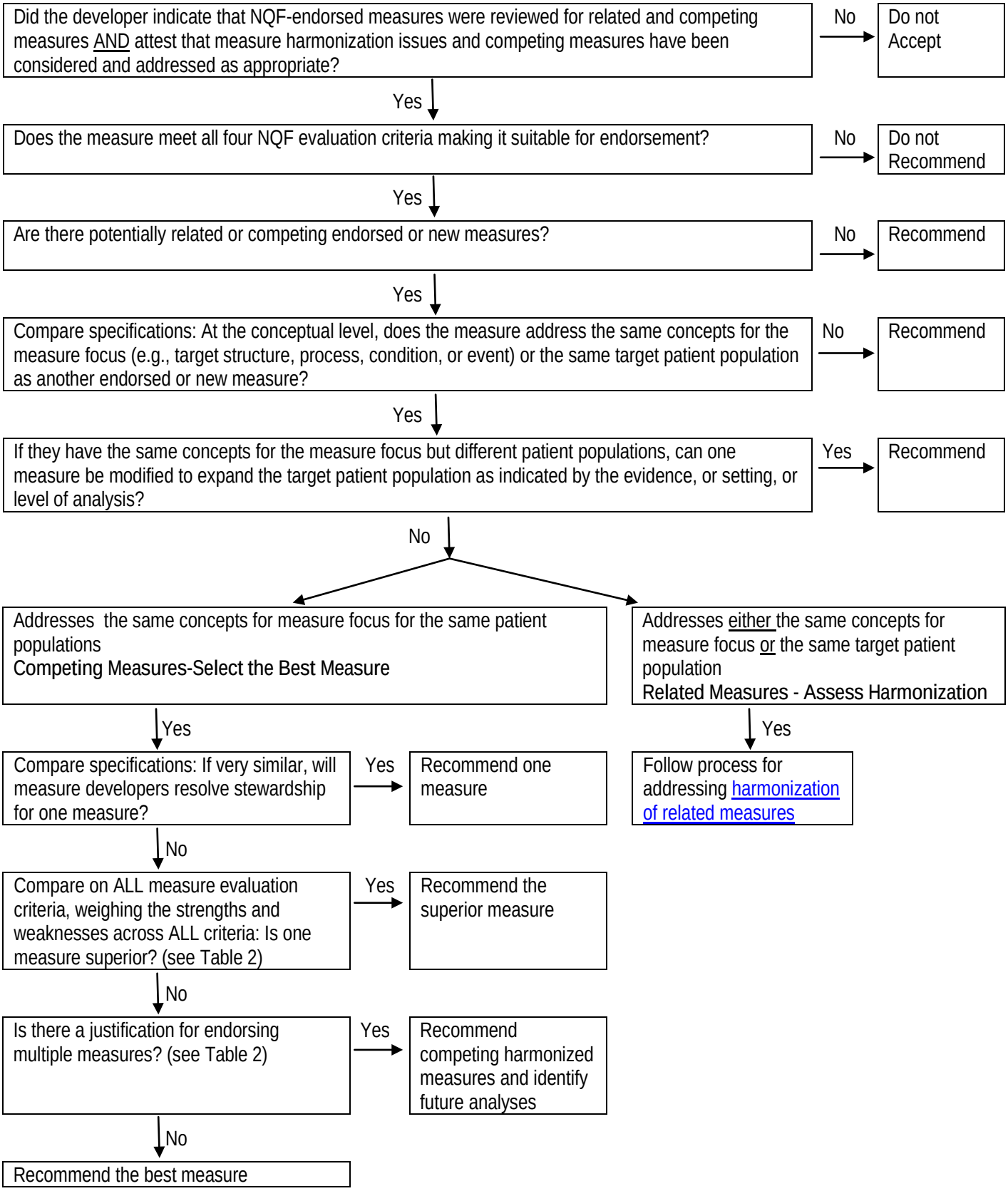

Table 2 provides an approach to the evaluation of competing measures for superiority or justification for multiple measures.

Table 2. Evaluating Competing Measures for Superiority or Justification for Multiple Measures

Steps	Evaluate Competing Measures
1. Determine if need to compare measures for superiority	Work through the steps in the algorithm (Figure 1) to determine if need to evaluate competing measures for superiority (i.e., two or more measures address the same concepts for measure focus for the same patient populations)
2.Assess Competing Measures for Superiority by weighing the strengths and weaknesses across ALL NQF evaluation criteria	Because the competing measures have already been determined to have met NQF's criteria for endorsement, the assessment of competing measures must include weighing the strengths and weaknesses across ALL the criteria and involves more than just comparing ratings. (For example, a decision is not based on just the differences in scientific acceptability of measure properties without weighing the evaluation of importance to measure and report, usability, and feasibility as well.) Impact, Opportunity, and Evidence—Importance to Measure and Report: Competing measures generally will be the same in terms of the measure focus addressing a high-impact aspect of healthcare (1a) and evidence for the focus of measurement (1c). However, due to differences in measure construction, they could differ on alignment with national health goals/priorities or opportunity for improvement. • Compare measures on alignment with national health goals/priorities (1a) • Compare measures on opportunity for improvement (1b) Reliability and Validity—Scientific Acceptability of Measure Properties: • Compare evidence of reliability (2a1-2a2) • Compare evidence of validity, including threats to validity (2b1-2b6) Untested measures cannot be considered superior to tested measures because there would be no empirical evidence on which to compare reliability and validity. (However, a new measure, when tested, could ultimately demonstrate superiority over an endorsed measure and the NQF endorsement maintenance cycles allow for regular submission of new measures.) Compare and identify differences in specifications All else being equal on the criteria and subcriteria, the preference is for: • Measures specified for the broadest application (target patient population as indicated by the evidence, settings, level of analysis) • Measures that address disparities in care when appropriate Usability: • Compare evidence of use and usefulness for public reporting, including availability of data for reporting performance results • Compare evidence of use and usefulness for quality improvement All else being equal on the criteria and subcriteria, the preference is for: • Measures that are publicly reported • Measures with the widest use (e.g., settings, numbers of entities reporting performance

Steps	Evaluate Competing Measures
	results) • Measures that are in use over those without evidence of use Feasibility: • Compare the ease of data collection/availability of required data • Compare the potential for inaccuracies, errors, and unintended consequences <u>All else being equal on the criteria and subcriteria</u>, the preference is for: • Measures based on data from electronic sources • Clinical data from EHRs • Measures that are freely available After weighing the strengths and weaknesses across ALL criteria, identify if one measure is clearly superior and provide the rationale based on the NQF criteria.
3.If a competing measure does not have clear superiority, assess justification for multiple measures	If a competing measure does not have clear superiority, is there a justification for endorsing multiple measures? Does the added value offset any burden or negative impact? Identify the value of endorsing competing measures Is an additional measure necessary? • to change to EHR-based measurement; • to have broader applicability (if one measure cannot accommodate all patient populations; settings, e.g., hospital, home health; or levels of analysis, e.g., clinician, facility; etc.); • to increase availability of performance results (if one measure cannot be widely implemented, e.g., if measures based on different data types increase the number of entities for whom performance results are available) <u>Note</u>: Until clinical data from electronic health records (EHRs) are widely available for performance measurement, endorsement of competing measures based on different data types (e.g., claims and EHRs) may be needed to achieve the dual goals of 1) advocating widespread access to performance data and 2) migrating to performance measures based on EHRs. EHRs are the preferred source for clinical record data, but measures based on paper charts or data submitted to registries may be needed in the transition to EHR-based measures. Is an additional measure unnecessary? • primarily for unique developer preferences Identify the burden of endorsing competing measures Do the different measures affect interpretability across measures? Does having more than one endorsed measure increase the burden of data collection? Determine if the added value of endorsing competing measures offsets any burden or negative impact? • If yes, recommend competing measures for endorsement (if harmonized) and provide the rationale for recommending endorsement of multiple competing measures. Also, identify analyses needed to conduct a rigorous evaluation of the use and usefulness of the measures at the time of endorsement maintenance. • If no, recommend the best measure for endorsement and provide rationale.

NQF Measure Evaluation Criteria

The NQF measure evaluation criteria were recently modified and selection of the best measure from among competing measures is addressed after the other criteria. Each measure is first evaluated individually and must be determined to be suitable for endorsement before it is compared to competing measures.

Determination of the best measure should be based on the evaluation criteria of *Importance to Measure and Report, Scientific Acceptability of Measure Properties, Usability,* and *Feasibility*. In the absence of empirical data to compare the measures, the Steering Committee will need to compare not only its evaluation ratings but also the information submitted in support of the criteria. The comparison will require expert judgment and requires weighing the strengths and weaknesses across all the criteria. For example, slightly lower reliability, but much greater feasibility might indicate the more feasible measure should be selected.

If the measures are determined to be conceptually the same, then generally they would be expected to be evaluated equally on the subcriteria under *Importance to Measure and Report,* i.e., impact, opportunity for improvement, and evidence supporting the focus of measurement. However, they could differ on opportunity for improvement depending on whether they are new measures or have been in use. For new measures, opportunity for improvement generally will be the same because it is based on epidemiologic and research data. However, measures in use and at the time of endorsement maintenance may differ in opportunity for improvement (e.g., one may be "topped out" in terms of performance). When measures are essentially the same on the criterion *Importance to Measure and Report,* the determination of the best measure to recommend for endorsement would be made based on the remaining criteria.

If the Steering Committee is unable to identify the best (superior) measure, multiple endorsed measures may be acceptable and the Steering Committees needs to identify the additive value of endorsement of more than one measure (see Table 2). That is, does having multiple measures add enough value to offset any potential negative impact? The Steering Committee will need to provide a rationale for recommending multiple competing measures and also identify analyses

for evaluation and identification of the best measure can be made at the time of endorsement maintenance.