

TO: Consensus Standards Approval Committee (CSAC)

FR: Helen Burstin, MD, MPH and Taroon Amin, PhD, MPH

RE: Intended Use Advisory Panel Recommendations

DA: November 10, 2015

The CSAC will review recommendations from the Intended Use Advisory Panel at its November 18-19 in-person meeting.

This memo includes the background of the project, the Intended Use Advisory Panel's charge, and the Panel's final recommendations.

Accompanying this memo is the comment table that lists the 106 comments received from 23 organizations on the draft report and preliminary recommendations. The Advisory Panel thoroughly considered the comments provided and updated their recommendations accordingly. The recommendations provided in this memo reflect the revisions to the Panel recommendations made after input from the public and member comments.

CSAC ACTION REQUIRED

The CSAC will consider the initial recommendations from the Panel, public and member comments, and final recommendations from the Panel to determine a path forward for NQF.

Introduction

Over the past several years, various stakeholder groups have questioned whether NQF should incorporate the specific intended or actual use(s) of a measure (e.g., in particular federal quality programs) as part of the endorsement process, possibly via expansion of the [NQF measure endorsement criteria](#). This would move beyond the current Usability and Use criterion, which requires evaluation of the extent to which potential audiences (e.g., consumers, purchasers, providers, policymakers) 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.

This effort by the NQF Intended Use Advisory Panel seeks to consider the merit of, and the various approaches to, considering a measure's specific intended or actual use(s) as part of the measure endorsement process.

Background

Consensus Task Force

In 2012, NQF's Board of Directors (BoD) empaneled a Consensus Task Force (CTF) that included members of the NQF BoD, the Consensus Standards Approval Committee (CSAC), and individuals from the NQF membership. During its two-year tenure, the CTF reviewed the NQF endorsement

process and recommended several specific enhancements. As part of its final set of recommendations, in the fall of 2014, the CTF advised the NQF BoD to convene an Advisory Panel to consider transitioning from a binary endorsement decision (endorsed/not endorsed) to a more nuanced recommendation of endorsement. The CTF recommended that this Advisory Panel consider two potential options:

1. **Endorsement of measures for a specific intended or actual use(s):** Endorsement for specific purposes (e.g., internal quality improvement, public reporting, and payment) would assume that all measures may not be suitable for all potential uses. Currently, NQF endorses measures for quality improvement and accountability applications. Accountability applications generally have been defined as the use of performance measures for comparative purposes, rather than solely for internal quality improvement. Endorsement for intended use would potentially allow NQF to hold measures used for different purposes to different standards and would recognize that different stakeholders may have different priorities.
2. **Distinguish levels or grades of endorsement** (independent of use): Levels of endorsement would allow NQF to apply the same criteria to all measures, but to grade endorsed measures differently. Endorsement would change from a simple yes/no to an incremental scale; measures could move up or down the scale, with periodic evaluations to maintain endorsement based on additional testing or experience. Measures could be graded based on how well they meet NQF endorsement criteria, along with experience.

The Intended Use Advisory Panel

Acting on the recommendations of the CTF, in December 2014 the NQF BoD approved the convening of an Intended Use Advisory Panel. Members of this Panel were selected to represent expertise in all aspects of quality measurement, including measure developers, program implementers, providers, and others who are measured by or use NQF-endorsed performance measures.

The Intended Use Advisory Panel was charged with:

- Discussing several critical topic areas, including identifying various use cases for NQF-endorsed measures, distinguishing among the use cases, and identifying the need, if any, for different measure attributes, depending on the specific intended or actual measure use(s);
- Determining whether the NQF measure endorsement criteria requires updating;
- Proposing a path forward on whether, and if so, how, to incorporate the specific use of measures in the endorsement process.

Accordingly, this Advisory Panel did not design new processes nor develop new policies for NQF. Instead, their contributions should be viewed as foundational and strategic recommendations that will initiate and facilitate continued dialogue with NQF members and other stakeholders on this topic via various public commenting opportunities. The Advisory Panel considered public and member comments received prior to finalizing its recommendations and presenting them to the CSAC and NQF BoD.

The Intended Use Advisory Panel met via four web-meetings, as follows:

- June 8 & 10, 2015: Oriented to panel charge and key considerations outlined by the NQF BoD
- July 10, 2015: Considered the various uses for NQF-endorsed performance measures
- July 29, 2015: Considered how NQF endorsement criteria might vary based on the various use cases
- October 20, 2015: Reviewed public and member comments on draft report and updated recommendations

Panel Recommendations

The final five recommendations that emerged from the Advisory Panel deliberations are summarized below.

Recommendation 1: NQF should not try to distinguish between the measurement needs for pay-for-performance, public reporting, and other types of accountability applications.

As part of their consideration of whether NQF should endorse measures for specific use, the Advisory Panel considered at length the question of whether there are differences in the measurement needs among the various accountability applications (e.g. pay-for-performance, public reporting).

Several members of the Panel argued that there are differences based on the potential risk of misclassifying providers in particular programs and therefore in the subsequent financial impact to providers. Some Panel members noted that measures used for pay-for-performance programs, particularly programs that are designed as penalty programs, should demonstrate the highest adherence to requirements for measure testing and evidence. Members agreed that there is a considerable impact to the health care system if measures used in payment programs inappropriately misclassify providers and take away resources that are needed to support performance improvement.

Other members of the Panel argued that most accountability applications have a financial impact on providers, regardless of whether they have direct penalties. They noted that pay-for-performance programs, but also public reporting, health plan network design, and other programs, have the potential to have a financial impact on providers. The use of performance measures can also have a financial impact on providers by informing consumer choice on the selection of providers and directing patients away from low-performing providers.

Others argued that identifying the measurement needs of these applications based on financial impact to providers might be misguided. They maintained that the risk of misclassification of providers has both a financial impact and quality of care impact for patients, consumers, and purchasers. These panel members noted that the scientific rigor used to examine candidate measures should be the same regardless of the particular use as many of the accountability applications drive purchasing and selection decisions by purchasers and consumers. They

believe that measures that are endorsed by NQF should produce reliable and valid information, regardless of the intended use of the measure.

Recommendation 2: Create a “NQF+” designation for endorsed measures that exceed the criteria for endorsement in key areas, and include a requirement for vetting by those being measured

The Advisory Panel recommended introducing a “NQF+” designation for measures that have exceeded the criteria for endorsement in key areas. The following characteristics should be used to identify measures that achieve the “NQF+” designation:

- Meet or exceed evidence for measure focus sub-criterion without an exception;
- High or moderate certainty or confidence that the measure is reliable, as demonstrated by reliability testing of the measure score;
- High or moderate certainty or confidence that the measure is valid, as demonstrated by empirical validity testing of the measure score (i.e., not via face validity only); and
- Confirmation that the candidate measure is well vetted in real world settings by those being measured.

The first three elements needed for the NQF+ designation signify that a measure meets a higher bar for several key elements of the NQF endorsement criteria. To ensure greater consistency in Standing Committee ratings, it is recommended that NQF technical staff/consultants provide a preliminary analysis and rating on those must-pass criteria with detailed algorithms for rating (i.e., evidence, reliability, and validity). Increased transparency by using the “NQF+” designation will allow users to understand how well the measures meet these key criteria.

In addition to exceeding the criteria for endorsement in key areas, the Panel recommended including an evaluation of the measure experience in order to confer the “NQF+” designation. Specifically, this new element can help support the validity of the measure by demonstrating that a measure has been vetted in real world settings by those being measured, ensuring that the measure can be adopted and sustained in diverse contexts without undue burden or unintended negative consequences. This can be demonstrated when those being measured have been given:

- Measure performance results and data,
- Assistance with interpreting the measure results and data, and
- An opportunity to provide feedback on the measure performance and implementation by the community of entities being measured.

Although NQF staff will identify those candidate measures that meet the “NQF+” designation, Standing Committees will make the final decision regarding endorsement recommendations, including conferring of the NQF+ designation.

Recommendation 3: NQF endorsement should focus on endorsement of measures intended for accountability applications, such as public reporting, and payment applications

The Panel recommended that the NQF endorsement process should support the use of measures for accountability applications and performance improvement. The Panel agreed that measures that are endorsed for accountability applications can be used for internal quality

improvement (QI); however NQF should make it clear that it does not endorse measures for QI-only. Measures for QI-only are used by providers to evaluate their performance over time or among a group of providers, and are not typically used publicly for comparative purposes. The Panel agreed that QI-only measures are exceedingly important and that efforts to develop meaningful measures that facilitate both local improvement and shared learning should be accelerated. However, they did not support development of a new process by NQF to endorse or otherwise approve QI-only measures. The Panel acknowledged the public and member comments that supported the use of the National Quality Measures Clearinghouse, developed by the Agency for Healthcare Research and Quality (AHRQ) as a resource for identifying measures suitable for QI-only.

Recommendation 4: Encourage the Measures Applications Partnership (MAP) to consider how the NQF+ designation can be used in the selection of measures for programs

The Advisory Panel encouraged the Measures Applications Partnership (MAP) to consider how the “NQF+” designation can be used when selecting individual measures for specific programs. For example, in an effort to align program and measure attributes, the MAP may determine that an individual program requires “NQF+” measures. The Advisory Panel generally agreed that the MAP Coordinating Committee would be most appropriate to develop an approach for applying the “NQF+” designation in their future work.

Recommendation 5: Pursue future work to consider the interaction between program attributes and individual measure attributes

The Advisory Panel urged NQF to pursue future work to define key measure attributes and program attributes, examine their interaction, and give program implementers guidance on which measures may be better suited for implementation in specific programs based on program characteristics. The Panel identified a preliminary set of measure attributes that should be considered, including; (1) inclusion and exclusion criteria, (2) potential for misclassification based on reliability and validity testing results, and (3) the precision of the risk adjustment models. Further, they identified an initial set of program attributes to consider, including (1) methods used to define performance categories (e.g. measure score thresholds), (2) whether or not statistical tests are used to distinguish between performance categories and the approach to those tests, and (3) the nature of the financial incentive (e.g. tied to performance or improvement, upside or downside risk or both, etc.).

This work should begin with identifying and categorizing measure and program attributes, and move further to provide guidance on the interaction of these various elements. Such work would advance health care measurement science by identifying more precisely the likelihood that a measure will perform well in the context of a given program design. An examination of measure attributes and program attributes likely would reveal other key principles for the endorsement and application of measures. These key principles can help to create a test environment in which measures and programs may be more precisely matched to drive health system performance improvement.

Appendix A: Panel Members and NQF Staff

PANEL MEMBERS

Helen Darling, MA

National Business Group on Health
Chair, National Quality Forum Board of Directors
Washington, DC

Elizabeth E. Drye, MD, SM

Yale School of Medicine, Center for Outcomes Research & Evaluation
New Haven, CT

Lee A. Fleisher, MD

Perelman School of Medicine, University of Pennsylvania
Philadelphia, PA

Don Goldmann, MD

Institute for Healthcare Improvement
Cambridge, MA

Kate Goodrich, MD, MHS

Centers for Medicare & Medicaid Services
Washington, DC

Bruce L. Hall, MD, MBA, PhD

BJC Healthcare
Saint Louis, MO

Mary Beth Landrum, PhD

Harvard Medical School
Boston, MA

Beth A. McGlynn, PhD, MPP

Kaiser Permanente Center for Effectiveness & Safety Research
Pasadena, CA

Jonathan Perlin, MD, MSHA, PhD, FACP

Hospital Corporation of America
Nashville, TN

Andrew Ryan, PhD

University of Michigan Center for Health Outcomes and Research

Ann Arbor, MI

Cristie Upshaw Travis, MSHA

Memphis Business Group on Health
Memphis, TN

Lina Walker, PhD

AARP Public Policy Institute
Washington, DC

Rachel Werner, MD, PhD

University of Pennsylvania
Philadelphia, PA

NQF STAFF

Helen Burstin, MD, MPH

Chief Scientific Officer

Taroon Amin, PhD, MPH

Consultant

Suzanne Theberge, MPH

Senior Project Manager

Poonam Bal, MSHA

Project Manager

Kaitlynn Robinson-Ector, MPH

Project Analyst

The Panel would like to acknowledge **Reva Winkler, MD, MPH; Karen Johnson, MA, Elisa Munthali, MPH, and Marcia Wilson, PhD** for the input and comments provided draft and final recommendations.