NATIONAL QUALITY FORUM

TO: NQF Members

FR: Consensus Standards Approval Committee

SU: Consideration of revisions to Consensus Development Process

DA: April 6, 2011

As a part of the National Quality Forum's (NQF's) ongoing commitment to enhance the timeliness of the Consensus Development Process (CDP) and to ensure that endorsed measures and practices continue to meet Members' needs while remaining important, scientifically acceptable, useable, and feasible, we are proposing the following revisions and additions:

- Enhancement of the nine-step CDP
- Addition of an "inactive" endorsement status for measures that have achieved the highest level of performance yet still meet the remaining measure evaluation criteria
- eMeasure review process and timeline
- Guidance on competing measures and selection of the best measure

These changes were discussed by the Consensus Standards Approval Committee (CSAC) at the March 2011 meeting. Before presenting these recommendations to the NQF Board of Directors, we are seeking input from NQF Members and the public during this comment period.

Details for each of the proposed changes are included within this document, and comments can be submitted via online submission process. Comments **must be submitted using the online submission process by 6:00 pm ET on April 25, 2011.**

If you have any questions, please contact Helen Burstin, MD, at (202) 783-1300 or via e-mail at hburstin@qualityforum.org**.**

PROPOSED ENHANCEMENTS TO THE NQF CONSENSUS DEVELOPMENT PROCESS

EXECUTIVE SUMMARY

Healthcare legislation passed in the past two years (the Health Information Technology for Economic and Clinical Health Act and the Affordable Care Act) established a demanding schedule for the implementation of public reporting programs, value-based payment programs, and incentives for "meaningful use" of health information technology. All of these programs require standardized performance measures. The U.S. Department of Health and Human Services (HHS) has asked the National Quality Forum (NQF) to identify ways to enhance the timeliness of its measure endorsement process.

NQF currently follows a nine-step Consensus Development Process (CDP) when considering measures for endorsement. CDP project durations range from 12 to 15 months depending upon the volume and complexity of the measures and the ease of obtaining multi-stakeholder support. The Consensus Standards Approval Committee (CSAC) proposes three changes to the CDP that would collectively reduce the average project timeline by 3 to 4 months without compromising the CDP's integrity:

- Replace project-specific Steering Committees with term-limited topic-specific committees;
- Solicit measures earlier based on tentative project schedule; and
- Shorten the voting period to 15 days.

Member and public comments will be summarized and shared with the NQF Board of Directors before it makes final decisions regarding these enhancements.

BACKGROUND

The CDP has been utilized to endorse more than 600 measures over the past decade. In 2007, significant modifications were made to the CDP. Most notably, the CSAC was established and

charged with reviewing and recommending measures for endorsement to the Board of Directors. Since that time, the basic structure of the nine-step process has not changed significantly, but there have been very important improvements (summarized in a CDP year-in-review document), often sparked by Member and public input.

These enhancements have certainly strengthened the CDP, but some also have lengthened the time required to complete the process. In particular, the vetting of potential committee members for conflicts of interest and the posting of proposed Committee slates have added several weeks to the process.

In order to streamline the process, NQF worked with experts in Six Sigma Lean Processing and conducted a Value Stream Mapping process to identify and eliminate waste, cut lead times, and improve the CDP's quality. NQF also incorporated suggestions from the external evaluation of the CDP by Mathematica Policy Research. Through these efforts, NQF has identified additional opportunities to enhance both the CDP's integrity and timeliness. NQF staff has started to implement some of the changes that do not fundamentally change the CDP (e.g., earlier lead time for measure submission).

PROPOSED CHANGES

The current nine-step CDP includes:

1. **Call for Intent to Submit Measures.** Interested measure stewards are invited to notify NQF of their intent to submit measures for endorsement.
2. **Call for Nominations.** Nominations are open for 30 days for the multi-stakeholder committee that will oversee the project. After selection, NQF posts committee rosters on its website to solicit public comments on the composition of the panel and makes adjustments as needed to ensure balanced representation.
3. **Call for Measures.** Measures may be submitted during an open 30-day period through NQF's online submission form.
4. **Steering Committee Review.** The Steering Committee conducts a detailed evaluation of all submitted measures against the NQF evaluation criteria in open sessions.

5. **Public and Member Comment.** NQF solicits input to a draft report that outlines the steering committee's assessment of the measures for possible endorsement. The steering committee may request a revision to the proposed measures.
6. **Member Vote.** NQF asks members to review the draft report and cast their votes on the endorsement of measures.
7. **CSAC Review.** The CSAC deliberates on the merits of the measure and the issues raised during the review process, and makes a recommendation on endorsement to the Board of Directors.
8. **Board Ratification.** The Board provides final review and ratification of the measures for endorsement.
9. **Appeals.** During a 30-day period, anyone can appeal the Board's decision.

After reviewing the current process, three changes are being proposed:

- **Committee Nomination and Appointment Process.** Replace project-specific Steering Committees with term-limited topic-specific Committees.
- **Measure Solicitation.** Solicit measures earlier based on a tentative annual project schedule.
- **Member Voting**. Reduce the voting period to 14 days.

Committee Nomination and Appointment Process

Prior to the HHS contract that started in 2009, NQF operated with a great deal of uncertainty regarding resources for proposed projects. Consequently, work was organized on a project-by-project basis with no comprehensive schedule. NQF appointed project-specific Steering Committees, with the nominations process commencing when project funding had been secured.

During the past year, NQF established a three-year schedule for Endorsement Maintenance projects across 22 cross-cutting and condition-specific areas. The combination of longer-term planning and dedicated resources now provides the opportunity to move some of the more time-intensive steps of the CDP to pre-work. Specifically, NQF proposes to establish term-limited Standing Committees corresponding to the 22 identified areas. Committee members will serve

two-year terms, and the Committees will be responsible for handling endorsement and measure maintenance, as well as ad hoc and expedited project work in their designated areas.

Periodically, NQF will issue a simultaneous Call for Nominations across the CDP endorsement and maintenance areas planned for the coming months (e.g., in the next six months). Nominators will be able to see the projects, including descriptions of the project scope, and identify appropriate nominees across the range of projects. The Call for Nominations for the full range of 2011 endorsement and maintenance areas will be open for 45 days, rather than the current 30 days for individual projects.

Measure Solicitation

The projected project schedule over a three-year period also will provide measure developers with clear timelines for measure submission. NQF proposes to post measure submission deadlines for all projects for the upcoming year. This schedule generally will provide *more* time for measure developers to submit, with a minimum submission period of 30 days. This capacity for early submission also will eliminate the need for a Notice of Intent to Submit Measures. NQF is also developing the ability for measure developers to initiate a measure submission form without a formal Call for Measures. This will allow developers significant lead time for submission as well as an opportunity for NQF to view the measure pipeline prior to formal submission.

Public Comment

The comment period has a high level of Member and public participation. No changes are being proposed to the current 30-day comment period.

Member Voting

Given the significant opportunities for Member engagement throughout the process, the voting process mainly offers an opportunity to assess the extent of support across stakeholder groups, which is important input into CSAC and Board deliberations. To further inform Members on the status of measures, enhance Member engagement, and achieve broad consensus, NQF will

sponsor project-specific webinars with the Steering Committee co-chairs and a CSAC representative to review comments and final adjudication by the Committee. Webinars will provide Members with an opportunity to engage on the measures prior to the voting period and the CSAC and Board endorsement decisions. Shortly after the webinar, the voting period will commence. NQF recommends that the Member voting period be reduced from 30 to 15 days.

PROPOSED ESTABLISHMENT OF INACTIVE ENDORSEMENT STATUS

Given the number of publicly reported measures with high levels of performance, NQF is facing a situation where reliable and valid measures of great importance may not retain endorsement due to the lack of a performance gap. The purpose of an inactive endorsement status is to retain endorsement of reliable and valid quality performance measures that have overall high levels of performance with little variability so that performance could be monitored in the future if necessary to ensure that performance does not decline. This status would apply only to highly credible, reliable, and valid measures that have high levels of performance due to quality improvement actions (often facilitated or motivated through public reporting and other accountability programs). The key issue for continued endorsement is the opportunity cost associated with measuring things that are at high levels of performance – rather than focusing on areas where there is really a gap in care. Establishing an "inactive endorsement" status is one way to retain these measures in the NQF Portfolio to be used periodically for monitoring, while also communicating to potential users that the measures no longer address high leverage areas for accountability purposes. As discussed below, inactive status would only be assigned to measures with consistently high levels of performance that satisfy specified criteria.

Measures with High Levels of Performance - Recommendations from the Evidence Task Force

The Evidence Task Force addressed measures with high levels of performance with little variability as follows. The report and recommendations that follow were approved by the Board in 2010 and are being implemented this year.

When a measure undergoes review for continued endorsement, an issue that sometimes arises is whether the measure is "topped out," meaning there are high levels of performance with little variation and, therefore, little room for further improvement.

The Task Force did not recommend specific quantitative thresholds for identifying conformance with the subcriteria of high impact (1a) and opportunity for improvement (1b). Threshold values for opportunity for improvement would be difficult to standardize and depends on the size of the

population at risk, the effectiveness of an intervention, and the consequences of the quality problem. For example, even modest variation would be sufficient justification for some highly effective, potentially life-saving treatments (e.g., certain vaccinations) that are critical to the public health.

The Task Force noted that, at the time of endorsement maintenance review, if measure performance data indicate overall high performance with little variation, then justification would be required for continued endorsement of the measure. The CSAC added that the default action should be to remove endorsement unless there is a strong justification to continue endorsement. If a measure fails opportunity for improvement (1b), then it does not pass the threshold criterion, *Importance to Measure and Report,* and is therefore not suitable for endorsement.

Task Force Recommendations related to opportunity for improvement (1b) include the following:

- At the time of initial endorsement, evidence for opportunity for improvement generally will be based on research studies, or on epidemiologic or resource use data. However, at the time of review for endorsement maintenance, the primary interest is on the endorsed measure as specified, and the *evidence for opportunity for improvement should be based on data for the specific endorsed measure.*
- When assessing measure performance data for opportunity for improvement, the following factors should be considered:
    - number and representativeness of the entities included in the measure performance data;
    - data on disparities; and
    - size of the population at risk, effectiveness of an intervention, likely occurrence of an outcome, and consequences of the quality problem.
- At the time of review for endorsement maintenance, an overall high level of performance with little variation in the endorsed measure scores should result in removal of endorsement. If other evidence (e.g., epidemiologic or research) is consistent with the measure performance data, then it confirms the lack of opportunity for improvement. If

other evidence is not consistent with the measure performance data, then it is suggestive of potential problems with the measure as specified.

- In exceptional situations, a strong justification for continuing endorsement could be considered (e.g., ***evidence*** that overall performance will likely deteriorate if not monitored and of the magnitude of potential harm if outcomes deteriorate while not being monitored).

Criteria for Assigning Inactive Endorsement Status to Measures with High Levels of Performance

Rarely is there evidence that performance will deteriorate if a measure is not monitored; therefore, some additional criteria are needed. The CSAC identified the following criteria to be used when there are concerns that performance will deteriorate, but no evidence. These criteria are intentionally rigorous so that the use of inactive endorsement status is by exception.

- Evidence for measure focus – moderate to high ratings for quantity, quality, and consistency as described in the Evidence Task Force report (Table 4). There should be strong direct evidence of a link to a desired health outcome; therefore, there would be detrimental consequence on patient health outcomes if performance eroded.
- Generally measures more distal to the desired outcome with only indirect evidence of influence on the outcome would not qualify for inactive endorsement status.
  For example:
  - A measure focus is about assessing blood pressure (BP), but the direct evidence is for the link between BP level or a specific treatment to morbidity and mortality.
  - A measure is about assessing HbA1C, but the direct evidence is for HbA1C level or specific treatment leading to morbidity and mortality.
- Generally measures more distal to the desired outcome when another more proximal measure is available, would not be eligible for inactive endorsement status.
  For example:
  - A VTE measure focused on whether VTE prophylaxis was ordered when there is another measure focused on whether VTE prophylaxis was administered.

- Reliability – high rating as described in the Measure Testing Task Force report (Table2). Reliability has been demonstrated for both the data elements and measure scores.
- Validity – high rating as described in the Measure Testing Task Force report (Table2). Validity has been demonstrated for both data elements and the measure score (face validity not acceptable).
- Demonstrated usefulness for improving quality (e.g., data on trends of improvement and scope of patients and providers included)
- Demonstrated use of the measure (e.g., specific programs and scope of patients and providers included; would not grant inactive endorsement status for a measure that has not been used)
- The reason for high levels of performance is better performance, not an issue with measure construction/specifications (e.g., "documentation")

eMEASURES: PROPOSED TIMETABLE AND PROCESS

The U.S. healthcare system is making major investments in health information technology (health IT). Over the coming five years, many if not most healthcare providers will acquire electronic health records (EHRs), and many consumers will likely begin using personal health records (PHRs). Efforts are under way to enable interoperability, including the Office of the National Coordinator's NHIN Direct and the development of health information exchanges in Beacon and other communities. This migration to health IT, although currently very uneven, has already started to open up important opportunities to measure and improve care longitudinally, across the entire patient-focused episode, and to capture patient-reported outcomes (e.g., health functioning, health behaviors).

During the past 18 months, NQF and many measure stewards have been involved in efforts to rapidly "retool" existing measures for use on an electronic platform. As part of the HHS contract, NQF has worked with measure stewards to retool an initial set of more than 100 performance measures, many to be used for Health Information Technology for Economic and Clinical Health (HITECH) Act incentive payments linked to "meaningful use" of EHRs. In a recent child health quality measures project, NQF received, for the first time, measures submitted with specifications for EHRs.

NQF has been laying the groundwork for eMeasures endorsement for some time. The NQF Testing Task Force report released in September 2010 specifies requirements for testing new and retooled e-measures. The Quality Data Model (QDM; formerly Quality Data Set) specifies the types of data that need to be captured in EHRs to support quality measurement and is an essential building block for both performance measures and EHRs. NQF, with support from HHS, subcontracted with the Iowa Foundation for Medical Care to develop a Measure Authoring Tool that will help developers generate specifications for eMeasures in a consistent fashion. The tool is expected to be publicly available to measure developers in 2012.

Because the pace of migration to EHRs and PHRs will vary across geographic areas and

providers, the portfolio of NQF-endorsed® measures likely will need to include performance measures for use with different data platforms (e.g., paper records and EHRs) for at least the coming five years. In some instances, the portfolio may include a measure with two sets of specifications—one for paper records and another for EHRs (e.g., primary PCI within 90 minutes). In other instances, the portfolio may include two measures that address the same topic area but are substantively different because the availability of EHRs presents opportunities to measure in a better way. In neither instance should it be inferred that performance results derived from different data platforms are comparable. Users of the measures would need to conduct additional analyses to determine if such is the case.

Migration Plan

Currently about one-fifth of the measures in the NQF portfolio include specification for EHRs. Working with measure stewards, HHS, and others, NQF has developed a plan for further migration of the portfolio.

Timing

Starting March 2012, NQF will require that all newly submitted measures, and currently endorsed measures going through maintenance review, include specifications for EHRs or PHRs (if appropriate). Before that time, NQF will accept eMeasures for consideration in its endorsement projects, but EHR or PHR specifications will not be a requirement.

Format

As noted above, the Measure Authoring Tool should be ready for widespread use in January 2012. Starting March 2012, measure stewards will be required to use the Measure Authoring Tool when submitting measures to ensure that specifications are developed in a consistent fashion. Before that time, NQF will work with stewards as needed to determine the best method of submission. Some stewards will participate in the beta testing of the Measure Authoring Tool and will be able to submit eMeasures as a part of that activity. Regardless of whether the measures are submitted using the electronic tool or through other means, all measures will need to adhere to the standardized Health Quality Measure Format (HQMF).

Types of Applicable Measures

All measures relying on clinical data sources, such as medical record reviews and specialized data collection strategies (e.g., clinical registries or Category II CPT coding), will be required to submit eMeasures.

This requirement will not apply to measures relying on claims or administrative data, although it is anticipated that many of these measures eventually will be supplanted by eMeasures that take advantage of clinically rich information available in EHRs, as well as cost and other administrative information.

This requirement also will not apply to measures derived from patient surveys or reports. NQF is assessing the potential to capture patient-reported outcomes and other data using PHRs (or other health IT tools). Undoubtedly, this will be a very important area for future eMeasure development, but the measures and software tools are not yet ready for widespread application.

Requirements for eMeasure Testing by Stewards

The NQF Testing Task Force report made a clear distinction between the testing requirements for endorsed measures that have been re-specified for EHRs and testing requirements for newly submitted eMeasures. For endorsed measures that are retooled for EHRs before their regularly scheduled maintenance review, testing will focus on a crosswalk of the EHR measure specifications (QDM data elements, code lists, and measure logic) to the endorsed measure specifications. For newly submitted eMeasures or measures undergoing maintenance, the testing guidance is comparable to new measures based on other data platforms. The testing requirement for eMeasures is summarized in Table 1.

NQF Review Process for eMeasures

The eMeasures include a measure description in human readable form and a technical (XML) component. Steering Committees will evaluate the measure description for content using the NQF evaluation criteria. The measure's technical component will be reviewed by NQF health IT

staff to ensure adherence to the QDM and other guidelines, but primary reliance will be based on the "checks and balances" built into the Measure Authoring Tool and evidence of measure testing submitted by stewards.

312 **Table 1: Evaluation of Reliability and Validity of Measures Specified for EHRs**

	New Measure Specified for EHR		Modifications for Endorsed Measure *Re-specified* for EHRs
Rating	**Reliability Description and Evidence**	**Validity Description and Evidence**	
High	All EHR measure specifications are unambiguous[+] and include only data elements from the Quality Data Model (QDM)***** including quality data elements, code lists, and measure logic; **OR** new data elements are submitted for inclusion in the QDM; **AND** Empirical evidence of reliability of both data element **AND** measure score within acceptable norms: • Data element: reliability (repeatability) assured with computer programming—**must test data element validity** **AND** • Measure score: appropriate method, scope, and reliability statistic within acceptable norms	The measure specifications (numerator, denominator, exclusions, risk factors) reflect the quality of care problem (1a,1b) and evidence cited in support of the measure focus (1c) under *Importance to Measure and Report*; **AND** Empirical evidence of validity of both data elements **AND** measure score within acceptable norms: • Data element: validity demonstrated by analysis of agreement between data elements electronically extracted and data elements visually abstracted from the entire EHR with statistical results within acceptable norms; **OR** complete agreement between data elements and computed measure scores obtained by applying the EHR measure specifications to a simulated test EHR data set with known values for the critical data elements; **AND** • Measure score: appropriate method, scope, and validity testing result within acceptable norms; **AND** Identified threats to validity (lack of risk adjustment/stratification, multiple data types/methods, systematic missing or "incorrect" data) are empirically assessed and adequately addressed so that results are not biased	The EHR measure specifications use only data elements from the Quality Data Model (QDM)***** and include quality data elements, code lists, and measure logic; **AND** Crosswalk of the EHR measure specifications (QDM quality data elements, code lists, and measure logic) to the endorsed measure specifications demonstrates that they represent the original measure, which was judged to be a valid indicator of quality; **AND** Analysis of comparability of scores produced by the retooled EHR measure specifications with scores produced by the original measure specifications demonstrated similarity within tolerable error limits
Moderate	All EHR measure specifications are unambiguous[+] and include only data elements from the QDM;* **OR** new data elements are submitted for inclusion in the QDM; **AND** Empirical evidence of reliability within acceptable norms for either data elements **OR** measure score as noted above	The measure specifications reflect the evidence cited under *Importance to Measure and Report* as noted above; **AND** Empirical evidence of validity within acceptable norms for either data elements **OR** measure score as noted above; **OR** Systematic assessment of face validity of measure score as a quality indicator (as described in Table A-3) explicitly addressed and found substantial agreement that ***the scores obtained from the measure as specified will provide an accurate reflection of***	The EHR measure specifications use only data elements from the QDM as noted above **AND** Crosswalk of the EHR measure specifications as noted above demonstrates that they represent the original measure **AND** For measures with time-limited status, testing of the original measure and evidence ratings of moderate

	New Measure Specified for EHR		Modifications for Endorsed Measure *Re-specified* for EHRs
Rating	**Reliability Description and Evidence**	**Validity Description and Evidence**	
		quality and can be used to distinguish good and poor quality **AND** Identified threats to validity noted above are empirically assessed and adequately addressed so that results are not biased	for reliability and validity as described in Table 2.
Low	One or more EHR measure specifications are ambiguous[+] or do not use data elements from the QDM*; **OR** Empirical evidence of unreliability for either data elements **OR** measure score—i.e., statistical results outside of acceptable norms	The EHR measure specifications do not reflect the evidence cited under *Importance to Measure and Report* as noted above; **OR** Empirical evidence (using appropriate method and scope) of invalidity for either data elements **OR** measure score— i.e., statistical results outside of acceptable norms **OR** Identified threats to validity noted above are empirically assessed and determined to bias results	The EHR measure specifications do not use only data elements from the QDM; **OR** Crosswalk of the EHR measure specifications as noted above identifies that they do not represent the original measure **OR** For measures with time-limited status, empirical evidence of low reliability or validity for original time-limited measure
Insufficient evidence	Inappropriate method or scope of reliability testing	Inappropriate method or scope of validity testing (including inadequate assessment of face validity as noted above) **OR** Threats to validity as noted above are likely and are NOT empirically assessed	Crosswalk of the EHR measure specifications as noted above was not completed OR For measures with time-limited status, inappropriate method or scope of reliability or validity testing for original time-limited measure

313

DRAFT 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 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. NQF's goal is to endorse the best measure and minimize confusing or conflicting information.

Table 2. 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.
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 "best in class" determination at the time of three-year maintenance.

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. If the competing measure also is a new submission, the developers can be asked to address that question after the committee determines that both meet the evaluation criteria.

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. Table 3 provides an approach to the evaluation of competing measures for superiority or justification for multiple measures.

Figure 1. Addressing Competing Measures in the NQF Evaluation Process

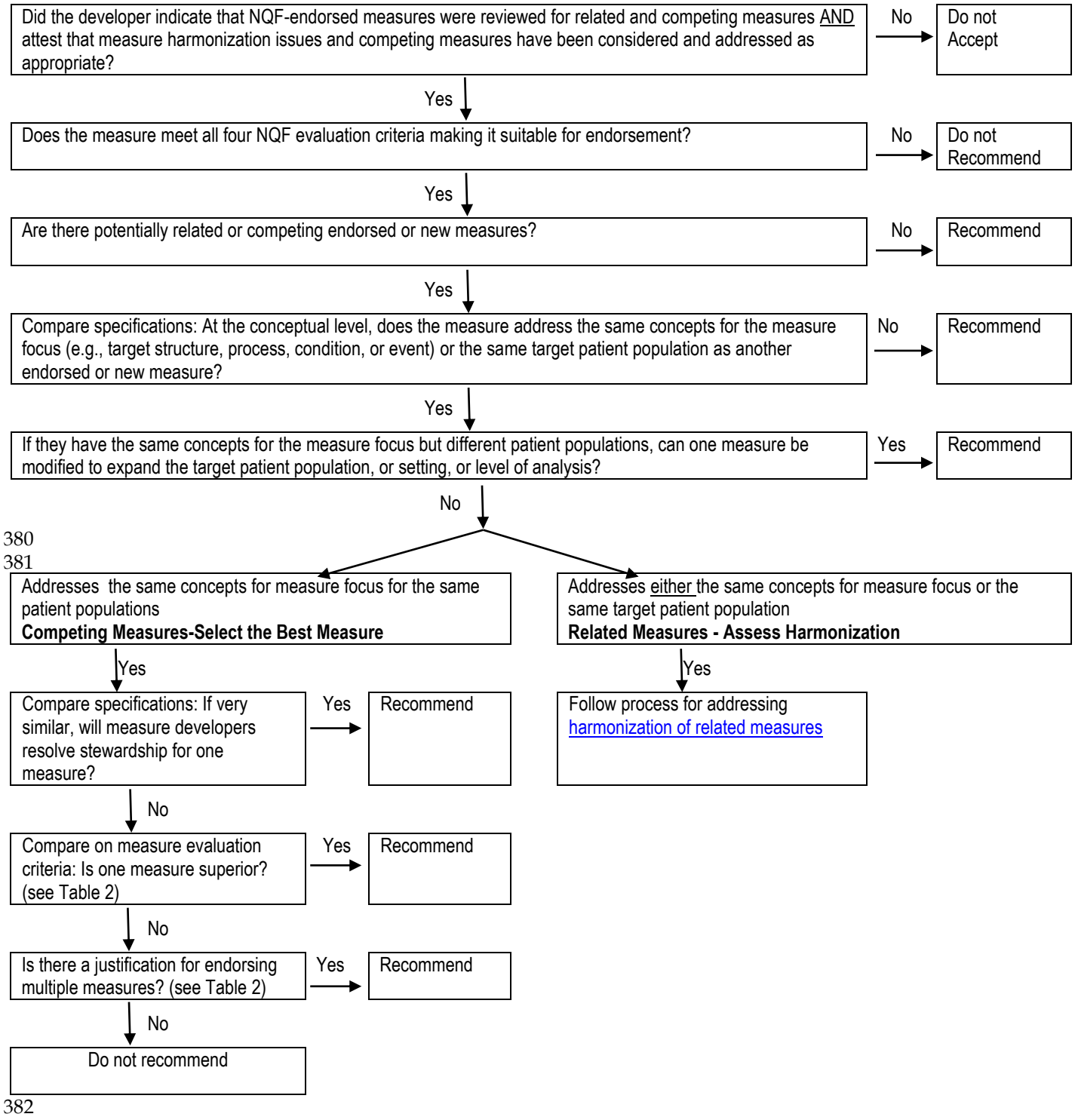

Table 3. Evaluating Competing Measures for Superiority or Justification for Multiple Measures

	Evaluate Competing Measures
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)
Assess competing measures for superiority on NQF Evaluation Criteria and Subcriteria	The comparison will require expert judgment and may involve considerations of pros and cons related to all the criteria. **Impact, Opportunity, and Evidence—Importance to Measure and Report:** Competing measures generally will be the same in terms of impact (1a) and evidence (1c) for the focus of measurement. • Compare measures on opportunity for improvement (1b). For new measures, this generally will be the same. However, measures in use or at the time of endorsement maintenance may differ in opportunity for improvement (e.g., one may be overall high levels of performance ("topped out). **Reliability and Validity—Scientific Acceptability of Measure Properties:** • Compare evidence of reliability (2a) • Compare evidence of validity (2b) 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 and the NQF endorsement maintenance cycles allow for regular submission of new measures.) Compare and identify differences in specifications. All else being equal, the preference is for: • Measures with the broadest application (target patient population, settings, level of analysis) • Measures that address disparities in care when appropriate **Usability:** • Compare evidence of use and usefulness for public reporting • Compare evidence of use and usefulness for quality improvement All else being equal, the preference is for: • Measures that are publicly reported • Measures with the widest use (e.g., settings, numbers of entities reporting performance results) • Measures that are in use over those without evidence of use **Feasibility:** • Compare the ease of data collection • Compare the potential for inaccuracies, errors, and unintended consequences All else being equal, the preference is for: • Measures based on data from electronic sources • Measures that are freely available

	Evaluate Competing Measures
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? **Value** 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). Is an additional measure unnecessary? • primarily for unique developer preferences **Burden** Do the different measures affect interpretability across measures? Does having more than one endorsed measure increase the burden of data collection? Measures based on different data types may provide added value if as noted above, the additional measure allows transition to EHR-based measurement or increases the number of individuals and entities for whom performance results are available. A rationale for recommending endorsement of multiple competing measures must be provided. Identify analyses needed to conduct a rigorous evaluation of the use and usefulness of the measures at the time of endorsement maintenance.

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 evaluated individually and must be determined to be suitable for endorsement before 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 may involve considerations of the pros

and cons related to all the criteria. For example, slightly lower reliability but much greater feasibility might indicate that 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. 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.