# NATIONAL QUALITY FORUM

TO: NQF Members

FR: Karen Pace, PhD, MSN

RE: Review of *Guidance for Evaluating Usability and Use of Performance Measures*

DA: October 3, 2011

**Background**

The National Quality Forum's (NQF) mission to improve the quality of American healthcare includes endorsing national consensus standards for measuring and publicly reporting on performance. Performance measures considered for endorsement are evaluated against four major criteria: Importance to Measure and Report, Scientific Acceptability of Measure Properties, Usability, and Feasibility. Historically, NQF's work has revolved around endorsing performance measures useful both for quality improvement and accountability, with an emphasis on transparency and public reporting. With the passage of health reform legislation and today's quality environment, the needs of the Department of Health and Human Services (HHS) and other stakeholders are such that the NQF portfolio will need to be broad enough to support additional core accountability applications, including: value-based payment, health IT incentive payments, accreditation, and regulation.

The Usability Task Force was charged with the following tasks.

- review and refine the NQF Usability criterion and subcriteria;
- develop operational guidance related to the measure evaluation criteria for Usability;
- identify the types of information measure stewards will be expected to submit to NQF at the time of endorsement and maintenance to demonstrate usability; and
- discuss whether measure developer recommendations for reporting performance results (e.g., classification methods used for public reporting and other accountability applications) should be reviewed in the measure evaluation process.

The Task Force's recommendations are included in the draft document, *Guidance for Evaluating Usability and Use of Performance Measures*. The draft report is posted on the NQF web site for review and comment only—not voting.

You may post your comments and view the comments of others on the NQF website.

**NQF Member comments must be submitted no later than 6:00 PM ET, October 24, 2011; public comments are due 6:00 PM ET, October 17, 2011.**

NQF is now using a program that facilitates electronic submission of comments on this draft report. **All comments must be submitted using the online submission process.**

Supporting documents related to your comments may be submitted by e-mail to kpace@qualityforum.org with "Usability Report" in the subject line and your contact information in the body of the e-mail.

Thank you for your interest in the NQF's work. We look forward to your review and comments.

# NATIONAL QUALITY FORUM

## GUIDANCE FOR EVALUATING

## USABILITY AND USE OF PERFORMANCE MEASURES

**Draft Report for Review and Comment**
**10/03/11**

# Guidance for Evaluating Usability and Use of Performance Measures

## CONTENTS

## INTRODUCTION

The National Quality Forum's (NQF) mission to improve the quality of American healthcare includes endorsing national consensus standards for measuring and publicly reporting on performance. Performance measures considered for endorsement are evaluated against four major criteria: Importance to Measure and Report, Scientific Acceptability of Measure Properties, Usability, and Feasibility.

The Usability criterion originally was intended to determine whether users of a measure–consumers, purchasers, providers, and policymakers—would be able to understand the performance results and find them useful for decision making related to accountability and improvement. At the Consensus Standards Approval Committee's March 2011 retreat, the committee discussed at length the need to adapt the Usability criterion to capture the full range of accountability uses for endorsed measures (e.g., selection, value-based payment, accreditation, health IT incentive programs). The committee also expressed interest in further delineating NQF expectations of measure owners to demonstrate that their measure is being used and the results are useful, either at the time of initial endorsement or by the time of endorsement maintenance.

### Task Force Charge

The Usability Task Force was charged with the following tasks.

- review and refine the NQF Usability criterion and subcriteria;
- develop operational guidance related to the measure evaluation criteria for Usability;
- identify the types of information measure stewards will be expected to submit to NQF at the time of endorsement and maintenance to demonstrate usability; and
- discuss whether measure developer recommendations for reporting performance results (e.g., classification methods used for public reporting and other accountability applications) should be reviewed in the measure evaluation process.

## BACKGROUND

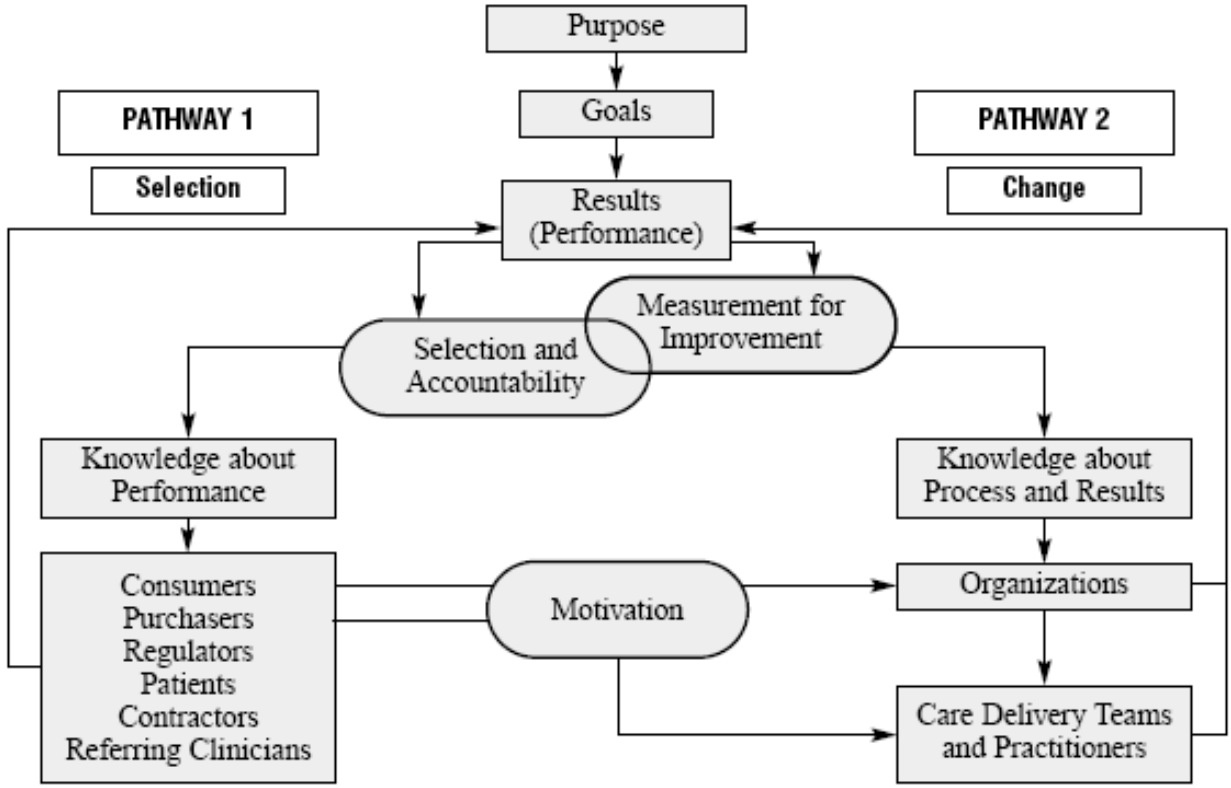
Historically, NQF's work has revolved around endorsing performance measures useful both for quality improvement and accountability, with an emphasis on transparency and public reporting. In October 2009, the NQF Board of Directors affirmed there is a general expectation that performance results from NQF-endorsed© measures will be used in public reporting programs, thus providing transparency and supporting the broadest set of applications, and that NQF should assess the "actual use and usefulness" of endorsed measures at the time of the three-year maintenance review.

Public reporting continues to be of great interest and until recently, the primary focus for accountability. However, with the passage of health reform legislation and today's quality environment, the needs of the Department of Health and Human Services (HHS) and other

stakeholders are such that the NQF portfolio will need to be broad enough to support additional core accountability applications, including: value-based payment, health IT incentive payments, accreditation, and regulation. The additional accountability applications complement selection of healthcare providers through public reporting with identification of healthcare entities for specific rewards or penalties. The goal is to align incentives to encourage and reward the provision of high-quality and efficient healthcare. What expectations should be regarding actual use in these various programs at the time of the three-year review or how to evaluate usefulness to the decision makers and ultimately for improvement is less clear.

Figure 1 illustrates the foundational concepts for using measurement to facilitate the goal of patients receiving high-quality, efficient healthcare through selection and accountability (which requires access to performance results by consumers, purchasers, and others) and changes in care leading to improvement.[1, 2] The term *accountability* will be used throughout this report because it is the broader term and encompasses selection. The term *performance improvement* will be used throughout this report to denote the change pathway that leads to improvement.

**Figure 1. Pathways from Measurement to Improvement**[1]




As interest in using measures for different applications has intensified, and the number of measures in the NQF portfolio has grown, it has become apparent that selecting measures for use

in a specific application is a complicated undertaking. In response to provisions in the Affordable Care Act, NQF initiated the Measure Applications Partnership (MAP) in 2011 for the explicit purpose of providing input to HHS and private-sector leaders on selecting performance measures for various accountability programs. Selection for various applications will build on the foundation of NQF endorsement.

## Issues Related to Evaluating Usability

Several issues have challenged the evaluation of the current Usability criterion.

- Measure developers have sometimes struggled to perform basic testing of reliability and validity and have reported they did not have resources to test the understanding and usefulness of performance results for various accountability applications or quality improvement.
- Implementing measures may be accomplished by some other entity than the developer/steward submitting the measure for endorsement, and the implementers would be in the best position to demonstrate usability.
- At the time of endorsement maintenance review, some endorsed measures are not in use or there is little or no information about use. Again, the measure developer/steward often has no mechanism or authority to initiate use of a measure.
- Specifications for NQF-endorsed measures need to be made publicly available. However, the data needed to implement a measure often is owned or collected by other entities. Occasionally, the developer/steward of the measure also has sole control of the data needed to compute and report on a performance measure, and some question whether NQF should endorse such measures without a commitment and plan for public reporting.
- There is some sentiment that endorsement should not be continued for measures that are not in use; however, there also is concern that good measures not be lost only because they are not yet implemented. There is little experience as yet regarding how long it takes to achieve use.
- Although transparency of performance information is critical to support accountability and selection, there are various degrees of transparency as well as a variety of accountability functions that can help drive improvement without performance scores necessarily being publicly reported.
- More accountability functions (e.g., payment, accreditation, professional certification) are dependent on performance measures but are not explicitly addressed in the current criteria.
- There is some concern that failing to require that all NQF-endorsed measures be reported on public websites will result in very limited information being available to support selection (i.e., the “slippery slope”).
- There also is concern that too much emphasis on public reporting will result in overload of patients/consumers and that the focus should be on providing the right information to

consumers that will help them make key decisions about their choice of providers, treatment options, etc.

Additionally, several other issues were identified as potentially applicable to the Usability criterion.

- NQF endorsement is for a specified measure and has not included how the measure results are classified or reported (e.g., using stars to indicate ranking, stating whether results are above or below average, using confidence intervals). Occasionally, some measures have been submitted with methods for classifying the results, but NQF has required those methods to be separate from the endorsed measure. The rationale for this position is twofold: 1) a measure may be used in more than one application, and classifying and formatting results should be tailored to the specific application (e.g., a "star" presentation may be most understandable for patients, while a numeric score may be preferred for payment applications; and 2) the NQF endorsement criteria, process, and committee appointments have not been designed to determine the best reporting approach. Given this position, the issue has been raised as to whether NQF should include "reporting guidance" along with an endorsed measure.
- The criterion for unintended consequences (4c) has been under the Feasibility criterion, but it has been suggested that it might be more appropriate under Usability.
- Currently disparities are addressed under performance gap (1b) and measure specifications to detect disparities (2c), but it has been suggested that disparities might be considered under Usability.

## RECOMMENDATIONS

The Task Force identified some definitions and principles that guided its discussion and the recommendations that follow.

### Definitions

Terms such as *accountability* and *public reporting* have been used inconsistently. Therefore, the task force recommended definitions to facilitate standard terminology and understanding.

**Accountability:** An obligation or willingness to accept responsibility for performance.

**Accountability Applications:** Use of performance results for 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, health IT incentives, performance-based payment, network inclusion/exclusion).

**Public Reporting:** Making comparative performance results for identifiable, accountable entities freely available (or at nominal cost) to the public at large (generally on a public website).

**Selection:** Use of performance results to make or affirm choices regarding providers of healthcare or health plans (e.g., an individual needing surgery to choose a surgeon; an employer to choose which health plan to offer; a health plan to choose which specialists to empanel; a family doctor to choose to which oncologist to refer a cancer patient; an employee or Medicaid enrollee choosing a health plan during open enrollment).

**Transparency:** Extent to which performance results for identifiable, accountable entities are *disclosed and available* outside of the organizations or practices whose performance is measured. There are degrees of transparency ranging from making performance results available only to a few selected staff within an organization to reporting the results to the public at large. The capability to verify the performance results adds significantly to measure transparency. *The following descriptions of disclosure and availability of performance results are listed in Table 1 from least to most transparency.*

### Table 1. Degrees of Transparency

| | |
|---|---|
| Not Transparent | • Performance results are *neither* disclosed nor available outside the organization or practice whose performance is being measured:<br>o available only to selected staff (e.g., quality department)<br>o shared only within the organization or practice<br>o reported confidentially to a third party for benchmarking |
| | • Performance data or results are reported to a third party for some accountability application, but generally are not publicly available (e.g., to an insurance plan to maintain preferred provider status or payment incentives)<br>• Performance results are self-reported on the organization's own website without comparative information<br>• Performance results reported and comparative performance results for identifiable, accountable entities available:<br>o only to members of a defined group (e.g., members of a health plan)<br>o to anyone for a cost<br>• Public reporting - performance results reported and comparative performance results for identifiable, accountable entities available freely (or at nominal cost) to anyone |
| Most Transparent | • Performance results are reported and comparative performance results for identifiable, accountable entities available freely (or at nominal cost) to anyone (i.e., public reporting) with the additional availability of Health Insurance Portability and Accountability Act (HIPAA)-compliant patient-level data for verification and analysis or reanalysis |

**Usable:** Capable of being used by intended audiences; convenient and practicable for use.

**Useful:** Capable of being put to use and serviceable for an end or purpose.

## Principles

The following principles provide a foundation for a criterion regarding usability of NQF-endorsed measures.

- Performance measurement facilitates the goal of high-quality, efficient healthcare for all Americans through two pathways: 1) improvement through changes in care initiated by healthcare providers and 2) accountability/selection by making information available to consumers, referring clinicians, and others involved in decisions about the selection of clinicians and providers. Accountability and selection aim to create an environment that enables and rewards improvement through aligning payment, public reporting, and quality oversight programs.
- NQF strives to endorse measures that are useful for *both* accountability and improvement to maximize the influence on the goal of high-quality, efficient healthcare for all Americans.
- To achieve maximal effect on quality healthcare and health, over time, NQF-endorsed measures should be used in all applications for which they provide useful information.
- Public disclosure of performance results not only is necessary for some types of selection such as consumer choice, but it also ensures accountability and provides external motivation for performance improvement. NQF encourages transparency of performance results.
- Measure developers may not be the implementers of performance measures for accountability/selection or quality improvement programs and also may not have access to the required data or information about use of the measure. The NQF-Quality Alliance Steering Committee (QASC) report encourages collaboration between developers and potential implementers of performance measures; otherwise, resources and efforts for developing and testing measures could be wasted if measures are not implemented.
- The NQF criteria of *Importance to Measure and Report* (i.e., high impact, opportunity for improvement, and evidence) and *Scientific Acceptability of Measure Properties* (i.e., reliability and validity) 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., if reporting methods or classification methods obscure differences in performance).
- The NQF criterion of Feasibility, particularly regarding the data required to implement a performance measure, also influences usability. However, feasibility issues may be mitigated or the benefit of measuring performance may be seen to outweigh associated burden.

## I. Recommendations for Measure Evaluation Criteria for Usability

A central question the task force discussed is whether measures ever fail to be endorsed *only* because of failure to meet the Usability criterion. If no measures fail on this criterion, then it may

not be a criterion. In other words, won't all measures that are determined to meet Importance to Measure and Report and Scientific Acceptability of Measure Properties be usable to some audience? To date, measures have not failed to be endorsed based *solely* on the Usability criterion. Questions that have arisen under Usability, such as usefulness of a measure for performance improvement, often relate to other criteria, such as validity or evidence. Some measures could potentially fail on the current Usability criterion because understandability or interpretability was not demonstrated, but usually steering committees have not seen that as a fatal flaw. Instead, it has been viewed as a correctable problem that can be addressed through the language used to explain the measure, which can be tested using cognitive interviews and focus groups.

Some measures could potentially fail endorsement because they are not in use at the time of endorsement maintenance. When a measure is not in use, sometimes it is because of problems related to other criteria, such as opportunity for improvement, evidence, reliability, validity, or unintended consequences. Sometimes a measure may not be in use because the measure steward also controls and limits access to performance results or the underlying data. In some cases, however, implementation depends on external factors beyond the steward's or developer's control (e.g., a measure is specified for electronic health records (EHRs), but EHRs are not yet widely adopted). The task force noted that although measure developers may not be the implementers, if they have not been engaged with potential implementers from the onset of measure development and testing, resources for measure development, testing, and endorsement may be wasted when measures are not put into use.

The task force identified that the concepts of usability and usefulness are related to a specific purpose. The general purpose of measures endorsed by NQF is to facilitate high-quality, efficient healthcare for all Americans. In general, measures that meet the NQF criteria for Importance to Measure and Report and Scientific Acceptability of Measure Properties should be theoretically usable for both accountability and performance improvement. The task force did not recommend different criteria for specific applications (e.g., payment incentives vs. public reporting). MAP will address the selection of specific endorsed measures for specific programs.

Understanding and interpretability are related to a specific audience and conditions of implementation (e.g., language used and how the results are displayed), and the task force agreed these factors should not be included under the Usability criterion. Several other NQF projects addressed guidance on reporting, which should be followed to help improve understanding and thus usability for key audiences (Reporting Results to the Public in Comprehensive Framework for Hospital Care Performance Evaluation[3], Guidelines for Consumer-Focused Public Reporting[4], and Public Reporting of Patient Safety Event Information[5]. The recommendations from these projects are provided in the appendix (Table A-2) and include:

- tailor reporting to the intended audience and specific purpose;

- use a transparent process and include input from the intended audience;
- provide contextual information;
- use consistent, simple, and familiar language;
- present and explain data clearly and objectively in ways that facilitate interpretation;
- identify and use effective designs and format; and
- regularly reassess and obtain feedback.

The earlier guidance suggests grouping information into categories such as "better" or "average" but did not address the methodological issues involved in making the categories, such as statistical analyses, or how to convey certainty around performance scores.

Ultimately, the goal for NQF-endorsed measures is to facilitate high-quality, efficient healthcare, which requires that the endorsed measures be in **use** both internally for improvement and externally for accountability. Therefore, the task force recommended that the Usability criterion be modified to include the concepts of use and progress toward achieving the goal of high-quality, efficient healthcare as presented in Table 2. Because the extent of use for internal quality improvement would be difficult to quantify, actual use is focused on accountability applications. Although transparency through public reporting is preferred, other accountability applications are recognized.

The task force also discussed a few additional concepts for potential consideration under the Usability criterion: unintended consequences, disparities, and methods for classifying performance.

- It agreed that unintended negative consequences should be considered under Usability along with the evidence of use and influence on quality. The task force clarified that the negative consequences should be to individuals or populations. (Issues regarding fair comparisons among the entities whose performance is being measured would relate to the measure's validity.) The task force noted it would not be feasible to request evidence that no adverse consequences occurred; however, the potential for unintended negative consequences should be considered in measure development, and this type of information should be solicited from users of endorsed measures. Reports of negative consequences should be accompanied with data specifying the nature of the consequence, the affected party, the number of people affected, and the severity of the impact.
- The task force did not think disparities should be addressed under Usability and agreed it be addressed early in the evaluation criteria. Currently assessment of disparities in care is part of performance gap (1b) under the first threshold criterion of Importance to Measure and Report and specifications to detect disparities (2c). A concurrent project on disparities will be making some recommendations about identifying disparities-sensitive measures.
- Finally, the task force agreed that guidance on reporting performance results such as methods for classifying performance results (e.g., stars) should not be considered under Usability

because it is not a core part of the measure construction and depends on context. NQF measure endorsement should focus on the performance measure rather than on methods of reporting. However, the task force noted that the way in which performance results are reported can affect understanding or even the validity of the conclusions made. It agreed it is important topic that NQF should identify the pros and cons of NQF involvement in reporting guidance. NQF's role could range from guidance, as in the earlier projects, to identifying additional principles for reporting based on the growing body of evidence about reporting on performance, to evaluating specific reporting guidance for each measure based on a set of criteria.

## Table 2. Evaluation Criteria for Usability and Use

| Current Criteria | DRAFT Modifications |
|---|---|
| Condition for Consideration<br>C. The intended use of the measure includes both public reporting and quality improvement. | Condition for Consideration<br>C. The intended use of the measure includes both accountability applications[1] and performance improvement to achieve high-quality, efficient healthcare. |
| 3. Usability<br>Extent to which intended audiences (e.g., consumers, purchasers, providers, policymakers) can understand the results of the measure and find them useful for decisionmaking.<br><br>3a. Demonstration that information produced by the measure is meaningful, understandable, and useful to the intended audiences for public reporting (e.g., focus group, cognitive testing) or rationale;<br><br>AND<br><br>3b. Demonstration that information produced by the measure is meaningful, understandable, and useful to the intended audiences for informing quality improvement[16] (e.g., quality improvement initiatives) or rationale. | 3. Usability and Use<br>Extent to which potential audiences (e.g., consumers, purchasers, providers, policymakers) are using or could use performance results for both accountability and performance improvement[2] to achieve the goal of high-quality, efficient healthcare for individuals or populations.<br><br>3a. A plausible rationale describes how the performance results could be used to achieve the goal of high-quality, efficient healthcare through both accountability and performance improvement;<br><br>over time, progress on achieving the goal of high-quality, efficient healthcare[3] for individuals or populations is demonstrated.<br><br>AND<br><br>3b. A credible plan[4] and commitments for implementation of the performance measure for at least one accountability application and path to transparency[5] of performance results represents the opportunity to influence the goal of high-quality, efficient healthcare;<br><br>over time, use in accountability applications with progress toward transparency is demonstrated.<br><br>AND<br><br>3c. The benefits of the performance measure to achieving high-quality, efficient healthcare for individuals or populations outweigh potential unintended negative consequences to individuals or populations. |
| Note<br>16. An important outcome that may not have an identified improvement strategy still can be useful for informing quality improvement by identifying the need for and stimulating new approaches to improvement. | Criteria Notes<br>1. Accountability applications are the use of performance results for 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, health IT 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.<br>2. An important outcome that may not have an identified improvement strategy still can be useful for informing quality improvement by identifying the need for and stimulating new approaches to improvement.<br>3. Examples of demonstrated progress on quality include evidence of improved performance and/or increased numbers of individuals receiving high-quality healthcare.<br>4. Credible plans should identify the specific entities involved and address mechanisms for data aggregation; intended audience; specific program and purpose; reporting methods; responsibilities; timeline; and pathway to achieve public reporting.<br>5. Transparency is the extent to which performance results for identifiable, accountable |

| Current Criteria | DRAFT Modifications |
|---|---|
| | entities are *disclosed and available* outside of the organizations or practices whose performance is measured. There are degrees of transparency ranging from making performance results available only to a few selected staff within an organization to reporting the results to the public at large. The capability to verify the performance results adds significantly to measure transparency. Public reporting is making comparative performance results for identifiable, accountable entities freely available (or at nominal cost) to the public at large (generally on a public website). |

## II. Recommendations for Evaluating Usability and Use

The goal of NQF-endorsed measures is to facilitate high-quality, efficient healthcare through their widespread adoption and use for accountability and performance improvement. Therefore, resources for measure development and endorsement should be focused on measures that are being used or will be. Usability and Use should be evaluated after the other three major criteria—Importance to Measure and Report, Scientific Acceptability of Measure Properties, and Feasibility. If the other criteria are met (i.e., the measure addresses a high-impact aspect of healthcare, with a performance gap and evidence base and is reliable, valid, and feasible), then a measure is almost certain to be potentially usable.

Usability is a hypothetical characteristic of a measure that can be evaluated at the time of initial endorsement. At the time of endorsement maintenance, attention should be turned to observed use of a measure and progress toward achieving high-quality healthcare. If a measure is already in use at the time of initial endorsement, use and achievement of high-quality healthcare could be evaluated at that time. In addition to the information submitted by the measure developer, implementation comments from the field will help identify use or reasons for lack of use or lack of progress in achieving high-quality healthcare. On evaluation for endorsement maintenance, lack of use or lack of progress in achieving high-quality healthcare may be a signal for problems related to the other criteria, which should be re-examined if indicated. For example:

- Is there little opportunity for improvement (criterion 1b)?
- Has the evidence changed and no longer supports the focus of measurement (criterion 1c)?
- Does the evidence link the measured process or structure to desired outcomes (criterion 1c)?
- Are there problems with reliability (criterion 2a) or validity (criterion 2b)?
- Are there issues with feasibility (criterion 3), such as delayed adoption of, or capture of required data in, electronic records, data collection burden, or privacy concerns?

If the other criteria are clearly met, then what other reasons explain lack of use? Do the reasons indicate a justification to retain endorsement or to remove endorsement?

- Do unintended negative consequences to individuals outweigh the benefit?

- To what extent is the measure steward or those entities being measured responsible for lack of access to performance results or the data needed for others to implement the measure?
- Are there other external factors delaying the measure's implementation (e.g., competing priorities, funding, legislative mandates)?

Although measures need to be used to influence quality, the task force agreed that setting a specific deadline by which measures must be in use to retain endorsement was not possible at this time. The amount of time required to implement a measure depends on factors such as whether data are already being collected and if there are established systems for aggregating, analyzing, and reporting data. Additional time may be needed to pilot test data presentations. Other external factors also may slow implementation, such as funding or competing priorities. Some measures may be ahead of their time, for example, if a measure is specified for electronic records but there is slow adoption of electronic records. However, the task force agreed that if the reason performance results are not in use for some accountability application is primarily due to actions or policies of the measure developer, steward, or owners of the data, then continued NQF endorsement probably is not warranted. Assessment of Usability and Use will require the judgment of NQF multi-stakeholder steering committees. The amount of time needed to demonstrate improvement also is difficult to predict at this time and may vary by topic and type of measure.

At this time, it also may not be appropriate to set public reporting as an absolute requirement for continued endorsement. There could be some measures that are not useful for public reporting, but are useful for other accountability applications and contribute to achieving high-quality, efficient healthcare. However, perspectives often vary on whether measures are useful for public reporting; for example, providers and consumers may have different opinions on whether measures are considered too technical or too complicated for consumers to understand. Additionally, as stated in the principles, public reporting serves other purposes beyond consumer choice such as ensuring accountability and providing external motivation for performance improvement. Therefore, statements that endorsed measures are not useful for public reporting should be based on some data or testing that demonstrates a measure is not useful or could not be made useful through translation of technical terms or appropriate framing, with information on how to interpret and use the data. With more experience over time, the need for implementation timeframes and deadlines or a definitive requirement for public reporting should be reassessed.

This guidance for evaluating Usability and Use is consistent with the recent guidance for evaluating competing measures—that is, competing measures would be compared on all the criteria and subcriteria, including Usability and Use. If measures were considered equal on Importance to Measure and Report, Scientific Acceptability of Measure Properties, and Feasibility, a measure in use or with credible commitments for implementation would be

considered superior to one that is not in use. If there are differences between competing measures on the criteria and subcriteria, steering committees need to weigh the strengths and weaknesses across all the criteria. If a competing measure does not have clear superiority, then steering committees need to assess justification for multiple measures.

Table 3 presents guidance for evaluating the criterion of Usability. Most measures that have passed the preceding criteria would be expected to pass this criterion.

## Table 3. Guidance for Evaluating Usability and Use

| Criterion | Indications that Criteria are Met and Questions for Further Evaluation |
|---|---|
| In Use for Accountability | • For new measures on initial endorsement that are not currently being used in an accountability application: at a minimum, there is a credible plan and commitment to use the performance measure in at least one accountability application and path to transparency. Credible plans should identify the specific entities involved and address mechanisms for data collection, data aggregation; intended audience; specific program and purpose; reporting methods; responsibilities; timeline; and path to achieve public reporting.<br>• Over time, use in accountability applications with progress toward transparency is demonstrated by providing specific accountability program information including: name, purpose, geographic area, link to web page, numbers of entities, and patients included.<br>o If measure is not in use, why not?<br>Review the submission and implementation comments:<br>▪ Are there issues with other threshold criteria that should be re-examined: performance gap, evidence, reliability, validity?<br>▪ Is feasibility the issue, and if so, can it be resolved?<br>▪ Are there other external factors delaying the measure's implementation (e.g., competing priorities, funding, legislative mandates)?<br>▪ Do unintended negative consequences to patients outweigh the benefit (see below)?<br>▪ To what extent is the measure developer, steward, or accountable entities being measured responsible for lack of access to performance results or the data needed for others to implement the measure (e.g., restrictions or limitations on release of performance results or use of data, relies on voluntary participation)?<br>o If not publicly reported, a credible plan is provided (including the plan elements identified above), or testing or other data indicates the measure is not useful for public reporting.<br>▪ Is there ample justification for not publicly reporting the performance results?<br>• Do the reasons for nonuse indicate a justification to retain endorsement or to remove endorsement? |
| Benefits vs. Unintended Negative Consequences | • For new measures on initial endorsement at a minimum, potential unintended negative consequences to individuals or populations related to the performance measure are identified (if any), and a plausible rationale explains why they are inconsequential or how they can be or were minimized.<br>• Over time with use, benefits of the performance measure to achieving high-quality healthcare to individuals or populations outweigh unintended negative consequences to individuals or populations (e.g., data on improvement outweigh reported/ substantiated negative consequences to individuals or populations). |
| Achievement of High-Quality Healthcare | • For new measures on initial endorsement at a minimum, a plausible rationale describes how the performance results could be or were used to achieve high-quality healthcare for individuals or populations through both accountability and performance improvement.<br>• Over time, progress on achieving the goal of high-quality healthcare for individuals or populations is demonstrated (e.g., evidence of improved performance and/or increased numbers of individuals directed/selecting/receiving high-quality healthcare)<br>o If progress on quality is not demonstrated, why not?<br>Review the submission and implementation comments:<br>▪ Is the measure in use?<br>▪ Has there been enough time since implementation to have an effect on quality?<br>▪ Are there issues with other threshold criteria that should be re-examined, such as performance gap, evidence, reliability, validity?<br>▪ Is feasibility the issue, and if so, can it be resolved? |

| Criterion | Indications that Criteria are Met and Questions for Further Evaluation |
|---|---|
| | • Do the reasons for lack of progress on achieving high-quality healthcare indicate a justification to retain or remove endorsement? |

## III. Recommendations for Measure Submission Items for Usability and Use

The information requested on the measure submission form needs to be modified as indicated in Table 4 to be consistent with the changes to the criteria.

### Table 4. Measure Submission Items

| DRAFT Modified Criteria | Proposed Measure Submission Items to Evaluate the Criteria |
|---|---|
| Condition for Consideration<br>C. The intended use of the measure includes both accountability applications[1] and performance improvement to achieve high-quality healthcare. | C.1. Intended Purpose/Use (*Check all the purposes or uses for which the measure is intended—must include at least one accountability application (AA) and at least one performance improvement (PI) purpose*):*<br>• Public Reporting (AA)<br>• Public Health/Disease Surveillance<br>• Payment Program (AA)<br>• Regulatory and Accreditation Programs (AA)<br>• Professional Certification or Recognition Program (AA)<br>• Quality Improvement with Benchmarking (external benchmarking to multiple organizations) (PI)<br>• Quality Improvement (internal to the specific organization) (PI)<br>• Other |
| 3. Usability and Use<br>Extent to which potential audiences (e.g., consumers, purchasers, providers, policymakers) are using or could use performance results for both accountability and performance improvement[2] to achieve the goal of high-quality, efficient healthcare for individuals or populations.<br><br>3a. A plausible rationale describes how the performance results could be used to achieve the goal of high-quality healthcare through | *3.1. Current Use *(check all the current uses; for any that are checked, provide a URL for the specific program)*<br>• Public Reporting _____<br>• Public Health/Disease Surveillance_____<br>• Payment Program _____<br>• Regulatory and Accreditation Programs _____<br>• Professional Certification or Recognition Program _____<br>• Quality Improvement with Benchmarking (external<br>• benchmarking to multiple organizations) _____<br>• Quality Improvement (Internal to the specific organization) _____<br>• Other _____<br>• Not in use<br>• Use unknown<br><br>*3.2 If not currently in use or if use is unknown, explain the reasons.<br><br>*3.3 Are there any policies or actions of the developer/steward or accountable entities that restrict access to performance results? (*Some examples include but are not limited to agreement for confidentiality, requires individual voluntary agreements to release performance results, data not available to others to implement the measure* )<br><br>3a1. Provide a rationale that describes how the performance results could be used to achieve the goal of high-quality, efficient healthcare through both |

| DRAFT Modified Criteria | Proposed Measure Submission Items to Evaluate the Criteria |
|---|---|
| both accountability and performance improvement;<br><br>over time, progress on achieving the goal of high-quality, efficient healthcare[3] for individuals or populations is demonstrated.<br><br>AND | accountability and improvement.<br><br>*3a2. Provide data that demonstrate progress on achieving the goal of high-quality, efficient healthcare for individuals or populations. (*Not required for initial endorsement unless available*)<br>• Source of data<br>• Geographic area and number and percentage of accountable entities and patients included<br>• Progress (trend in performance, number and percentage of people receiving high-quality healthcare) |
| 3b. A credible plan[4] and commitments for implementation of the performance measure for at least one accountability application and path to transparency[5] of performance results represents the opportunity to influence the goal of high-quality healthcare;<br><br>over time, use in accountability applications with progress toward transparency is demonstrated.<br><br>AND | 3b1. If not in use, provide a plan and commitments for implementing the measure for at least one accountability application and path to transparency of performance results. (*The plan should include specific entities involved and address mechanisms for data collection, data aggregation; intended audience; specific program and purpose; reporting methods; responsibilities; timeline; and path to achieve public reporting.*)<br><br>*3b2. Provide specific information on use of the measure for each accountability application identified in 3.1. (*Not required for initial endorsement unless already in use*)<br>• Name of program and sponsor (URL provided in 3.1)<br>• Purpose<br>• Geographic area and number and percentage of accountable entities and patients included |
| 3c. The benefits of the performance measure to achieving high-quality, efficient healthcare for individuals or populations outweigh potential unintended negative consequences to individuals or populations. | 3c1. Identify potential unintended negative consequences to individuals or populations related to the performance measure (if any) and explain why they are inconsequential or how they can be minimized.<br><br>*3c2. Identify if any unintended negative consequences to individuals or populations have been reported or substantiated and describe how benefits outweigh the negative unintended consequences.<br><br>** Input from stakeholders on these items should be solicited on measures undergoing endorsement maintenance review.* |

## NOTES

1. National Quality Forum (NQF), *A National Framework for Healthcare Quality Measurement and Reporting*, Washington, DC: NQF, 2002.

2. Berwick DM, James B, Coye MJ, Connections between quality measurement and improvement, *Med Care*, 2003;41(1 Suppl):I30-I38.

3. NQF, *A Comprehensive Framework for Hospital Care Performance Evaluation: A Consensus Report*, Washington, DC: NQF, 2003.

4. NQF, *National Voluntary Consensus Standards for Hospital Care 2007Guidelines for Consumer-Focused Reporting*, Washington, DC: NQF, 2007.

5. NQF, *National Voluntary Consensus Standards for Public Reporting of Patient Safety Event Information*, Washington, DC: NQF, 2010.

## APPENDIX A—SUPPLEMENTAL INFORMATION

### Table A-1. Current Measure Submission Items

| |
|---|
| C.1. Purpose/Use (Check all the purposes and/or uses for which the measure is intended – must include public reporting and at least one quality improvement purpose):*<br>Public Reporting<br>Public Health/Disease Surveillance<br>Payment Program<br>Regulatory and Accreditation Programs<br>Professional Certification or Recognition Program<br>Quality Improvement with Benchmarking (external benchmarking to multiple organizations)<br>Quality Improvement (Internal to the specific organization) |
| 3.1. Current Use *(check all that apply; for any that are checked, provide the specific program information in the following questions)*<br>Public Reporting<br>Public Health/Disease Surveillance<br>Payment Program<br>Regulatory and Accreditation Programs<br>Professional Certification or Recognition Program<br>Quality Improvement with Benchmarking (external benchmarking to multiple organizations)<br>Quality Improvement (Internal to the specific organization)<br>Not in use<br>Use unknown<br><br>3a.1. Use in Public Reporting -disclosure of performance results to the public at large *(If used in a public reporting program, provide name of program(s), locations, Web page URL(s). If not publicly reported in a national or community program, state the reason and plans to achieve public reporting, potential reporting programs or commitments, and timeline, e.g., within 3 years of endorsement)*<br><br>3a.2. Provide a rationale for why the measure performance results are meaningful, understandable, and useful for public reporting. *If usefulness was demonstrated (e.g., focus, group, cognitive testing) describe the data, method and results.*<br><br>3b.1. Use in QI *(If used in quality improvement program, provide name of program(s), locations, Web page URL (s))*<br><br>3.2. Use for other Accountability Functions *(payment, certification, accreditation) (If used in a public accountability program, provide name of program(s), locations, Web page URL(s))*<br><br>4c.1. Identify susceptibility to inaccuracies, errors, or unintended consequences of measurement identified during testing and/or operational use and strategies to prevent, minimize, or detect. If audited, provide results. |

## Table A-2. Previous NQF Guidance on Reporting Performance

| A Comprehensive Framework for Hospital Care Performance Evaluation (2003) | National Voluntary Consensus Standards for Consumer-Focused Reporting (2007) | National Voluntary Consensus Standards for Public Reporting of Patient Safety Event Information (2010) |
|---|---|---|
| 5a. Source and use of reports.<br>ii. Hospital performance reports must appeal to and take into account the needs of each of the following unique stakeholder audiences:<br>a) public/consumers;<br>b) purchasers;<br>c) clinicians and providers;<br>d) policymakers; and<br>e) accreditors/regulators.<br>Different audiences may require different formats and levels of detail. All audiences should always be able to access public reports prepared for other audiences. | 1. Identify the purpose of the web-based report, its intended main consumer audience(s), and how the report will be made known to the audience; also identify secondary audiences and how their unique needs will be addressed.<br>1a. Identify the nature and purpose of the report (what it will be about and what is to be accomplished by producing it).<br><br>1b. Identify the main consumer audiences for the report and describe their characteristics, their knowledge about the subject matter of the report, their information interests and needs, and how they will be expected to learn about and use the web-based report. (In planning for use, provide for layering of information that permits the user to drill down to the technical details.)<br><br>1c. Identify secondary audiences for the report, such as healthcare providers and policymakers, and describe how their report-specific interests and needs differ from those of the main consumer audiences. Determine how the report will accommodate the secondary audiences (such as allowing users to drill down to the technical details about measurement and statistical comparisons). | 1. Identify the purpose of the report, its intended main consumer audience(s), and how the report will be made known to the audience; also identify secondary audiences and how their unique needs will be addressed<br>1a. Identify the nature and purpose of the report (what it will be about and what is to be accomplished by producing it). Whenever possible, the purpose should include accountability, learning, and consumer decision-making.<br><br>1b. Identify the main consumer audiences for the report and describe their characteristics, their knowledge about the subject matter of the report, their information interests and needs, and how they will be expected to learn about and use the report. (In planning for use, provide for layering of information that permits the user to drill down to the technical details.)<br><br>1c. Identify secondary audiences for the report, such as healthcare providers and policymakers, and describe how their report-specific interests and needs differ from those of the main consumer audiences. Determine how the report will accommodate the secondary audiences (such as allowing users to drill down to the technical details about measurement and statistical comparisons). |
| 5a. Source and use of reports.<br>i. The entities producing reports of hospital performance should have the same general characteristics as data management/ analysis entities (i.e., independent, objective, and removed from any conflicts of interest). They should accept responsibility for establishing policies that guide the development of report content and format, | 2. Develop the web-based report using a transparent process that involves consumers and other relevant stakeholders.<br>2a. Identify the various stakeholders for the web-based report (these include, at a minimum, the developers and sponsors of the report, the main consumer audiences and organizations that represent these audiences, and | 2. Develop the report using a transparent process that involves consumers and other relevant stakeholders.<br>2a. Identify the various stakeholders for the report (these include, at a minimum, the developers and sponsors of the report, the main consumer audiences and organizations that represent these audiences, and the |

| A Comprehensive Framework for Hospital Care Performance Evaluation (2003) | National Voluntary Consensus Standards for Consumer-Focused Reporting (2007) | National Voluntary Consensus Standards for Public Reporting of Patient Safety Event Information (2010) |
|---|---|---|
| report production and distribution, and tasks involving education and diffusion of information. (This entity may use contractors, vendors, or agents to perform some or all of these tasks.)<br><br>5c. Verification of results.<br>i. Individual hospital results should be shared with that hospital by the reporting entity in advance of publishing and distributing the results publicly.<br><br>ii. Reporting entities should address individual concerns raised by hospitals about their results in an equitable manner that balances the needs of the community and hospitals with the goals of reporting.<br><br>iii. Reporting entities should establish policies and procedures for mediation with hospitals when individual concerns raised by hospitals about their results cannot be resolved.<br><br>iv. Reporting entities should be held accountable for errors in the reports that they publish. When such instances occur, reporting entities should, at a minimum, publicly retract the mistake and produce and distribute an errata sheet with subsequent distribution of that report.<br><br>v. Any self-reported results that are published by the hospital should be distinguished from externally validated/verified results published by the reporting entity.<br><br>vi. Reporting entities must distinguish NQF-endorsed measures from non-endorsed measures and explain why they are additionally reporting non-endorsed measures (e.g., measure is mandated by state law, measure is | the entities that are being measured and compared), and clarify their roles and responsibilities.<br><br>2b. Establish governance and decision-making rules.<br><br>2c. Provide an opportunity for the entities that are being measured and compared to preview their data and comment on the data's accuracy before the report is released; errors/misconceptions should be corrected and policies and procedures for mediation established.<br><br>2d. Involve consumers in the development and refinement of the report by seeking their input into the report design and getting their feedback on draft versions of language and data displays. Conduct usability/ease-of-use testing with consumers before the report is released, and then collect their feedback after the launch to help evaluate it. | entities that are being measured and compared), and clarify their roles and responsibilities.<br><br>2b. Establish governance and decision-making rules.<br><br>2c. Provide an opportunity for the entities that are being measured and compared to preview their data and comment on the data's accuracy before the report is released; errors or misconceptions should be corrected and policies and procedures for mediation established.<br><br>2d. Encourage organizations (healthcare organizations and/or providers) to describe, either as a part of or accessible from the public report, how these data may be used or have been used to improve safety.<br><br>2e. Involve consumers in the development and refinement of the report by seeking their input into the report design, where appropriate, and getting their feedback on draft versions of language and data displays. Conduct usability/ease-of-use testing with consumers before the report is released, and then collect their feedback after the launch to help evaluate it. |

| A Comprehensive Framework for Hospital Care Performance Evaluation (2003) | National Voluntary Consensus Standards for Consumer-Focused Reporting (2007) | National Voluntary Consensus Standards for Public Reporting of Patient Safety Event Information (2010) |
|---|---|---|
| being pilot tested). | | |
| | 3. At the beginning of the report, set the stage by communicating what quality is, how quality varies, and how making quality comparisons can be of value to consumers.<br>3a. Provide a brief introduction about healthcare quality.<br><br>3b. Explain that quality varies within and across institutions and how the report can be used to make quality comparisons.<br><br>3c. Use consistent, simple, and familiar language to discuss quality and provide examples that will resonate with the main consumer audiences. | 3. The report should establish a context by describing what patient safety is, including understanding the nature of patient safety events, explaining where the measures are in their development or evolution (i.e., how the measures may or may not be used for comparison across organizations over time—their robustness/usefulness). Reporters should consider linking to well-accepted national sources such as AHRQ, CDC, or NQF to accomplish this.<br>3a. Define terms.<br><br>3b. Explain adverse events in healthcare and how they can occur, and provide resources/links to consumer and patient-oriented resources (such as government and nonprofit sources) on topics such as infections, falls, pressure ulcers, safe surgery, medication use, and more.<br><br>3c. Discuss preventability of patient safety events and how the consumer can learn more about best practices to improve safety and about their role in improving safety.<br><br>3d. Explain how the report can be used to understand patient safety in healthcare organizations or providers.<br><br>3e. Use consistent, simple, and familiar language to discuss safety and provide examples that will resonate with the main consumer audiences. |
| | 4. Ensure that the measures included in a consumer-focused public report are meaningful to consumers, transparent, and meet widely accepted, rigorous criteria, including important, scientifically acceptable, feasible, and usable. | 4. Ensure that the measures included in a consumer-focused public report are meaningful to consumers, transparent, and meet widely accepted, rigorous criteria, including important, scientifically acceptable, feasible, and usable. |

| A Comprehensive Framework for Hospital Care Performance Evaluation (2003) | National Voluntary Consensus Standards for Consumer-Focused Reporting (2007) | National Voluntary Consensus Standards for Public Reporting of Patient Safety Event Information (2010) |
|---|---|---|
| | 4a. Because measures inherently have components that affect the way they should be reported, be clear about types of conclusions that can be reached.<br><br>4b. In choosing measures to be reported, take into account that the best measures:<br>i. are relevant to the healthcare-related concerns of the consumer audience;<br>ii. demonstrate variation and reflect care that those being measured can impact; and<br>iii. provide information that reflects the overall quality of care provided by the institutions included in the report (providing additional information about limited dimensions of care for specialty institutions is acceptable). | 4a. Provide context regarding the benefits and limitations of use of these data—make clear what they do and do not convey.<br><br>4b. In choosing measures to be reported, take into account that the best measures:<br>i. are relevant to the healthcare-related concerns of the public;<br>ii. provide information that reflects the safety of care provided by the organizations included in the report (while patient safety measures may reflect harm, they may not reflect improvements that have been made to reduce recurrence, and organizations should be encouraged to provide data of the efforts to reduce recurrence.); and<br>iii. are objective, valid, reliable, methodologically sound, feasible, transparent, verifiable, and represent consensus among stakeholders, including consumers and professionals. |
| 5b. Report generation<br>i. Reports prepared for consumers should include two components: a summary of the measure results and a technical supplement.<br>a) The summary of measure results should include:<br>i) annual results, appropriately risk adjusted and in composite form (as appropriate), for each measure in the endorsed set, unless a measure's specification necessitates less frequent measurement/reporting periods;<br>ii) guidance on how to interpret and use the results as well as the data's limitations; and<br>iii) reporting entity information (name, address, contact telephone number, fax number, and e-mail address). | 5. Present and explain the data clearly and objectively in ways that help consumers understand and use the information.<br>5a. Help consumers quickly and easily arrive at correct and meaningful conclusions.<br>i. Display data in formats that have been shown to be evaluable. This means summarizing and displaying the data for the viewer in a way that facilitates interpretation (e.g., summary scores, labels).<br>ii. To help users make correct interpretations, report measures in a consistent way so that, within a report, either a high score or a low score consistently indicates better performance.<br>iii. Make presentations of information more vivid and | 5. Present and explain the data clearly and objectively in ways that help consumers understand and use the information. For each measure to be included, a determination should be made whether it is appropriately displayed as a rate, as low frequency, and, in some cases whether the measure should be included in a composite.<br>5a. Help consumers to quickly and easily understand each measure and to use the information to aid in decision-making.<br>i. Display data in formats that have been shown to be evaluable. This means summarizing and displaying the data for the viewer in a way that facilitates interpretation (e.g., summary scores, labels, trends) without conveying |

| A Comprehensive Framework for Hospital Care Performance Evaluation (2003) | National Voluntary Consensus Standards for Consumer-Focused Reporting (2007) | National Voluntary Consensus Standards for Public Reporting of Patient Safety Event Information (2010) |
|---|---|---|
| b) The technical report should include:<br>i) trended annual results for at least three years for each endorsed measure;<br>ii) detailed measure definitions;<br>iii) measure specifications; and<br>iv) risk-adjustment methodologies applied, including limitations of risk adjustment.<br><br>ii. Results should be summarized using a standardized approach to composite measure development (an aggregate index for each group [or groups] of related measures). Development of such a standardized approach is a high priority. Until this approach exists, results should be reported individually for each measure.<br><br>iii. In presenting comparative results, the following should be taken into account:<br>a) Results for individual hospitals should be presented in comparison with local, regional, and national averages.<br>b) Reports should be presented based on a single, evidence-based template for reporting measures to consumers. This template should be voluntarily adopted on a national level by any reporting entity providing hospital performance results and should be clearly identified when used. Establishing such a template should be a high priority.<br>c) Evaluable} formats that have been tested to show that consumers can quickly and easily identify top choices should be used. A simple and attractive design, based on evidence of what is most likely to be understood by consumers and used for choice (e.g., legends, graphic aids, easy-to-decipher visual cues, and same-page displays) should be employed.<br>d) Reports should be published in print and electronic | compelling by including anecdotes or stories to illustrate the meaning of the data.<br>iv. Take advantage of web-based capabilities for subordinating and sorting information in order to make it responsive to the needs of users; that is, offer options that allow users to select which parts of the information they want to see and how they want to see it (e.g., listed in order of performance or alphabetically, shown in summary format or in detailed breakdowns).<br><br>5b. In presenting comparative quality information:<br>i. use tools and methods such as rank ordering, color coding, and/or symbols that help users discern performance variation and quickly determine their best options;<br>ii. when possible, include benchmarks to provide users a better context for making comparisons and using the information;<br>iii. provide risk-adjusted rates and grouping of information into categories such as ”better,” ”average” within standardized categories (such as by disease or by institution), when appropriate, and provide a simple explanation of why this was done; i.e., to make the comparisons fair and meaningful;<br>iv. label indicators using everyday language (not clinical or technical terms);<br>v. ensure that comparisons are accurate and supportable; and<br>vi. whenever possible, limit the use of statistics and terms that are difficult for most consumers to understand.<br><br>5c. In presenting data from composite measures:<br>i. where measures are interpretable at the individual measure level, report all measures that comprise the | misleading comparisons.<br>ii. To help users make correct interpretations, report measures in a consistent way so that, within a measure/group of measures, either a high score or a low score consistently indicates better performance.<br>iii. Make presentations of information more vivid and compelling by including anecdotes, stories, or case studies to illustrate the meaning of the data.<br>iv. Consider ancillary content to help consumers understand safe care (e.g., safe surgery checklist) and what they can do to contribute to improved safety.<br><br>5b. Use approaches such as those listed below to present comparative patient safety information.<br>i. Use tools and methods such as rank ordering, color coding, or symbols that help users to discern meaningful performance variation and quickly determine their best options.<br>ii. When possible, include context for making comparisons and using the information.<br>iii. Where applicable and appropriate, provide risk-adjusted rates and grouping of information into categories such as “better” and “average” within standardized categories (such as by disease or by institution) and provide a simple explanation of why this was done (e.g., to make the comparisons fair and meaningful).<br>iv. Label indicators using everyday language (not clinical or technical terms).<br>v. Ensure that comparisons are reasonable and supportable.<br>vi. Whenever possible, limit the use of statistics and terms that are difficult for most consumers to understand.<br><br>5c. Composite measures, if used, should be clinically |

| A Comprehensive Framework for Hospital Care Performance Evaluation (2003) | National Voluntary Consensus Standards for Consumer-Focused Reporting (2007) | National Voluntary Consensus Standards for Public Reporting of Patient Safety Event Information (2010) |
|---|---|---|
| formats. Electronic reports are useful in that they enable "drill down" and user self-customization. Electronic results should be analyzed and displayed in two different manners: by hospital and by condition (when applicable).<br><br>iv. Regarding the sample size for reporting:<br>a) There must be a minimum of 30 annual cases in the denominator of a measure for the reporting entity to report hospital results on that measure. When insufficient case volume prevents a reporting entity from reporting individual hospital results, it should be determined by the reporting entity whether aggregating data at a higher level (e.g., all small hospitals in a region) might be useful to consumers or other stakeholder audiences.<br>b) When a hospital has insufficient case volume to meet the minimum threshold of 30 cases, the reporting entity should report, in a manner that is understood by consumers, that there is insufficient data to indicate that there are too few cases for the measure to be reported with sufficient precision/confidence.<br>c) If a hospital does not admit patients with a particular diagnosis or does not perform a particular procedure being measured, the reporting entity should report, in a manner that is understood by consumers, that the measure is not applicable, in order to indicate the service is not provided by the reporting hospital.<br><br>v. Measures should be reported by race/ethnicity (consistent with NQF's report, *Improving Healthcare Quality for Minority Patients*), age, and gender of patient subpopulations, as well as for the hospital population as a whole.<br><br>vi. Reports should be translated by the reporting entity | composite without adding or deleting any individual component or make any change to the composite transparent (at a layer down from the initial data display); and<br>ii. report results for the composite and for each component measure (at a layer down from the initial composite data display).<br><br>5d. In providing contextual information/decision support:<br>i. provide a clear contextual framework as part of the report introduction;<br>ii. make sure that key messages are included in the data display;<br>iii. whenever data are missing, provide a specific explanation for this and make the distinction clear between data that are missing because of small numbers (too few to report) and data that are missing because of refusal to provide the data;<br>iv. make information understandable by using everyday words and language;<br>v. use consumer testing to verify that the language and displays provided in the report are easy for the intended consumer audiences to understand and use (provide translations into languages other than English, if needed); and<br>vi. use reasonably current data, and display the dates/period that are covered by the data.<br><br>5e. In presenting technical documentation:<br>i. include detailed measure definitions, specifications, and risk-adjustment methods;<br>ii. include resource information such as identification of the measure developer, sources of data, and interpretation guides; and | coherent, actionable, and transparent.<br>i. Explain what a composite is and how it is constructed (in consumer language).<br>ii. Give examples to demonstrate how a composite may accurately reflect underlying safety or how it may fail to give an accurate depiction (e.g., if it averages widely varying results).<br>iii. Where measures are interpretable at the individual measure level, report all measures that comprise the composite without adding or deleting any individual component, or ensure transparency in the composite (at a layer down from the initial data display).<br>iv. Report results for the composite and for each component measure (at a layer down from the initial composite data display).<br><br>5d. Provide context for low-frequency events.<br>i. Explain how low-frequency events are identified, collected, and displayed and how patient confidentiality is maintained.<br>ii. Discuss the use of low-frequency events in assessing quality and safety of healthcare provider.<br>iii. Retain and make accessible reports from year to year. In doing so, it would be appropriate to provide information about variation over time.<br><br>5e. Provide context for adverse events displayed by rates.<br>i. Explain measures of adverse events that are calculated as rates.<br>ii. Discuss the use of rates in assessing quality and safety of a healthcare provider.<br>iii. Retain and make accessible reports from year to year. In doing so, it would be appropriate to provide information |

| A Comprehensive Framework for Hospital Care Performance Evaluation (2003) | National Voluntary Consensus Standards for Consumer-Focused Reporting (2007) | National Voluntary Consensus Standards for Public Reporting of Patient Safety Event Information (2010) |
|---|---|---|
| into the key languages read by the communities served by the hospitals whose results are reported. Reports and related materials for consumers and patients should be written at the sixth-grade reading level.<br><br>vii. Costs of hospital performance measurement reporting should be shared among purchasers, providers, and other groups (e.g., consumers, employers). Burden reduction for hospitals should be achieved through consensus and standardization of measures and reporting methods and by the use of technology (e.g., electronic medical record), not by reducing the availability of data relevant to consumers and purchasers.<br><br>5d. Distribution and dissemination of reports.<br>i. The frequency of reports and the data used to prepare them should be as follows:<br>a) Published reports to consumers should be updated at least annually unless the specifications of a measure necessitate data reporting less frequently (e.g., percent of low-risk patients who received urine protein testing or dilated eye exam within the past two years).<br>b) The most recent data published should be no more than two years old.<br>c) An aggregate mean and comparison for each composite measure should be reported to the public.<br>d) Because multiple years of data will not be available initially, technical reports containing trended data should have one year of data the first year and build in subsequent years to no fewer than three years of data. The absence of three years of data and the reason behind these more limited, available trends should be noted in the reports.<br>e) As new measures are added or existing measures are | iii. provide details about methodology. | about variation over time.<br><br>5f. In providing contextual information/decision support:<br>i. provide a clear contextual framework as part of the report introduction;<br>ii. make sure that key messages are included in the data display;<br>iii. make clear that reports of low-frequency/rare events are different from rates—distinguish between appropriate uses of different kinds of data;<br>iv. provide a specific explanation for any missing data and make the distinction clear between data that are missing because of small numbers (i.e., events that occur so infrequently that meaningful comparisons cannot be drawn from rate calculations) and data that are missing because of refusal to provide the data;<br>v. make information understandable by using everyday words and language;<br>vi. use consumer testing to verify that the language and displays provided in the report are easy for the intended consumer audiences to understand and use (in addition to English, provide content in the key languages of the consumer audiences);<br>vii. use most current data available, and display the dates/period that are covered by the data;<br>viii. provide context of comparison to peers, to self over time, and to optimum performance (policy goals); and<br>ix. clearly explain risk stratification, that is, where it is done, why it is important.<br><br>5g. In presenting technical documentation, address verifiability, reliability, validity, data sources, and data collection (e.g., self-reported versus IT system-generated; voluntary versus mandatory, etc.). |

| A Comprehensive Framework for Hospital Care Performance Evaluation (2003) | National Voluntary Consensus Standards for Consumer-Focused Reporting (2007) | National Voluntary Consensus Standards for Public Reporting of Patient Safety Event Information (2010) |
|---|---|---|
| modified or terminated, there may be discontinuity of data elements in the technical report. In these instances, reports should indicate what has occurred to result in discontinuity.<br><br>ii. A variety of secondary distribution channels and vehicles (e.g., unions, local businesses, providers, libraries, media outlets, speakers bureaus, and other regional or local organizations) should make reports available once published by reporting entities. | | i. Include detailed measure definitions, specifications, and risk-adjustment methods.<br>ii. Describe verifiability of the data (if any) through audits, reviews, cross-checking with other data sources, or attestation by the provider.<br>iii. Define data sources, quality control, and the data collection process.<br>iv. Explain whether data are collected as part of a legal or accreditation mandate, or on a voluntary basis.<br>v. Include resource information, when available, such as identification of the measure developer, sources of data, and interpretation guides.<br>vi. Provide complete details about methodology. *(The report should not use any measures or data that lack complete transparency as to methodology.)* |
| | 6. Ensure that report design and navigation features enhance report usability. Design features should be used to:<br>6a. organize information in a way that lets users know what is available and lets them make their own choices;<br><br>6b. provide an engaging format and include intuitive and consistent navigation tools that are placed in consistent locations;<br><br>6c. make the report easy to skim and build in layering to provide the capability to drill down to information and to navigate back out;<br><br>6d. seek feedback and test the design and navigation with the intended audiences; and<br><br>6e. provide users a way to print the information in | 6. Ensure that report design and navigation features enhance report usability. Web-based reports are recommended because of their design, display, and navigation capabilities.<br>6a. organize information in a way that lets users know what is available and lets them make their own choices;<br><br>6b. provide an engaging format and include intuitive and consistent navigation tools that are placed in consistent locations;<br><br>6c. make the report easy to skim and build in layering to provide the capability to drill down to information and to navigate back out;<br><br>6d. seek feedback and test the design and navigation with the intended audiences;<br><br>6e. provide users a way to print the information in |

| A Comprehensive Framework for Hospital Care Performance Evaluation (2003) | National Voluntary Consensus Standards for Consumer-Focused Reporting (2007) | National Voluntary Consensus Standards for Public Reporting of Patient Safety Event Information (2010) |
|---|---|---|
| | understandable and usable formats. | understandable and usable formats;<br><br>6f. make it easy to locate/access ancillary information (in a contextually relevant way); and<br><br>6g. encourage consumer interaction through an easy-to-use comment feature (e.g., e-mails, FAQs, etc.). |
| 5e. Consumer research.<br>The following areas should become priorities for research, development, and further investigation to fully inform the improvement of approaches to consumer reporting:<br><br>i. the way in which consumers access and use reported results to determine how best to support consumers' uses of reports (e.g., research should be undertaken to understand the various audiences for hospital care performance reports, such as patients, surrogates of patients, and family members, and their use/s of the information, as well as the "tipping point"—the extent of effort required to affect those who are most likely to want the information and act on it);<br><br>ii. the most appropriate, evaluable approaches and formats for presenting reports to consumers;<br><br>iii. the most appropriate method/s of developing composite results for consumer reporting; and<br><br>iv. the effectiveness of reporting comparative results to consumers. | 7. Regularly review and assess reports to ensure their effectiveness, usability, and currency<br>7a. Conduct assessments of the use and impact of reports.<br><br>7b. Use a combination of methods to obtain and use feedback from the intended consumer audiences and the institutions that are the subject of the reporting.<br><br>7c. Involve stakeholders in revisions and seek their feedback after the report undergoes significant changes.<br><br>7d. Use what is learned to help inform and drive the improvement and usefulness of performance measures and the field of consumer public reporting. | Regularly review and assess reports to ensure their effectiveness, usability, and currency.<br>7a. Define the intended impact of the report, and measure usage/penetration and impact against that goal.<br><br>7b. Use a combination of methods such as population-based surveys, focus groups, and direct consumer reports, which may be conducted internally or externally, to obtain and use feedback from the intended consumer audiences and the institutions that are the subjects of the reporting.<br><br>7c. Involve stakeholders in revisions and seek their feedback after the report undergoes significant changes.<br><br>7d. Use what is learned, including identification of unintended consequences of report publication, to help inform and drive the improvement and usefulness of performance measures and the field of consumer public reporting. |

## APPENDIX B—TASK FORCE MEMBERS

**Christopher Queram, MA (Chair)**
President/CEO, Wisconsin Collaborative for Healthcare Quality, Middleton, WI

**Paul Conlon, PharmD, JD**
SrVP, Clinical Quality and Patient Safety, Trinity Health, Novi, MI

**Carol Cronin, MSW, MSG**
Executive Director, Informed Patient Institute, Annapolis, MD

**R. Adams Dudley, MD, MBA**
Professor of Medicine and Health Policy, University of California at San Francisco, San Francisco, CA

**Judith Hibbard, DrPH**
Professor, Dept. of Planning, Public Policy & Management, University of Oregon, Eugene, OR

**Dana Safran, Sc.D**
Senior Vice President/Performance Measurement & Improvement, Blue Cross Blue Shield of Massachusetts, Boston, MA

**Patrick Romano, MD, MPH**
Professor of Medicine and Pediatrics, UC Davis Division of General Medicine, Sacramento, CA

**John Santa, MD, MPH**
Director, Consumer Reports Health Ratings Center, Yonkers, NY

**Shoshanna Sofaer, MPH, DrPH**
Professor, Robert Luciano Chair of Health Care Policy, Baruch College of CUNY, New York, NY