Measure Feedback Loop—Usability and Use Final Report

TECHNICAL REPORT

JULY 3, 2019

NATIONAL QUALITY FORUM

This report is funded by the Department of Health and Human Services under contract HHSM-500-2017-00060I 75FCMC18F0007.

CONTENTS

EXECUTIVE SUMMARY

NQF evaluates measures for endorsement against five hierarchical criteria: Importance to Measure and Report, Scientific Acceptability of Measure Properties, Feasibility, Usability and Use, and Related and Competing Measures. NQF designed the Usability and Use criterion specifically to address the usability of the measure and to gain an understanding of how the measure is currently used or is intended to be used.

The consideration of feedback on the use of NQF-endorsed measures is also integral to the evaluation of feasibility and importance. In addition to the feedback that is collected through the submission form completed by the measure developer, NQF solicits feedback through other channels to support the Measure Feedback Loop Committee deliberations on Usability and Use in the endorsement process. The Consensus Development Process (CDP), the Measure Applications Partnership (MAP), and the NQF measure feedback tool all currently channel feedback including public comments into the NQF endorsement process. However, each of these channels comes with benefits and limitations, and there remains opportunity for improving them to ensure that the CDP measure feedback loop is effective.

With guidance from the Measure Feedback Loop Committee, this report explores the current application of the Usability and Use criteria (where measures are evaluated on the basis of implementation information), current practices for collecting feedback, challenges associated with each of these practices, recommendations for improving them, and new potential approaches for collecting feedback.

The Measure Feedback Loop Committee's recommendations centered on six key areas: (1) modifying the Usability and Use criteria and NQF **measure submission form**, (2) improving accessibility of commenting tools and opportunities to submit comments, (3) facilitating communication of feedback throughout the loop (defined in **Appendix A**), (4) targeting outreach to key stakeholders, (5) classifying feedback into key domains, and (6) developing guidance for measure developers. In conjunction with the prior environmental scan, this report will inform the Measure Feedback Loop Committee's future recommendations for feedback loop options and an implementation plan for a future measure feedback loop pilot.

BACKGROUND

The National Quality Forum's (NQF) Consensus Development Process (CDP) endorses performance measures that multistakeholder expert committees have judged to have met a series of criteria. Measure endorsement remains a vital step in the quality measurement enterprise (**Figure 1**). The CDP also serves as a mechanism for collecting and reviewing feedback on measures under consideration for endorsement. This information is collected from various stakeholder groups and through other measure evaluation processes conducted by NQF (e.g., the Measure Applications Partnership). Measures that are submitted to NQF for consideration for endorsement are evaluated against NQF's Measure Evaluation Criteria[1]: (1) Importance to Measure and Report, (2) Scientific Acceptability of Measure Properties, (3) Feasibility, (4) Usability and Use, and (5) Related and Competing Measures.

To understand the impact of measures on users and implementers, NQF incorporates an assessment of use, usability, feasibility, and opportunity for improvement into its criteria, and NQF assesses feedback for each of these elements. NQF designed the Usability and Use criterion specifically to address the usability of the measure and to gain an understanding of how the measure is currently used or is intended to be used. This criterion evaluates 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. Usability and Use are evaluated separately, and each has two subcriteria:

Usability

1. **Improvement**: Progress toward achieving the goal of high-quality, efficient healthcare for individuals or populations is demonstrated
2. **Unintended consequences**: Benefits outweigh harms: Progress toward achieving the goal of high-quality, efficient healthcare for individuals or populations is demonstrated

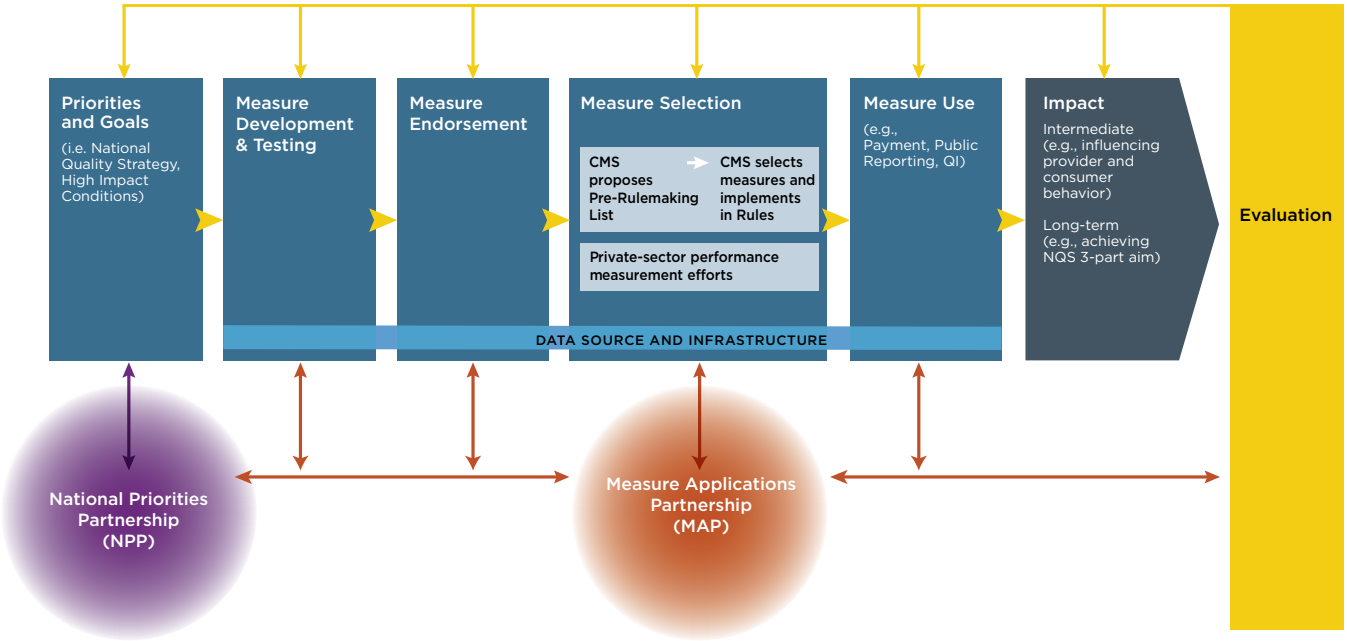
Use

1. **Accountability and transparency**: Performance results are used in at least one accountability application within three years after initial endorsement and are publicly reported within six years after initial endorsement
2. **Feedback on the measure by those being measured or others**
    a. Those being measured have been given performance results or data, as well as assistance with interpreting the measure results and data
    b. Those being measured and other users have been given an opportunity to provide feedback on the measure performance or implementation
    c. This feedback has been considered when changes are incorporated into the measure

Per the guidance of the 2012 Use and Usability Task Force, Usability and Use are evaluated last in the hierarchy because it is assumed that a measure should be usable if it passes the prior criteria. In addition to Usability and Use, there are other criteria intended to solicit information related to feedback for the measures. Within the Importance criteria, evaluators must consider whether the measure's performance scores over time demonstrate that the implementation of the measure has had some impact on quality and performance and whether there remains a gap in performance that warrants measurement. Within Feasibility, evaluators are asked to assess whether the data collection strategy for implementing the measure presents any challenges to collecting the necessary data to implement the measure (e.g., costs, missing data).

Quality Measurement Enterprise

FIGURE 1. THE QUALITY MEASUREMENT ENTERPRISE



The evaluation of measures for endorsement does not solely depend on information provided by the measure developer in the measure submission form. Comments submitted during the measure endorsement process and the Measure Applications Partnership (MAP) process are also collected and collated for consideration during evaluation. NQF's online **measure feedback tool** is another source of information used to collect feedback on measures that is considered during evaluation. Any comments related to the specifications or experience with the measure that are submitted through the tool are provided to the CDP standing committee for consideration during its evaluation.

The measure submission form (guided by NQF evaluation criteria), public comments, and the NQF measure feedback tool are all channels currently in use that feed into the NQF endorsement process. However, each has its benefits and limitations, and there remains opportunity for improving these channels to ensure the measure feedback loop is effective.

To guide this effort, NQF convened the Measure Feedback Loop Committee (**Appendix C**), which consists of 21 representatives of various stakeholder groups, including measure developers, implementers, clinicians, NQF member organizations, and patient advocates. Many representatives also bring perspectives from prior involvement with NQF's measure endorsement process and their own experiences with measure implementation and feedback loops. With guidance from the Committee, this report will explore the current application of the Usability and Use criteria, current practices for collecting feedback, challenges associated with each of these channels, and recommendations for improving them.

PURPOSE

This report summarizes the current processes and channels for soliciting and collecting feedback for the Consensus Development Process. This report also provides an analysis of the Usability and Use criteria, as well as the other NQF criteria that rely on measure feedback. Together with the findings from the previous environmental scan report, the recommendations for improving current feedback loops within NQF processes will be considered as the foundation for designing potential feedback loops and a plan to implement a pilot of a selected process.

HISTORY OF NQF USABILITY AND USE CRITERIA

NQF established its Measure Evaluation Criteria to guide expert committees assessing measures for endorsement on the key aspects of a performance measure that must be met to enable broad application, implementation, and comparisons. These evaluations rely on information submitted by the measure developer organization via the NQF **Measure Submission Form**. This submission form was designed to collect all of the information necessary for NQF's multistakeholder expert committees to apply the evaluation criteria and make recommendations for endorsement. The submission form mirrors the evaluation criteria and is designed to capture enough information to inform recommendations without burdening measure developers unnecessarily.

The evaluation criteria reflect best practices in measure development and have evolved over time to maintain relevancy in a dynamic quality measurement landscape and to meet various stakeholder needs. NQF's Usability and Use criterion has also evolved significantly since its creation. Multistakeholder input and critical milestones in performance measurement guided each iteration. NQF's original definition of the Usability criterion focused on the intelligibility of the measure results and was described as:

> *Extent to which intended audiences (e.g., consumers, purchasers, providers, policymakers) can understand the results of the measure and find them useful for decision making.*

This description of the criterion reflects the relatively early stage of healthcare performance measure implementation and endorsement. As robust national reporting programs were not yet widely adopted, NQF approached the question of usability of a measure from a theoretical perspective, without requiring much evidence of effective implementation or public reporting. The summary below describes the major milestones in the refinement of the criterion to its current state.

Expansion of the Usability Criterion (2012)

In 2012, NQF convened a Usability Task Force to review and refine the Usability criterion, in response to the ongoing challenge of evaluating measures that are not implemented at the time of endorsement maintenance (typically three

years after initial endorsement), and to broaden the scope of evaluating measure use beyond public reporting functions, to include value-based payment, health information technology incentive payments, accreditation, and regulation.[2] The Task Force modified the Usability criterion, expanding the scope to include demonstration of "Use" as a requirement. This expansion was intended to emphasize NQF's desire that endorsed measures are being used and are included in an accountability application within three years after endorsement, and are publicly reported within six years after endorsement. Within "Usability," the Task Force added a subcriterion requiring developers to demonstrate that the measure results could be used to facilitate improvement in healthcare quality, and recommended that potential unintended consequences be evaluated as part this criterion.

The Task Force issued several recommendations vital to understanding NQF's approach to feedback on measures in use at that time. First, the Task Force suggested that at the time of submission, new performance measures should not be expected to supply use and performance results. They also recommended that Usability and Use be evaluated last, as a measure found to meet the first three criteria (Importance to Measure and Report, Scientific Acceptability, and Feasibility) is "almost certain to be potentially usable." The Task Force cited dueling concerns about requiring public reporting of all NQF-endorsed measures. Some worried that an excess of measure results to consult when selecting a provider would overwhelm the public, while others worried that very few measures would be publicly reported without this requirement.

The Task Force emphasized that an accountability application is a necessary precondition to the successful use of a measure because measures endorsed by NQF are intended to help consumers make selections and to drive healthcare improvement by ensuring accountability and providing for incentive structures. The Task Force also identified several barriers to public reporting, including the then-limited adoption of electronic health record systems, limitations on access to data, or even variations in funding or competing priorities by program administrators. The Task Force recommended that developers engage with potential implementers during development and testing of measures and before submission for endorsement.

Emphasizing Usability and Use for Maintenance Measures (2015)

Every three years, endorsed performance measures must be re-submitted to NQF for re-evaluation against the endorsement criteria in order to maintain endorsement. In August 2015, NQF restructured the endorsement maintenance process to shift the emphasis of the re-evaluation of maintenance for certain criteria. The goal of the restructuring is to emphasize those aspects of measure submissions that are most likely to change in between evaluations, and to provide an opportunity to review public and NQF member comments received on the measures since the prior endorsement.

Although the criteria themselves did not change, the emphasis on Importance to Measure and Report and Scientific Acceptability of Measure Properties was decreased, particularly when there were no material changes to the evidence and when the developer did not update testing. In these cases, the CDP Standing Committee would no longer be required to re-adjudicate or vote on these criteria. However, there was a shift to greater emphasis on the Usability and Use criterion. At that time, standing committees had indicated a desire to have more comprehensive discussions of "what has been learned about previously endorsed measures," including "how has the measure been used, and unintended consequences"—effectively, measure feedback.

Assessment of Intended Use (2016)

In 2012, NQF's Board of Directors convened a Consensus Task Force (CTF) that included members of the NQF Board, the Consensus Standards Approval Committee (CSAC), and representatives from the NQF membership. During its two-year tenure, the CTF reviewed the NQF endorsement process and recommended several enhancements. The CTF recommended that NQF convene an Advisory Panel to consider whether measures should be endorsed for specific use cases and assess the viability of transitioning from a binary endorsement decision (endorsed/not endorsed) to a more nuanced recommendation of endorsement based on how well measures met the criteria. In 2015, NQF convened the Intended Use Advisory Panel to assess how a measure's specific intended use (e.g., internal quality improvement, public reporting, value-based purchasing) should be considered in its endorsement recommendation.

This Panel examined whether NQF-endorsed measures may only be applicable in certain use cases. Some stakeholders expressed concern that misclassifying providers in some accountability programs, especially those with payment penalties, could have severe consequences, by misallocating resources that could be directed towards patient care. Ultimately, in the 2016 report, the Panel recommended that NQF endorsement should not distinguish based on the intended use or application of the measure and that the evaluation criteria should be applied equally to all measures regardless of their intended use. However, the Panel recommended that a new Endorsement+ designation be created to identify measures that have exceeded the criteria for endorsement in key areas, included usability and use. Specifically, the Panel highlighted the measure being "well-vetted in real world settings by those being measured and others" as a key component. Although NQF did not ultimately adopt the Endorsement+ designation, the recommendation did lead to the creation of a feedback-focused subcriterion, described below.

Addition of a Subcriterion Focused on Feedback on Measures (2016)

With inputs from the Intended Use Advisory Panel, NQF's Consensus Standards Approval Committee (CSAC) recommended adding an additional subcriterion to Usability and Use to assess the degree to which a measure is "well-vetted in real world settings by those being measured and other users." The CSAC is an advisory committee whose members are appointed by the NQF Board of Directors. It reviews measure endorsement recommendations of multistakeholder NQF standing committees, which are convened in topical areas to review and recommend submitted standards (measures) for endorsement. The feedback collected by developers for the evaluation of the measure would serve to demonstrate that the measure can be implemented without undue burden and unintended consequences. The key elements of this subcriterion were described in the following ways:

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.

Including this new subcriterion was intended to be a "signal to the field about its importance." Indeed, this was NQF's first effort to collect measure feedback as part of the measure submission, and not solely as part of the public and member comment periods. At the time of implementation, these same questions were asked of developers when they updated their endorsed

measures annually. This was intended to facilitate the collection of substantive measure feedback, without having to wait three or more years for the measure to be re-submitted to NQF for endorsement maintenance.

Use Becomes a "Must-Pass" Criterion (2017)

With greater emphasis on Usability and Use in the endorsement process, there was a desire to further align the criteria with this emphasis and demonstrate NQF's desire that endorsed measures be in use in some accountability application. There was consensus among the CSAC that at the time of endorsement maintenance, measure developers should be required to demonstrate that the measure has been implemented in a public reporting or other accountability program (e.g., value-based purchasing, physician profiling) since its endorsement. In order to pass the Use criterion and be re-endorsed, a maintenance measure would have to satisfactorily demonstrate its use. The CSAC recognized that this requirement would make it more difficult for some measures to maintain endorsement given the difficulty for some developers to obtain information on the use of their measures. However, even with these challenges, it was important for NQF-endorsed measures to meet this standard.

APPLYING THE NQF USABILITY AND USE CRITERIA

Principles for Evaluating Usability and Use

In the 2012 *Guidance for Evaluating Usability and Use of Performance Measures*,[2] the Usability Task Force outlined several principles designed to guide CDP standing committees and other stakeholders on the intent and application of the criteria. For this effort, the Measure Feedback Loop Committee was tasked with reviewing these principles within the context of the current performance measurement landscape to determine whether they still hold true and if any should be modified, added, or removed. The Measure Feedback Loop Committee agreed that the following principles remained relevant and reflect the appropriate guidance for evaluating Usability and Use for endorsement. Additional context and explanations of the applications of the principles can be found in the **2012 report**.

- Performance measurement facilitates achieving the goal of high-quality, efficient healthcare for all Americans through two pathways: (1) changes in care initiated by healthcare providers and (2) accountability/selection by making information available to consumers, referring clinicians, and others involved in selecting clinicians and providers.
- NQF strives to endorse measures that are useful for *both* accountability and improvement to maximize their influence on progress toward the goal of high-quality, efficient healthcare for all Americans.
- Public disclosure of performance results not only is necessary for some types of selection such as consumer choice, but also ensures accountability and provides external motivation for performance improvement (i.e., impetus to improve outside of drivers within the health system).

- The NQF criteria of Importance to Measure and Report and Scientific Acceptability of Measure Properties ensure that a measure is potentially useful for a variety of applications.
- Measures can be more or less useful to intended audiences depending on the conditions of implementation for a specific purpose (e.g., whether reporting methods or classification methods obscure differences in performance, like determining whether a three versus four star rating represents true differences in performance).
- The NQF criterion of Feasibility, particularly regarding the data required to implement a performance measure, also influences usability (e.g., data collection strategies that are difficult to implement may impact accuracy and usability of measure results). However, feasibility issues may be mitigated, or the benefit of measuring performance may outweigh associated burden.
- Evaluate potential usability for new measures and actual use and performance results of measures subject to endorsement maintenance.
- Set expectations for timeframes to achieve use in accountability applications and public reporting, but allow flexibility.
- The benefits of measurement in terms of facilitating improvement should outweigh evidence of unintended negative consequences.
- Address Usability and Use last in the hierarchy of the four major criteria, because if the other criteria are met, then a measure should be usable.
- To achieve maximal effect on quality healthcare and health, over time, NQF-endorsed measures should be used in all applications (e.g., decision making, public reporting, payment) for which they provide useful information.

Recommended Revision

The Measure Feedback Loop Committee recommended revisions to one principle.

Principle: Measure developers may not be responsible for implementing performance measures for accountability/selection or quality improvement programs and may not have access to the required data or information about measure use.

Suggested revision: Measure developers may not be responsible for implementing performance measures for accountability/selection or quality improvement programs and may not have access to the required data or information about measure use, but should be responsible for gathering and responding to feedback on their measures to the extent of their capabilities.

Rationale: The Measure Feedback Loop Committee recognized that the measure developer organizations are diverse in their infrastructure, capabilities, and intent for developing measures. Depending on the nature of the developer's relationship with the intended audience and users of the measure, they may have limited access to feedback from the implementers. Conversely, there are developers that may also be the implementer (e.g., Quality Clinical Data Registry) and will have the ability to collect substantial information and feedback on the implementation of the measure that should be considered during initial endorsement and maintenance. However, regardless of the proximity of the developer to the implementation, the Measure Feedback Loop Committee agreed that the measure developer has the responsibility to seek out information on the use and implementation of their measure and make changes to the measure as needed based on feedback received.

New Principle

The Measure Feedback Loop Committee recommended the addition of one principle for consideration in the application of the Usability and Use criterion.

New Principle: Usability should be considered in the context of the intended audience.

Rationale: The Committee noted the need for the developer to clearly define the audience for whom the measure is intended, in addition to the intended use, so that usability can be evaluated from the perspective of those for whom the measure is intended. The usability of a measure intended for internal quality improvement of certain physician practices may be perceived differently by the physicians being measured and a patient or consumer. This is particularly challenging for audiences like consumers, patients, or caregivers who seek useful information for healthcare decision making, but are relying on measure results intended for other purposes or audiences.

A Committee member suggested that measure developers conduct end user testing with the patients, families, and caregivers when developing public facing measures. Incorporating this perspective prior to publication may help clarify the results for members of the public. However, the Measure Feedback Loop Committee did recognize that measures often are used and implemented in applications other than those initially intended. The context of the intended audience is necessary in order for evaluators to weigh the relevancy of the feedback and make an informed endorsement decision.

For example, the Measure Feedback Loop Committee acknowledged the challenges that some providers face in identifying measures for their use that align with their practice and specialty. This becomes particularly challenging when certain measures that may be more relevant to one's practice are only available to report on through channels (e.g., clinical data registry) other than or in addition to what is currently used (e.g., federal programs). The burden of reporting through multiple channels outweighs the selection of more relevant measures. For providers that resort to reporting on measures that may be less relevant to their practice, the limited actionability of the measure results influences the perceived usability of the measure results. For the specialist who may manage some aspects of a patient's care but not others, it becomes less clear who is responsible for influencing the improvement of the measure score and how that provider can act to improve that score.

Usability Criterion

The Usability criterion is intended to assess whether the measure is helping to drive improvements in quality and health outcomes while not causing any unintended harms.

These objectives are described in more detail below:

- **Improvement**: Progress toward achieving the goal of high-quality, efficient healthcare for individuals or populations is demonstrated.
- **Benefits outweigh harms**: The benefits of the performance measure in facilitating progress toward achieving high-quality, efficient healthcare for individuals or populations outweigh evidence of unintended negative consequences to individuals or populations.

In order to assess this criterion, the developer should provide information on both actual and envisioned improvements, and unanticipated benefits demonstrated from the implementation of the measure. This is weighed with information provided by the developer on any actual or anticipated unintended consequences; the Measure Feedback Loop Committee rates Usability on a scale of high, moderate, or low. The CDP measure submission form questions related to showing improvements in quality enable the developer to demonstrate the scope and use of the measure, including a description of the trends in the performance results over time, the number

and percentage of people receiving high-quality healthcare, the geographic areas, and the number and percentage of accountable entities and patients reflected in the data. If a developer finds that no improvement has been realized since implementing the measure, the developer will be asked to provide a rationale. For new measures or those not in use at the time of initial endorsement, the developer is asked to provide a credible rationale that describes how the performance results could be used to drive improvements in quality, health outcomes, or efficiency.

Under the current process, responses to these questions in the submission form are solely provided by the measure developer. This presents several challenges. Often, this field is left blank, incomplete, or marked not applicable. When a response is provided, the response typically describes the process in which feedback is collected, but does not list any results. For new measures that have not yet been put into use, these items are particularly challenging to address. This lack of information is often attributed to measure developers' lack of knowledge of how the measure is being used and who is using it. Response to these questions would require feedback from implementers to gain a sense of the measure performance and the experience of those who have been using the measure. Recommendations for collecting additional feedback to fill this deficit in reported feedback are discussed in the **Recommendations for Improving Feedback Channels** section of the report.

Especially for maintenance measures, the information requested for the evaluation of Usability closely aligns with the data that are requested for the Importance criteria. The subcriterion of Opportunity for Improvement assesses the developers' demonstration of a gap in performance that warrants measurement. A measure that has been shown to drive improvement may also be closing the gap in performance, thus affecting the evaluation of its importance.

Use Criterion

The Use criterion became a "must-pass" criterion for maintenance measures in 2016. The objective of this criterion is two-fold—to assess transparency of the measure in accountability functions and to understand how the measure developer has solicited and adjudicated feedback. These objectives are described in more detail below.

- **Accountability and transparency**: Performance results are used in at least one accountability application within three years after initial endorsement and are publicly reported within six years after initial endorsement.
- **Feedback by those being measured or others**: Those being measured have been given results and assistance in interpreting results; those being measured and others have been given opportunity for feedback; the developers have considered the feedback.

In order to assess the adequacy of the measure submission against these objectives, the developer is asked to provide the sponsor and name of the program(s) in which the measure is being used, the purpose of the program, the geographic areas in which the measure is in use, the number and percentage of accountable entities, the number of patients included, and level of measurement (e.g., hospital, heath plan), and setting for each current use. If the measure is not currently publicly reported or used in another accountability application, the developer is asked to provide a rationale and a credible plan for implementation within a specified timeframe—any accountability application within three years and publicly reported within six years of initial endorsement.

For the second objective of the criterion, the developer is asked to summarize its activities in collecting and responding to feedback received on the measure in seven questions:

- Describe how performance results, data, and assistance with interpretation have been provided to those being measured or other users during development or implementation.

- How many and which types of measured entities and/or others were included? If only a sample of measured entities were included, describe the full population and how the sample was selected.
- Describe the process(es) involved, including when/how often results were provided, what data were provided, what educational/explanatory efforts were made, etc.
- Summarize the feedback on measure performance and implementation from the measured entities and others. Describe how feedback was obtained.
- Summarize the feedback obtained from those being measured.
- Summarize the feedback obtained from other users.
- Describe how the feedback has been considered when developing or revising the measure specifications or implementation, including whether the measure was modified and why or why not.

These questions provide an opportunity for the developer to describe its feedback collection process, the nature of the feedback received, and how the developer has tailored measure specifications, implementation guidance, or other measure features to respond to the feedback. However, similar to the challenges with the responses to the questions on Usability in the submission form, responses to the questions on Use are also typically deficient. The Measure Feedback Loop Committee had extensive discussion regarding the challenges developers face with providing adequate and meaningful information in response to feedback-related measure submission items. A Committee member noted that for measures addressing certain populations, such as pediatrics, there are significant gaps in opportunity for measures to be included in federal-level quality reporting programs. This lack of available programs can make fulfilling NQF's requirement on use in an accountability and a public reporting program challenging.

For developers that create measures that are not in use in federal programs, identifying entities that are using the measure also presents a challenge. As a result, it is possible for a measure developer to not receive a significant amount of feedback on its measure. Many developers may not know for whom or how their measures are being used. CDP Committees rate this criterion as either "pass" or "no pass." Since Use has become a "must-pass" criterion for maintenance measures, several measures have failed to obtain endorsement, or developers have chosen not to resubmit as a result of this requirement.

Committee Discussion and Recommendations

The Measure Feedback Loop Committee reviewed the Usability and Use criterion, its intent, and the submission form and discussed the challenges with applying the criteria and collecting the relevant information from measure developers as a primary source. In addition to the domains of feedback collected within the Usability and Use criterion, feedback is also collected to support the evaluation of the importance criteria and the feasibility criteria. The Measure Feedback Loop Committee considered the six domains of feedback collected across all of the criteria and determined that the scope of feedback sought through the current criteria is adequate.

- Opportunity for improvement (Importance to Measure and Report)
- Data collection strategy (Feasibility)
- Accountability and transparency (Use)
- Feedback from those being measured and others (Use)
- Improvement (Usability)
- Benefits outweigh harms (Usability)

The Measure Feedback Loop Committee underscored the importance of "closing the loop," or ensuring that endorsement committees have access to all feedback that has been gathered on a performance measure at the time of initial endorsement and maintenance reviews. The Committee identified two recommendations for consideration in modifying NQF's criteria and submission form to improve the feedback collected by developers.

The Measure Feedback Loop Committee discussed challenges that measure developers confront when they try to track and influence the use of their measure. Many developers do not have a mechanism to directly implement their performance measures in any setting, much less publicly report the results. In many cases developers may not know whether the measures they have published are in use in smaller regional or state programs. Typically, measures are published in the public domain without licensing requirements. This dilemma may also present for well-known measures in use when an established channel for sharing feedback is lacking.

These challenges present a misalignment of the locus of control of the developer regarding implementation of the measure and its responsibility to steward the measure. As such, developers face considerable challenges in collecting feedback from those being measured and other users. However, the Measure Feedback Loop Committee made an important distinction between the challenges a developer might face in collecting information about a measure's use, and the measure developer's responsibility to indicate the intended use of the measure as part of its submission. The Committee maintained that this remains the developer's responsibility, specifically stating the preferred data source, level of analysis, and intended audience. The Measure Feedback Loop Committee also acknowledged that NQF does hold a significant role in facilitating communication between users and developers through its channels as the steward of endorsement, particularly because this information is needed to make informed endorsement recommendations. The Committee's current recommendations will be paired with findings from the previous environmental scan report to develop options to strengthen the existing feedback loop process and foster new feedback loops, with the goal of informing measure evaluation deliberations in the CDP.

Recommendation 1: **Use should remain a "must-pass" criterion for maintenance measures, but with the opportunity to qualify for exceptions.** The Measure Feedback Loop Committee agreed that there may also be certain allowable justifications for a maintenance measure to not meet the requirement of use in an accountability application at three years after endorsement, and public reporting after six years. Many programs using performance measures may change measure sets from year to year in order to reflect contemporary priorities; performance measures discontinued for use may be picked up again in future reporting years. Programs may also shift measures in use based on changes to measure specifications and progress in meeting quality goals. However, the Measure Feedback Loop Committee emphasized the importance of furthering public understanding of performance measurement by reporting the rationale for these changes, and clearly disclosing where specifications that substantially affect the measure result may have changed.

Recommendation 2: **The subcriterion, feedback on the measure by those being measured or others, should be considered under Usability.** Currently this subcriterion is considered under Use. The focus of the Use criterion is to determine who is using the measure, who is impacted by it, how the measure is currently being used, and whether it is being used in an accountability application. The information sought by the feedback subcriterion is certainly impacted by use, but it is questionable whether it should be entangled with Use and also considered as "must-pass." The information collected by the developer to address this question speaks directly to the usability of the measure and should be considered as such.

OTHER CHANNELS FOR MEASURE FEEDBACK

In addition to what is collected through the submission form completed by the measure developer, NQF solicits feedback through other channels that are also used to support CDP committee deliberations on Usability and Use in the endorsement process. Information collected via the Measure Applications Partnership deliberations and commenting process, endorsement process commenting period, as well as NQF's online measure feedback tool, are all collated and provided to the relevant endorsement committees for consideration. A description of each feedback channel and how it collects and distributes information follows.

MAP Deliberations and Pre-rulemaking Public Comments

NQF convenes the MAP as a public-private partnership of healthcare stakeholders to provide input to the Department of Health and Human Services (HHS) on the selection of performance measures for public reporting and performance-based payment programs. During the annual MAP pre-rulemaking process, NQF solicits feedback on candidate measures by publicly publishing the list of quality and cost measures that HHS is considering for adoption through rulemaking under Medicare.

The MAP public comment process yielded 361 public comments on 40 measures in the 2018-2019 cycle. Feedback included comments on risk-adjustment methodology, specifications, exclusions, data sources, implementation, and possible measure burden. Feedback from the MAP deliberations focused on recommendations for multistakeholder review and concerns over how measures align across payment programs. The MAP commenting process frequently yields comments from stakeholders on unintended consequences of specific measures under consideration. However, the MAP process lacks any process that enables the developer to respond to the comments or that facilitates the communication of a response from the developer back to the commenter. The comments, along with a summary of the MAP deliberations, are sent directly to the MAP Workgroups for their consideration during the MAP process as part of the discussion guide. The recommendations from the MAP process are incorporated into the CDP preliminary analyses and sent to the CDP committee for consideration during endorsement review.

The timing of the MAP process faces challenges for multiple stakeholders involved, including NQF endorsement committees, CMS, and the measure developers. Many CMS programs require that measures in the program are NQF-endorsed. As a result, MAP often recommends measures with the condition that the developers submit them to NQF for endorsement consideration. However, as MAP recommendations are published annually on February 1, developers may not have had an opportunity to submit measures by the January 1 Intent to Submit deadline, meaning they would have to wait to submit for endorsement review until August 1 of that same year. In practice, an endorsement decision would not be made official until June of the following year, meaning developers must wait nearly a year and half between MAP and CDP verdicts. The misalignment of the timing of the MAP recommendations, the NQF submission process, and the final rule for the program, does not always allow developers to submit for endorsement with the recommended revisions in a timely manner.

Further, feedback provided during both the MAP and the CDP processes is collected after the measure developer has completed the measure development and testing process. Comments submitted that raise issues with specifications or other issues that require testing cannot be

completed within the CDP or MAP timelines. Developers are often faced with the challenge of reconvening their own expert panels used to guide measure development to vet the comments and recommendations in a timely manner and balance their resources with addressing high-priority issues in ongoing revisions to the measures.

Public Comments Received during Endorsement Consideration

In 2017, NQF extended its commenting period to include both the period prior to CDP standing committee evaluation, and after CDP standing committee evaluation—creating one continuous public commenting period spanning 16 weeks. The commenting period opens approximately three weeks prior to the standing committee evaluation meeting and closes 30 days after NQF staff posts the draft technical report on the NQF website. Announcements regarding NQF's public commenting periods are distributed via email to NQF members and to individual members of the public who have opted to receive email communications from NQF for specific topic areas or projects. In order to comment on any measure, a commenter must acquire an NQF login to access the commenting website. Commenters are required to log in to the site to ensure that comments can be attributed to the commenter. Comments are solicited through an open text box associated with each individual measure under consideration. Each measure evaluation call is also open to the public and allows for members of the public to provide an oral comment following CDP committee evaluation of the measure.

The public comment process of the CDP has yielded 229 comments on 56 measures from November 2017 to February 2019. NQF has noted a significant decline in the number of public comments received during the CDP process in recent years despite efforts to extend the period of commenting for each topic area. The cause of this decline is likely multifaceted and related to the changes to NQF's endorsement schedule and the significant increase in the volume of endorsement work. Public commenting during measure evaluation calls is also typically underutilized.

For those written comments that have been submitted, the common themes center on concerns about a measure's specifications, exclusions, risk-adjustment methodology, and data sources. Comments also frequently address issues of implementation and possible burden on clinicians and other care providers. The comments from the most recent commenting period are also collated and included in the materials for the CDP committee to consider during the evaluation of the measures. Those comments received prior to the committee's evaluation of a measure are collated for discussion, and the developer is not required to submit a written response due to the time limitations. Developers submit individual responses to comments received after a committee's deliberations and endorsement recommendations. Recommendations for collecting additional feedback to fill this deficit in reported feedback are discussed in the **Recommendations for Improving Feedback Channels** section of the report.

NQF Measure Feedback Tool

Prior to 2016, NQF members and the public could only comment on measures when they were actively under review in a CDP project. In 2016, NQF created an online **measure feedback tool** designed to collect feedback on endorsed measures from those using them at any point following the CDP endorsement. Comments submitted via this tool are entered in a separate online platform than that used for commenting during endorsement consideration. The tool provides a channel for receiving feedback on endorsed measures and sharing other comments received from others. The tool can be accessed online through NQF's homepage via the "NQF Work" menu item or via NQF's **Quality Positioning System** (QPS). This tool enables stakeholders to submit feedback on measures anytime, even when they are not under endorsement review.

Comments are collected via an open text box allowing the commenter to submit a comment on any topic. When the tool was launched, it was designed to capture unstructured comments in an effort to limit the burden of responding to structured questions or specific topic areas. Once comments are submitted, they can be viewed by others who access the measure comments. Currently there is no established, systematic mechanism for pushing comments back to the developers; developers must access the feedback on their measures on their own. Feedback submitted on the tool is collated for inclusion in the CDP committee materials at the time of endorsement maintenance consideration.

Since early 2017, NQF has received only 19 feedback submissions via this tool. Information received from the feedback tool includes comments on exclusions and other aspects of measure specifications, as well as questions on NQF processes. This feedback was submitted by health systems, advocacy groups, and specialty societies. NQF and the Measure Feedback Loop Committee recognize that this channel is significantly underused and could be enhanced with improved marketing, promotion, and communication and by improving visibility and access to the tool on the website. Recommendations for collecting additional feedback to fill this deficit in reported feedback are discussed in the next section of the report.

RECOMMENDATIONS FOR IMPROVING FEEDBACK CHANNELS

Improving Access and Opportunities to Submit

The Measure Feedback Loop Committee noted that although NQF had provided several avenues for measure developers and users to provide feedback on performance measures, there are barriers to accessing them that contribute to their limited use, particularly to those stakeholders less familiar with the field of performance measurement. For example, NQF-sponsored commenting features require that a commenter create an NQF login name and password. Although these accounts are free and include commenting privileges, users may forget a password when returning, or otherwise be confused by the sign-on process. While authentication features help attribute submitted feedback to a particular stakeholder perspective, and may help reduce the possibility of abuse of the feedback features, alternatives to the current authentication process could generate additional feedback for consideration in the CDP process.

The website design also presents some challenges. After signing onto the site, some features, such as the list of current projects and the QPS measure database tool, have been described as potentially confusing, especially when attempting to identify opportunities to submit feedback to active evaluation processes.

The Measure Feedback Loop Committee also noted that those being measured—as well as process managers and other individuals who are tasked with the implementation and operationalization of performance measures—would benefit from having more direct access to feedback submission tools. At present, in order for feedback on a measure to reach a CDP standing committee, the feedback reporter must know the measure is NQF-endorsed and be aware of the feedback tool and how to access it. In practice this appears to be quite rare, as so few have used the tool. The Measure Feedback Loop Committee recommended that NQF partner

with organizations implementing performance measures, including registries, the Centers for Medicare and Medicaid Services, and others, to add direct links to the NQF feedback tool or other feedback mechanisms that would be directly incorporated into the CDP measure evaluation. Achieving this may require overcoming some barriers, but opportunities to improve current infrastructure incrementally do exist.

Closing the Feedback Loop

As a neutral convener, NQF generally plays the role of facilitator by passing along information between entities. In the case of feedback, this means taking submitted feedback and conveying it to the developer, but without acquiring an independent response from the receiver. The Committee noted that when measure end-users submit feedback, developers should proactively respond as soon as possible.

First, users should receive a message that confirms that NQF has received their feedback, along with specific information about what has been done with it, and how and when they might anticipate a response. In some NQF tools, there is no notification for submitters that a comment has been received and logged in the system. Next, users should get confirmation that the feedback has been provided to an entity that is capable of responding to it, either with a modification to measure specifications, with additional implementation guidance or clarification, or with an explanation of why no change was made after considering the feedback. Confirmations of feedback submissions should clarify whether the feedback has been passed on to a developer, or to a standing committee reviewing the measure. Within the CDP evaluation process, comments that were submitted in prior review cycles often are not included in the materials for the committee's consideration. The inclusion of previous comments would assist the committee in understanding if the measure developer addressed the past concerns and issues since the measure's last endorsement review.

Targeted Outreach to Key Stakeholders

The Measure Feedback Loop Committee emphasized that it is imperative to promote opportunities for stakeholders to offer feedback on performance measures. Although NQF regularly promotes opportunities to comment on performance measures as they are undergoing evaluation in the CDP, these promotions only reach those who have already subscribed to NQF's contact lists. As a result, stakeholders who are less familiar with NQF's role in performance measure evaluation, but are nonetheless being measured by endorsed standards, may not realize they have an opportunity to note unintended consequences or unexpected benefits of measures.

The Committee noted certain stakeholder groups are particularly important to reach with these opportunities. In particular, these include hospital administrators, database coordinators, or other quality managers. These individuals will have their own perspective on measures, and are well-placed to collect feedback directly from frontline providers. The Committee suggested that NQF identify other channels that measure implementers are using to communicate about measure development and approval milestones (e.g., CMS's communications about the rulemaking process) and request that NQF's endorsement milestones also be included in these communications.

NQF and measure developer organizations should establish and standardize feedback channels with specialty societies, professional organizations, vendors, and relevant stakeholder groups to facilitate use and feedback on measures. Once collected, this feedback could be used in several ways. For example, if it warrants a re-examination of the measure (e.g., a significant unintended consequence has been identified), NQF would initiate an early review for maintenance of endorsement. Alternatively, if the feedback is best addressed through a clarification by the developer, NQF could facilitate connecting the feedback provider to the developer to resolve the issue.

By soliciting feedback from targeted organizations, developers and NQF can create the potential for consistent and reliable feedback from a known source. Developers also have the opportunity to seek out professional societies with registries to identify opportunities for implementation through this source or to promote the use of the measure thorough its membership and to use the organization as a channel for collecting feedback from its membership. While many of these organizations have key personnel that are engaged in NQF's endorsement work, often their engagement is limited to their specific stakeholder interests (e.g., hospital associations interested in exclusively hospital-level measures). Others suggested there may be opportunity to partner with electronic health record (EHR) vendors or other data management vendors; there may be opportunities for collecting measure feedback through the same interfaces where data elements are entered or submitted.

Classifying Feedback

The Measure Feedback Loop Committee noted that comments and other submitted feedback were generally entered as "free text" with no opportunity to classify the feedback or pair it with other comments that may have similar themes. Although the Committee did not define a classification scheme, they did suggest some approaches, such as a distinguishing comments or feedback related to clinical concerns about the measure and implementation and burden concerns. Others suggested structuring feedback around the NQF subcriteria that assess feedback (e.g., unintended consequences, improvement) to create a better link to a Standing Committee's evaluation. Likewise, the Committee noted the importance of attributing the feedback to a particular stakeholder group or perspective, in order to ensure that feedback was drawn from a representative group of measure users.

Developer Guidance on Collecting Feedback

Guidance should be established for developers on best practices and approaches to gathering and submitting feedback. Due to the various types and varying capabilities of measure developer organizations, the resources and approaches for collecting feedback also vary significantly. In order to help developers collect more meaningful feedback in the endorsement process, the Measure Feedback Loop Committee recommended that guidance be established for developers on strategies for how to approach the collection of feedback for their measures based on varying levels of resources and access to data. The Measure Feedback Loop Committee emphasized the importance of developers soliciting feedback directly from those implementing the measure and understanding any unintended consequences that have arisen.

NEXT STEPS

In conjunction with the environmental scan, this report will inform the Measure Feedback Loop Committee's future recommendations for the options and implementation plan for a measure feedback loop pilot. In a series of upcoming webinars, the Measure Feedback Loop Committee will discuss several potential options for the design of the feedback loop pilot. The options will include the impacts of cost, time, data quality, and various other outcomes to feedback. The Committee will review and discuss the elements of each option that make it feasible and valuable, and will consider how to address potential future barriers and ensure the success of the pilot. Once a pilot design is selected, an implementation plan for the measure feedback loop pilot will be developed. See the **Measure Feedback Loop project page** for a full list of the meeting dates for this project.

REFERENCES

1 National Quality Forum (NQF). *Measure Evaluation Criteria and Guidance for Evaluating Measures for Endorsement*. Washington, DC: NQF; 2018. http://www.qualityforum.org/WorkArea/linkit.aspx?LinkIdentifier=id&ItemID=88439.

2 National Quality Forum (NQF). *Guidance For Evaluating Usability and Use of Performance Measures*. Washington, DC: NQF; 2012. http://www.qualityforum.org/Publications/2012/04/Guidance_for_Evaluating_Usability_and_Use_of_Performance_Measures.aspx.

APPENDIX A:
Operational Definitions of Key Terms

Accountability applications: Uses of performance results about identifiable, accountable entities to make judgments and decisions as a consequence of performance, such as reward, recognition, punishment, payment, or selection (e.g., public reporting, accreditation, licensure, professional certification, health information technology incentives, performance-based payment, network inclusion/exclusion). Selection is the use of performance results to make or affirm choices regarding providers of healthcare or health plans.

Measure feedback: Information received or solicited on a measure following its implementation related to performance rates, measure feasibility, or Usability and Use of the measure, including unintended consequences.

Feedback Loop: The process by which feedback on an NQF-endorsed measure is relayed to NQF multistakeholder standing committees and measure developers by those who implement measures or use measure results for decision making and improving care. Those providing feedback should also receive a response to their feedback with the expectation that measure developers adjudicated that feedback considering whether revisions to the measure are needed.

Performance rate: Measure output.

Registry: A system for keeping an official list or record of health processes or outcomes.

Unintended consequences: A set of results due to measure implementation that was not intended as an outcome.

Implementation: A specified set of activities designed to put into practice an activity or program of known dimensions.

Feasibility: Extent to which the specifications, including the measure logic, require data that are readily available or could be captured without undue burden and can be implemented for performance measurement.

Dashboard: An information management tool that visually tracks, analyzes, and displays performance indicators, metrics, and key data points.

Usability and Use: 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.

Patient-reported outcomes: Any report of the status of a patient's health condition that comes directly from the patient, without interpretation of the patient's response by a clinician or anyone else.

APPENDIX B:
NQF Measure Submission Form—Usability and Use Section

For full measure submission form, please see NQF's **Submitting Standards page**.

Usability and Use

Extent to which intended audiences (e.g., consumers, purchasers, providers, policymakers) can understand the results of the measure and are likely to find them useful for decision making.

NQF-endorsed measures are expected to be used in at least one accountability application within 3 years and publicly reported within 6 years of initial endorsement in addition to performance improvement.

4a. Use

4.1. Current and Planned Use *(check all the current and planned uses; for any current uses that are checked, provide a program name and URL for the specific program)*

Intended Use	Specific Plan for Use	Current Use	For current use, provide Program Name and URL
a. Public Reporting			
b. Public Health/Disease Surveillance			
c. Payment Program			
d. Regulatory and Accreditation Programs			
e. Professional Certification or Recognition Program			
f. Quality Improvement with Benchmarking (external benchmarking to multiple organizations)			
g. Quality Improvement (Internal to the specific organization)			
h. Not in use			
i. Use Unknown			

Accountability/Transparency (measure evaluation criterion 4a1)

4a1.1. For each CURRENT use, checked above (update for maintenance of endorsement), provide:

- Name of program and sponsor
- Purpose
- Geographic area and number and percentage of accountable entities and patients included level of measurement and setting

4a1.2. If not currently publicly reported OR used in at least one other accountability application (e.g., payment program, certification, licensing) what are the reasons? (*e.g., Do policies or actions of the developer/steward or accountable entities restrict access to performance results or impede implementation?*)

4a1.3. If not currently publicly reported OR used in at least one other accountability application, provide a credible plan for implementation within the expected timeframes—any accountability application within 3 years and publicly reported within 6 years of initial endorsement. (*Credible plan includes the specific program, purpose, intended audience, and timeline for implementing the measure within the specific timeframes. A plan for accountability applications addresses mechanisms for data aggregation and reporting.*)

Feedback on the measure by those being measured or by others (measure evaluation criterion 4a2)

4a2.1.1. Describe how performance results, data, and assistance with interpretation have been provided to those being measured or other users during development or implementation.

How many and which types of measured entities and/or others were included? If only a sample of measured entities were included, describe the full population and how the sample was selected.

4a2.1.2. Describe the process(es) involved, including when/how often results were provided, what data were provided, what educational/explanatory efforts were made, etc.

4a2.2.1. Summarize the feedback on measure performance and implementation from the measured entities and others described in 4d.1.

Describe how feedback was obtained.

4a2.2.2. Summarize the feedback obtained from those being measured.

4a2.2.3. Summarize the feedback obtained from other users.

4a2.3. Describe how the feedback described in 4a2.2 has been considered when developing or revising the measure specifications or implementation, including whether the measure was modified and why or why not

4b. Usability

Improvement (measure evaluation criterion 4b1)

4b1. Refer to data provided in 1b but do not repeat here. Discuss any progress on improvement (trends in performance results, number and percentage of people receiving high-quality healthcare; geographic area and number and percentage of accountable entities and patients included.)

If no improvement was demonstrated, what are the reasons? If not in use for performance improvement at the time of initial endorsement, provide a credible rationale that describes how the performance results could be used to further the goal of high-quality, efficient healthcare for individuals or populations.

Unexpected findings (measure evaluation criterion 4b2)

4b2.1. Please explain any unexpected findings (positive or negative) during implementation of this measure including unintended impacts on patients.

4b2.2. Please explain any unexpected benefits from implementation of this measure.

APPENDIX C:
Measure Feedback Loop Committee and NQF Staff

COMMITTEE

Rose Baez, RN, MSN, MBA, CPHQ, CPPS (Co-chair)
Blue Cross Blue Shield Association
Chicago, Illinois

Edison Machado, MD, MBA (Co-chair)
IPRO
Lake Success, New York

Constance Anderson, BSN, MBA
Northwest Kidney Centers
Seattle, Washington

Robert Centor, MD, MACP
University of Alabama at Birmingham School of Medicine
Birmingham, Alabama

Elvia Chavarria, MPH
PCPI Foundation
Chicago, Illinois

Dan Culica, MD, PhD
Health and Human Services
Austin, Texas

Melody Danko Holsomback, BSN
Keystone ACO, Geisinger
Honesdale, Pennsylvania

Anne Deutsch, RN, PhD
RTI International
Chicago, Illinois

Tricia Elliott, MBA, CPHQ
The Joint Commission
Oakbrook Terrace, Illinois

Lee Fleisher, MD
University of Pennsylvania
Philadelphia, Pennsylvania

Mark E. Huang, MD
Shirley Ryan Abilitylab
Chicago, Illinois

Joseph Kunisch, PhD, RN-BC, CPHQ
Memorial Hermann Health System
Houston, Texas

Ekta Punwani, MHA
IBM Watson Health
Chicago, Illinois

Jill Shuemaker, RN, CPHIMS
The American Board of Family Medicine
Washington, District of Columbia

Heather Smith, PT, MPH
American Physical Therapy Association
Alexandria, Virginia

Deborah Struth, MSN, RN, PhD(c)
Oncology Nursing Society
Pittsburgh, Pennsylvania

Claire Noel-Miller, MPA, PhD
AARP
Washington, District of Columbia

Sue Sheridan, MIM, MBA
Society to Improve Diagnosis in Medicine (SIDM)
Washington, District of Columbia

Koryn Rubin, MHA
American Medical Association
Washington, District of Columbia

Elizabeth Rubinstein
Henry Ford Health System
Detroit, Michigan

Sara Toomey, MD, MPhil, MPH, MSc
Boston Children's Hospital
Boston, Massachusetts

NQF STAFF

Elisa Munthali, MPH
Senior Vice President, Quality Measurement

Ashlie Wilbon, MS, MPH, FNP-C
Senior Director

Jean-Luc Tilly, MPA
Senior Manager, Data Analytics

Madison Jung
Project Manager

Navya Kumar, MPH
Project Analyst

APPENDIX D:
Public Comments

Collette Pitzen

Minnesota Community Measurement

MN Community Measurement appreciates the recognition that the measure developer does not always possess information about where a measure that they steward is being used and/or publicly reported. Any efforts to gain information by the various feedback loops proposed is necessary and appreciated. Additionally, we support the committee's comments in Recommendation #1 that there may be certain allowable justifications for a maintenance measure that is not included in an accountability program within 3 years of endorsement and publicly reported after six years. It is our experience that measures do change based on changes in evidence and measures move into and out of programs for various reasons. Thank you for the opportunity to comment!

ISBN 978-1-68248-121-9

NATIONAL QUALITY FORUM
1030 15TH STREET, NW, SUITE 800
WASHINGTON, DC 20005

http://www.qualityforum.org