NATIONAL QUALITY FORUM

GUIDANCE FOR EVALUATING

USABILITY AND USE OF PERFORMANCE MEASURES

Approved by the Board
February 24, 2012

Guidance for Evaluating Usability and Use of Performance Measures

CONTENTS

EXECUTIVE SUMMARY

The National Quality Forum's (NQF's) three-part 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.

Several challenges prompted a review of the Usability criterion. Public reporting of performance results continues to be of great interest; however, with the passage of health reform legislation and today's quality environment, the NQF portfolio must be broad enough to support additional core accountability applications, such as value-based payment, health information technology incentive payments, accreditation, and regulation. NQF-endorsed© measures are intended to be used in both accountability and quality improvement and must be implemented in order to facilitate NQF's mission to improve the quality of healthcare. However, some measures are not implemented by the time of endorsement maintenance, and therefore continued endorsement is questioned. NQF convened a nine-member Usability Task Force to review and refine the NQF Usability criterion and subcriteria.

Recommendations

The Task Force recommended changes to the criterion, defined terms, identified key questions to guide evaluation, and suggested revisions to the measure submission items. The revised criterion of Usability and Use is provided in Table 1. The Task Force confirmed that Usability and Use apply to accountability/transparency and improvement and that actual use and demonstrated improvement are the ultimate demonstration of usability. The recommendations include:

- 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.
- Consider the positive and negative effects of measurement. 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 (after the two must-pass criteria of Importance to Measure and Report and Scientific Acceptability of Measure Properties and Feasibility) because if the other criteria are met, then a measure should be usable.
- Usability and Use is not a must-pass criterion. If a measure is not in use or demonstrating improvement, then the determination of its suitability for continued endorsement requires an assessment of the factors involved and judgment about its potential to be put into use or to improve care.

Table 1. Evaluation Criteria for Usability and Use

Condition for Consideration C. The intended use of the measure includes both accountability applications[1] and performance improvement to achieve high-quality, efficient healthcare.
4. 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[2] to achieve the goal of high-quality, efficient healthcare for individuals or populations. 4a. Accountability and Transparency[3] Performance results are used in at least one accountability application[1] within three years after initial endorsement and are publicly reported[3] within six years after initial endorsement (or the data on performance results are available).[4] If not in use at the time of initial endorsement, then a credible plan[5] for implementation within the specified timeframes is provided. AND 4b. Improvement[6] Progress toward achieving the goal of high-quality, efficient healthcare for individuals or populations is demonstrated.[6] If not in use for performance improvement at the time of initial endorsement, then a credible rationale describes how the performance results could be used to further the goal of high-quality, efficient healthcare for individuals or populations. AND 4c. 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 (if such evidence exists).
Criteria Notes 1. Accountability applications are the use 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. 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. 3. Transparency is the extent to which performance results about identifiable, accountable entities are *disclosed and available* outside of the organizations or practices whose performance is measured. Maximal transparency is achieved with public reporting defined as making comparative performance results about identifiable, accountable entities freely available (or at nominal cost) to the public at large (generally on a public website). *At a minimum, the data on performance results about identifiable, accountable entities are available to the public (e.g., unformatted database)*. The capability to verify the performance results adds substantially to transparency. 4. This guidance is not intended to be construed as favoring measures developed by organizations that are able to implement their own measures (such as government agencies or accrediting organizations) over equally strong measures developed by organizations that may not be able to do so (such as researchers, consultants, or academics). Accordingly, measure developers may request a longer timeframe with appropriate explanation and justification. 5. Credible plan includes the specific program, purpose, intended audience, and timeline for implementing the measure within the specified timeframes. A plan for accountability applications addresses mechanisms for data aggregation and reporting. 6. Demonstrated progress toward achieving the goal of high-quality, efficient healthcare includes evidence of improved performance and/or increased numbers of individuals receiving high-quality healthcare. Exceptions may be considered with appropriate explanation and justification.

INTRODUCTION

The National Quality Forum's (NQF's) three-part 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 decisionmaking related to accountability and improvement. During its March 2011 retreat, the Consensus Standards Approval Committee (CSAC) 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 information technology [IT] incentive programs). The CSAC also expressed interest in further delineating NQF expectations of measure stewards to demonstrate that their measures are 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

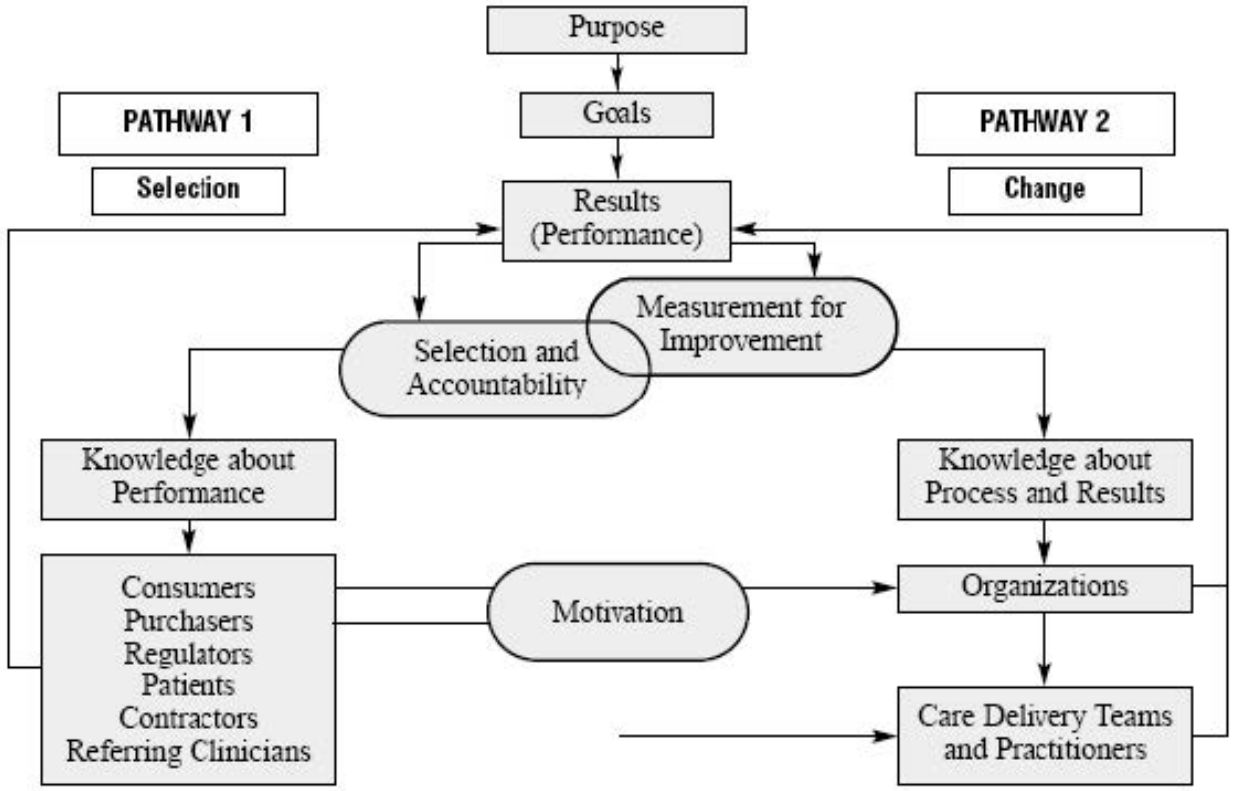
Historically, NQF's work has revolved around endorsing performance measures useful for both quality improvement and accountability, with an emphasis on transparency and public reporting. In October 2009, the NQF Board of Directors affirmed a general expectation that performance results from NQF-endorsed© 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 must 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. Less clear are the expectations regarding actual use in the various programs at the time of endorsement maintenance and the process for evaluating usefulness for decisionmakers and improvement.

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 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 about the selection of 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 insufficient resources to test the understanding and usefulness of performance results for various accountability applications or quality improvement.
- An entity other than the measure developer/steward may be the one implementing the measure and consequently is in the best position to demonstrate usability.
- At the time of endorsement maintenance, some 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 must be publicly available. However, the data needed to implement a measure often are owned or collected by other entities. Occasionally, the measure developer/steward has sole control of the data needed to compute and report on a performance measure; and some stakeholders question whether NQF should endorse such measures without a commitment and plan for public reporting.
- Some stakeholders think that endorsement should not be continued for measures that are not in use. However, they also acknowledge the concern that good measures could be lost simply because they have not yet been implemented. More experience is needed to determine how long it takes to achieve use.
- Although transparency of performance information is critical to supporting accountability and selection, various degrees of transparency as well as a variety of accountability functions can help to drive improvement without public reporting of performance scores.
- 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 concern that failure to require public reporting of all NQF-endorsed measures will result in very limited information being available to support selection (i.e., the “slippery slope”).
- There also is concern that excessive emphasis on public reporting will result in overload of patients/consumers. The focus should be on providing the right information to

consumers that will help them to 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 to be classified or reported (e.g., using stars to indicate ranking, stating whether results are above or below average, using confidence intervals). Occasionally, measures have been submitted with methods for classifying the results, but NQF requires those methods to be separated from the endorsed measure. The rationale for this position is twofold: 1) a measure may be used in more than one application, and the classification and format of 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, processes, 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 subcriterion for unintended consequences (4c) has been considered under Feasibility, but it has been suggested that it would be more appropriate under Usability.
- Disparities have been considered under performance gap (1b) and measure specifications to detect disparities (2c), but it has been suggested that disparities be considered under Usability.

RECOMMENDATIONS

The Task Force developed definitions and principles to guide its discussions and recommendations.

Definitions

Because terms such as *accountability* and *public reporting* have been used inconsistently, 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 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).

Public Reporting: Making comparative performance results about 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 choosing a surgeon; an employer choosing a health plan to offer; a health plan choosing specialists to empanel; a family doctor choosing an oncologist to refer a cancer patient; an employee or Medicaid enrollee choosing a health plan during open enrollment).

Transparency: Extent to which performance results about identifiable, accountable entities are *disclosed and available* outside of the organizations or practices whose performance is measured. The degrees of transparency are described in Table 2 and range 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.

Table 2. Degrees of Transparency

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

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 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. 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 their influence on progress toward 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 also ensures accountability and provides external motivation for performance improvement. NQF encourages transparency of performance results.
- 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. In its 2010 report, the NQF-Quality Alliance Steering Committee Workgroup encouraged collaboration between developers and potential implementers of performance measures because resources and efforts to develop and test measures could be wasted if the measures are not implemented.[3]
- 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 outweigh associated burden.

I. Recommendations for Measure Evaluation Criteria for Usability and Use

The Task Force discussed the central question: Do measures ever fail to be endorsed *only* because they fail to meet the Usability criterion? If no measures fail to meet this criterion, then it may not be a criterion. In other words, measures that meet Importance to Measure and Report and Scientific Acceptability of Measure Properties would be expected to be usable to some audience. To date, measures have not failed to be endorsed based solely on the Usability criterion. Issues 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 to meet the current Usability criterion because understandability or interpretability was not demonstrated, but usually steering committees have not viewed this as a fatal flaw. Instead, it has been viewed as a problem that can be corrected 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. A measure may not be in use because of problems related to other criteria, such as opportunity for improvement, evidence, reliability, validity, or unintended consequences, or because the measure steward controls and/or limits access to performance results or the underlying data. In some cases, however, implementation depends on external factors beyond the measure steward's or developer's control (e.g., a measure is specified for electronic health records [EHRs], but EHRs are not yet widely adopted). 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, then significant resources may be wasted if the measure is not put into use.

The Task Force determined that the concepts of usability and usefulness are related to a specific purpose. The general purpose of NQF-endorsed measures is to facilitate high-quality, efficient healthcare for all Americans. Theoretically, measures that meet the NQF criteria for Importance to Measure and Report and Scientific Acceptability of Measure Properties should be 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). Rather, MAP will address the selection of specific endorsed measures for specific programs.

The Task Force agreed that understanding and interpretability, which are related to a specific audience and implementation conditions (e.g., language used and how the results are displayed), should not be considered under the Usability criterion. Several other NQF projects produced guidance on reporting performance results to help improve understanding and thus usability for key audiences.[4-6] 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 prior guidance suggested grouping information into categories such as "better" or "average" but did not address the methodological issues involved in determining the categories, such as statistical analyses, or how to convey certainty around performance scores.

Because the goal for NQF-endorsed measures is to facilitate improvements in healthcare and health, they must 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 3. Although transparency through public reporting is still the goal, other accountability applications also are recognized.

The Task Force also discussed several additional concepts for potential consideration under the Usability criterion: unintended consequences, disparities, and methods for classifying performance.

- The Task Force agreed that unintended negative consequences should be considered under Usability along with the evidence of use and influence on quality. 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.) 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 information should be solicited from users of endorsed measures. Reports of negative consequences should be accompanied by evidence including the nature of the consequence, the affected party, the number of people affected, and the severity of the impact.
- The Task Force concluded that disparities in care should be addressed early in the evaluation criteria and therefore not under Usability. Currently disparities are assessed as 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 made recommendations about identifying disparities-sensitive measures.
- Finally, the task force agreed that a measure developer's guidance on reporting performance results such as methods for classifying performance results (e.g., stars) should not be considered under Usability because it depends on context and is not a core part of the measure construction. NQF measure endorsement should focus on the performance measure rather than on the methods for reporting performance results. However, the ways in which

performance results are reported can affect understanding or even the validity of the conclusions made. This is an important issue, and NQF should identify the pros and cons of including such reporting guidance as part of measure. NQF's role could range from providing general 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.

Comments Received on Proposed Evaluation Criteria

Comments were generally in favor of the recommended modifications to the evaluation criteria. Some commenters suggested that requirements for use and public reporting should be more stringent, while others suggested less emphasis on public reporting. NQF's CSAC recommended that the criteria set explicit expectations and specific timeframes for use in accountability applications and public reporting of performance results (or availability of data on performance results). However, application of the criteria should be flexible and based on an assessment of the reasons for lack of use or improvement and the likelihood of progress with more time.

Based on the feedback, the Task Force revised the criteria as follows:

- The subcriteria were reorganized to more clearly identify accountability and transparency versus improvement.
- Expectations for use were explicitly stated with timeframes; however, the need for flexibility was described in the explanatory notes.
- Expectations for public reporting of performance results (or availability upon request) were explicitly stated.

The Task Force cautioned that setting specific timeframes for use could have the unintended consequences of removing endorsement of good performance measures because of efforts to resist measurement and stifling the efforts of independent measure developers and innovations in measure development.

Table 3. Evaluation Criteria for Usability and Use—Prior and Modified

Prior Criteria	Modified Criteria
Condition for Consideration C. The intended use of the measure includes both public reporting and quality improvement.	Condition for Consideration C. The intended use of the measure includes both accountability applications[1] and performance improvement to achieve high-quality, efficient healthcare.
3. Usability Extent to which intended audiences (e.g., consumers, purchasers, providers, policymakers) can understand the results of the measure and find them useful for decisionmaking. 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; AND 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.	4. 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[2] to achieve the goal of high-quality, efficient healthcare for individuals or populations. 4a. Accountability and Transparency[3] Performance results are used in at least one accountability application[1] within three years after initial endorsement and are publicly reported[3] within six years after initial endorsement (or the data on performance results are available).[4] If not in use at the time of initial endorsement, then a credible plan[5] for implementation within the specified timeframes is provided. AND 4b. Improvement[6] Progress toward achieving the goal of high-quality, efficient healthcare for individuals or populations is demonstrated.[6] If not in use for performance improvement at the time of initial endorsement, then a credible rationale describes how the performance results could be used to further the goal of high-quality, efficient healthcare for individuals or populations. AND 4c. 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 (if such evidence exists).
Note 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 1. Accountability applications are the use 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. 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. 3. Transparency is the extent to which performance results about identifiable, accountable entities are *disclosed and available* outside of the organizations or practices whose performance is measured. Maximal transparency is achieved with

Prior Criteria	Modified Criteria
	public reporting defined as making comparative performance results about identifiable, accountable entities freely available (or at nominal cost) to the public at large (generally on a public website). *At a minimum, the data on performance results about identifiable, accountable entities are available to the public (e.g., unformatted database)*. The capability to verify the performance results adds substantially to transparency. 4. This guidance is not intended to be construed as favoring measures developed by organizations that are able to implement their own measures (such as government agencies or accrediting organizations) over equally strong measures developed by organizations that may not be able to do so (such as researchers, consultants, or academics). Accordingly, measure developers may request a longer timeframe with appropriate explanation and justification. 5. Credible plan includes the specific program, purpose, intended audience, and timeline for implementing the measure within the specified timeframes. A plan for accountability applications addresses mechanisms for data aggregation and reporting. 6. Demonstrated progress toward achieving the goal of high-quality, efficient healthcare includes evidence of improved performance and/or increased numbers of individuals receiving high-quality healthcare. Exceptions may be considered with appropriate explanation and justification.

II. Recommendations for Evaluating Usability and Use

The goal for 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 used. 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 is evidence based; is reliable and valid; and is 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 focus on the observed use of a measure and progress toward achieving high-quality, efficient healthcare. If a measure is already in use at the time of initial endorsement, then use and progress toward this goal can be evaluated at that time. In addition to the information submitted by the measure developer, comments from the field will help to identify use or reasons for lack of use or improvement. On evaluation for endorsement maintenance, lack of use or improvement may signal 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 health records, data collection burden, or privacy concerns?

If the other criteria are clearly met, then attention should shift to determining the reasons for lack of use and whether they indicate a justification to retain or remove endorsement:

- Does evidence of unintended negative consequences to individuals or populations outweigh the benefit?
- To what extent is the measure steward, measure developer, or the entities being measured responsible for lack of access to performance results or the data needed to implement the measure?
- To what extent are entities whose performance is being measured resisting performance measurement and/or reporting?
- Are there other external factors delaying the measure's implementation (e.g., competing priorities, funding, legislative mandates)?

Although measures must be in use to influence quality, setting a specific deadline by which measures must be in use to retain endorsement should be approached cautiously and with flexibility. The amount of time required to implement a measure depends on factors such as whether data are already being collected and whether systems for aggregating, analyzing, and reporting performance results are already established. Additional time may be needed to pilot test the presentation of performance results. External factors, such as limited funding or competing priorities, also may slow implementation. Some measures may be ahead of their time, for example, those that specified for electronic health records, whose adoption has been slow.

The Task Force agreed that if performance results are not available for use in an accountability application because of actions or policies of the measure developer or steward, then continued NQF endorsement may not be warranted. However, lack of use also could be due to the actions of parties external to the measure developer or steward that are directed at resisting performance measurement. Assessment of Usability and Use will require the judgment of NQF multi-stakeholder steering committees.

Public reporting, defined as making comparative performance results about identifiable, accountable entities freely available (or at nominal cost) to the public at large (generally on a public website), may not be an absolute requirement for every endorsed measure. Some measures may not be useful for public reporting but are useful for other accountability applications and contribute to improving health and healthcare. However, perspectives of the usefulness of measures for public reporting often vary. For example, providers and consumers may have different opinions on whether measures are too technical or complicated for consumers to understand. Additionally, as stated in the principles, public reporting serves purposes other

than 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 data or testing that demonstrate a measure is not useful or could not be made useful through translation of technical terms or appropriate framing and information on how to interpret and use the data. Additionally, the availability of resources for publicly reporting performance results may be limited. However, if a measure is used in an accountability application, then the data on performance results should be available to the public even if it is an unformatted database.

The amount of time needed to demonstrate improvement also is difficult to predict and may vary by topic or type of measure. With more experience over time, the criteria for specific timeframes, public reporting, and demonstrated improvement should be reassessed.

This guidance for evaluating Usability and Use is consistent with the recent guidance for evaluating competing measures—that is, competing measures should be compared on all the criteria and subcriteria, including Usability and Use. If measures are considered equal on Importance to Measure and Report, Scientific Acceptability of Measure Properties, and Feasibility, then measures should be compared on Usability and Use to determine superiority. For example, if all other criteria are equal, then a measure in use will be considered superior to one not in use. However, differences between competing measures on the criteria and subcriteria are likely, and the steering committee should weigh the strengths and weaknesses across all the criteria. If a competing measure does not have clear superiority, then the steering committee should assess justification for multiple measures.

Comments Received on Guidance for Evaluation

In its initial work the Task Force decided that a rating scale would be more complicated than useful. Therefore, the draft guidance included only the questions and factors that the Task Force thought steering committees should consider. However, some commenters found the proposed guidance table to be complicated, and the CSAC asked the Task Force to reconsider developing a rating scale to provide more specific guidance for evaluating Usability and Use. The Task Force evaluated some potential rating scales but affirmed its original decision for the following reasons:

- A measure is either in use or not, and improvement is either demonstrated or not. Currently there is no basis for identifying cut points on a rating scale for extent of use or improvement.
- A rating scale would need to be accompanied by some decision logic for interpreting the consequences of the ratings on accountability and improvement, which would increase complexity.

- The three subcriteria for Usability and Use are joined by "AND," so the intent is that all three criteria must be met. Failure to meet the criteria for Usability and Use will trigger an assessment of the reasons for lack of use, public reporting, or improvement and the context (e.g., external factors, existence of other comparable or related measures or a credible plan for implementation, and the strengths of the proposed measure) to determine whether a measure is suitable for endorsement.

Table 4 provides the key questions that must be addressed when evaluating Usability and Use and some of the implications of determining a measure's suitability for endorsement. A final recommendation for endorsement also is dependent on addressing measure harmonization and competing measures. The Task Force emphasized that endorsement decisions occupy a gray area and their consequences must be carefully examined and weighed.

Table 4. Key Questions for Evaluating Usability and Use

Subcriteria	Key Questions	Suitable for Endorsement?
4a, 4b, 4c	• Are all three subcriteria met? (4a—accountability/transparency, 4b—improvement, and 4c—benefits outweigh any unintended consequences)	If Yes, then the Usability and Use criterion is met, and if the other criteria (Importance to Measure and Report, Scientific Acceptability of Measure Properties, Feasibility) are met, then the measure is suitable for endorsement
4a. Accountability/ Transparency	• Is it an initial submission with a credible plan for implementation in an accountability application? • Is the measure used in at least one accountability application by three years? • Are the performance results publicly reported by six years (or the data on performance results are available)? If any of the above answers are "No": • What are the reasons (e.g., developer/steward, external factors)? • Is there a credible plan for implementation and public reporting?	If 4a and/or 4b are not met, then the Usability and Use criterion is not met, but the measure may or not be suitable for endorsement depending on an assessment of the following: • timeframe (initial submission, three years, six years, or longer); • reasons for lack of use in accountability application/public reporting (4a) and/or lack of improvement (4b); • credibility of plan for implementation for accountability/public reporting (4a) and/or credibility of rationale for improvement (4b); • strength of the measure in terms of the other three criteria (Importance to Measure and Report, Scientific Acceptability of Measure Properties, and Feasibility); and • strength of competing and related measures to drive improvement. Exceptions to the timeframes for accountability and public reporting (4a) OR demonstration of improvement (4b) require judgment and supporting rationale.
4b. Improvement	• Is it an initial submission with a credible rationale for improvement? • Has improvement been demonstrated (performance trends, numbers of people receiving high-quality, efficient healthcare)? If any of the above answers are "No": • What are the reasons? • Is there a credible rationale describing how the performance results could be used to further the goal of facilitating high-quality, efficient healthcare for individuals or populations? • Is the measure used in quality improvement programs?	
4c. Unintended negative consequences	• Is there evidence that unintended negative consequences to individuals or populations outweigh the benefits? For most measures, this will not be applicable and will not be a factor in whether a measure is recommended.	If Yes, then the Usability and Use criterion is not met and the measure is not suitable for endorsement regardless of evaluation of 4a and 4b.

III. Recommendations for Measure Submission Items for Usability and Use

The information requested on the measure submission form must be modified as indicated in Table 5 to be consistent with the changes to the criteria.

Table 5. Measure Submission Items

Modified Criteria	Proposed Measure Submission Items to Evaluate the Criteria
4. 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[2] to achieve the goal of high-quality, efficient healthcare for individuals or populations.	*4.1. Current Use *(Check all the current uses; for any that are checked, provide a URL for the specific program.)* • Public Reporting _____ • Public Health/Disease Surveillance_____ • Payment Program _____ • Regulatory and Accreditation Programs _____ • Professional Certification or Recognition Program _____ • Quality Improvement with Benchmarking (external benchmarking to multiple organizations) _____ • Quality Improvement (internal to the specific organization) _____ • Other _____ • Not in use • Use unknown 4.2. For each use, checked above, provide: • Name of program and sponsor • Purpose • Geographic area and number and percentage of accountable entities and patients included
4a. Accountability and Transparency[3] Performance results are used in at least one accountability application[1] within three years after initial endorsement and are publicly reported[3] within six years after initial endorsement (or the data on performance results are available).[4] If not in use at the time of initial endorsement, then a credible plan[5] for implementation within the specified timeframes is provided. AND	*4a.1. If not currently used in at least one accountability application, then identify the reasons (including any policies or actions of the developer/steward or accountable entities that restrict access to performance results or block implementation). 4a.2. If not currently used in at least one accountability application, provide a credible plan for implementation. *(Credible plan includes the specific program, purpose, intended audience, and timeline for implementing the measure within the specified timeframes. A plan for accountability applications addresses mechanisms for data aggregation and reporting.)* *4a.3. If not currently publicly reported, then identify the reasons (including any policies or actions of the developer/steward or accountable entities that restrict access to performance results or block implementation). 4a.4. If not currently publicly reported, provide a credible plan for public reporting or availability of data on performance results. *(Credible plan includes the specific program, purpose, intended audience, and timeline for implementing the measure within the specified timeframes. A plan for accountability applications addresses mechanisms for data aggregation and*

Modified Criteria	Proposed Measure Submission Items to Evaluate the Criteria
	reporting.)
4b. Improvement[6] Progress toward achieving the goal of high-quality, efficient healthcare for individuals or populations is demonstrated.[6] If not in use for performance improvement at the time of initial endorsement, then a credible rationale describes how the performance results could be used to further the goal of high-quality, efficient healthcare for individuals or populations. AND	4b.1. Provide a rationale that describes how the performance results are or could be used to achieve the goal of high-quality, efficient healthcare. *4b.2 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)* • Source of data • Geographic area and number and percentage of accountable entities and patients included • Progress (trends in performance results, number and percentage of people receiving high-quality healthcare) 4b.3. If no improvement demonstrated, then identify the reasons.
4c. 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 (if such evidence exists).	4c.1. Were any unintended negative consequences to individuals or populations identified during testing, or has evidence of unintended negative consequences to individuals or populations been reported since implementation? If so, then identify the negative unintended consequences and describe how benefits outweigh them or actions taken to mitigate them. * *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-Quality Alliance Steering Committee Workgroup. Enhancing Availability of Performance Information - Tab 6 Report to NQF Board 9/23/2010. *National Quality Forum* 2010; Available at: www.qualityforum.org/WorkArea/linkit.aspx?LinkIdentifier=id&ItemID=39504. Last accessed July 2011.

4. NQF, *A Comprehensive Framework for Hospital Care Performance Evaluation,* Washington, DC: National Quality Forum; 2003.

5. NQF, *National Voluntary Consensus Standards for Hospital Care 2007—Guidelines for Consumer-Focused Reporting,* Washington, DC: NQF; 2009.

6. 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): Public Reporting Public Health/Disease Surveillance Payment Program Regulatory and Accreditation Programs Professional Certification or Recognition Program Quality Improvement with Benchmarking (external benchmarking to multiple organizations) 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.)* Public Reporting Public Health/Disease Surveillance Payment Program Regulatory and Accreditation Programs Professional Certification or Recognition Program Quality Improvement with Benchmarking (external benchmarking to multiple organizations) Quality Improvement (internal to the specific organization) Not in use Use unknown 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 three years of endorsement.)* 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.)* 3b.1. Use in Quality Improvement *(If used in quality improvement program, provide name of program(s), locations, Web page URL(s).)* 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).)* 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[4]	National Voluntary Consensus Standards for Hospital Care—Guidelines for Consumer-Focused Reporting[5]	National Voluntary Consensus Standards for Public Reporting of Patient Safety Event Information[6]
5a. Source and use of reports. ii. Hospital performance reports must appeal to and take into account the needs of each of the following unique stakeholder audiences: a) public/consumers; b) purchasers; c) clinicians and providers; d) policymakers; and e) accreditors/regulators. 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. 1a. Identify the nature and purpose of the report (what it will be about and what is to be accomplished by producing it). 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.) 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 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. 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.) 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. 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. 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. 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[4]	National Voluntary Consensus Standards for Hospital Care—Guidelines for Consumer-Focused Reporting[5]	National Voluntary Consensus Standards for Public Reporting of Patient Safety Event Information[6]
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.) 5c. Verification of results. i. Individual hospital results should be shared with that hospital by the reporting entity in advance of publishing and distributing the results publicly. 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. iii. Reporting entities should establish policies and procedures for mediation with hospitals when individual concerns raised by hospitals about their results cannot be resolved. 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. v. Any self-reported results that are published by the hospital should be distinguished from externally validated/verified results published by the reporting entity. 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. 2b. Establish governance and decision-making rules. 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. 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. 2b. Establish governance and decision-making rules. 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. 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. 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[4]	National Voluntary Consensus Standards for Hospital Care—Guidelines for Consumer-Focused Reporting[5]	National Voluntary Consensus Standards for Public Reporting of Patient Safety Event Information[6]
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. 3a. Provide a brief introduction about healthcare quality. 3b. Explain that quality varies within and across institutions and how the report can be used to make quality comparisons. 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. 3a. Define terms. 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. 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. 3d. Explain how the report can be used to understand patient safety in healthcare organizations or providers. 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[4]	National Voluntary Consensus Standards for Hospital Care—Guidelines for Consumer-Focused Reporting[5]	National Voluntary Consensus Standards for Public Reporting of Patient Safety Event Information[6]
	4a. Because measures inherently have components that affect the way they should be reported, be clear about types of conclusions that can be reached. 4b. In choosing measures to be reported, take into account that the best measures: i. are relevant to the healthcare-related concerns of the consumer audience; ii. demonstrate variation and reflect care that those being measured can impact; and 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. 4b. In choosing measures to be reported, take into account that the best measures: i. are relevant to the healthcare-related concerns of the public; 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 iii. are objective, valid, reliable, methodologically sound, feasible, transparent, verifiable, and represent consensus among stakeholders, including consumers and professionals.
5b. Report generation i. Reports prepared for consumers should include two components: a summary of the measure results and a technical supplement. a) The summary of measure results should include: 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; ii) guidance on how to interpret and use the results as well as the data's limitations; and 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. 5a. Help consumers quickly and easily arrive at correct and meaningful conclusions. 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). 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. 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. 5a. Help consumers to quickly and easily understand each measure and to use the information to aid in decision-making. 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[4]	National Voluntary Consensus Standards for Hospital Care—Guidelines for Consumer-Focused Reporting[5]	National Voluntary Consensus Standards for Public Reporting of Patient Safety Event Information[6]
b) The technical report should include: i) trended annual results for at least three years for each endorsed measure; ii) detailed measure definitions; iii) measure specifications; and iv) risk-adjustment methodologies applied, including limitations of risk adjustment. 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. iii. In presenting comparative results, the following should be taken into account: a) Results for individual hospitals should be presented in comparison with local, regional, and national averages. 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. 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. d) Reports should be published in print and electronic	compelling by including anecdotes or stories to illustrate the meaning of the data. 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). 5b. In presenting comparative quality information: 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; ii. when possible, include benchmarks to provide users a better context for making comparisons and using the information; 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; iv. label indicators using everyday language (not clinical or technical terms); v. ensure that comparisons are accurate and supportable; and vi. whenever possible, limit the use of statistics and terms that are difficult for most consumers to understand. 5c. In presenting data from composite measures: i. where measures are interpretable at the individual measure level, report all measures that comprise the	misleading comparisons. 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. iii. Make presentations of information more vivid and compelling by including anecdotes, stories, or case studies to illustrate the meaning of the data. 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. 5b. Use approaches such as those listed below to present comparative patient safety information. 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. ii. When possible, include context for making comparisons and using the information. 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). iv. Label indicators using everyday language (not clinical or technical terms). v. Ensure that comparisons are reasonable and supportable. vi. Whenever possible, limit the use of statistics and terms that are difficult for most consumers to understand. 5c. Composite measures, if used, should be clinically

A Comprehensive Framework for Hospital Care Performance Evaluation[4]	National Voluntary Consensus Standards for Hospital Care—Guidelines for Consumer-Focused Reporting[5]	National Voluntary Consensus Standards for Public Reporting of Patient Safety Event Information[6]
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). iv. Regarding the sample size for reporting: 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. 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. 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. 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. 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 ii. report results for the composite and for each component measure (at a layer down from the initial composite data display). 5d. In providing contextual information/decision support: i. provide a clear contextual framework as part of the report introduction; ii. make sure that key messages are included in the data display; 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; iv. make information understandable by using everyday words and language; 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 vi. use reasonably current data, and display the dates/period that are covered by the data. 5e. In presenting technical documentation: i. include detailed measure definitions, specifications, and risk-adjustment methods; ii. include resource information such as identification of the measure developer, sources of data, and interpretation guides; and	coherent, actionable, and transparent. i. Explain what a composite is and how it is constructed (in consumer language). 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). 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). iv. Report results for the composite and for each component measure (at a layer down from the initial composite data display). 5d. Provide context for low-frequency events. i. Explain how low-frequency events are identified, collected, and displayed and how patient confidentiality is maintained. ii. Discuss the use of low-frequency events in assessing quality and safety of healthcare provider. iii. Retain and make accessible reports from year to year. In doing so, it would be appropriate to provide information about variation over time. 5e. Provide context for adverse events displayed by rates. i. Explain measures of adverse events that are calculated as rates. ii. Discuss the use of rates in assessing quality and safety of a healthcare provider. 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[4]	National Voluntary Consensus Standards for Hospital Care—Guidelines for Consumer-Focused Reporting[5]	National Voluntary Consensus Standards for Public Reporting of Patient Safety Event Information[6]
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. 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. 5d. Distribution and dissemination of reports. i. The frequency of reports and the data used to prepare them should be as follows: 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). b) The most recent data published should be no more than two years old. c) An aggregate mean and comparison for each composite measure should be reported to the public. 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. e) As new measures are added or existing measures are	iii. provide details about methodology.	about variation over time. 5f. In providing contextual information/decision support: i. provide a clear contextual framework as part of the report introduction; ii. make sure that key messages are included in the data display; iii. make clear that reports of low-frequency/rare events are different from rates—distinguish between appropriate uses of different kinds of data; 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; v. make information understandable by using everyday words and language; 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); vii. use most current data available, and display the dates/period that are covered by the data; viii. provide context of comparison to peers, to self over time, and to optimum performance (policy goals); and ix. clearly explain risk stratification, that is, where it is done, why it is important. 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[4]	National Voluntary Consensus Standards for Hospital Care—Guidelines for Consumer-Focused Reporting[5]	National Voluntary Consensus Standards for Public Reporting of Patient Safety Event Information[6]
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. 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. ii. Describe verifiability of the data (if any) through audits, reviews, cross-checking with other data sources, or attestation by the provider. iii. Define data sources, quality control, and the data collection process. iv. Explain whether data are collected as part of a legal or accreditation mandate, or on a voluntary basis. v. Include resource information, when available, such as identification of the measure developer, sources of data, and interpretation guides. 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: 6a. organize information in a way that lets users know what is available and lets them make their own choices; 6b. provide an engaging format and include intuitive and consistent navigation tools that are placed in consistent locations; 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; 6d. seek feedback and test the design and navigation with the intended audiences; and 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. 6a. organize information in a way that lets users know what is available and lets them make their own choices; 6b. provide an engaging format and include intuitive and consistent navigation tools that are placed in consistent locations; 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; 6d. seek feedback and test the design and navigation with the intended audiences; 6e. provide users a way to print the information in

A Comprehensive Framework for Hospital Care Performance Evaluation[4]	National Voluntary Consensus Standards for Hospital Care—Guidelines for Consumer-Focused Reporting[5]	National Voluntary Consensus Standards for Public Reporting of Patient Safety Event Information[6]
	understandable and usable formats.	understandable and usable formats; 6f. make it easy to locate/access ancillary information (in a contextually relevant way); and 6g. encourage consumer interaction through an easy-to-use comment feature (e.g., e-mails, FAQs, etc.).
5e. Consumer research. The following areas should become priorities for research, development, and further investigation to fully inform the improvement of approaches to consumer reporting: 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); ii. the most appropriate, evaluable approaches and formats for presenting reports to consumers; iii. the most appropriate method/s of developing composite results for consumer reporting; and iv. the effectiveness of reporting comparative results to consumers.	7. Regularly review and assess reports to ensure their effectiveness, usability, and currency 7a. Conduct assessments of the use and impact of reports. 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. 7c. Involve stakeholders in revisions and seek their feedback after the report undergoes significant changes. 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. 7a. Define the intended impact of the report, and measure usage/penetration and impact against that goal. 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. 7c. Involve stakeholders in revisions and seek their feedback after the report undergoes significant changes. 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
Senior Vice President, 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

Judith Hibbard, DrPH
Professor, Dept. of Planning, Public Policy and Management, University of Oregon, Eugene

Dana Safran, ScD
Senior Vice President/Performance Measurement and Improvement, Blue Cross Blue Shield of Massachusetts, Boston, MA

Patrick Romano, MD, MPH
Professor of Medicine and Pediatrics, University of California Davis, Division of General Medicine, Sacramento

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

NATIONAL QUALITY FORUM STAFF

Helen Burstin, MD, MPH
Senior Vice President, Performance Measures

Heidi Bossley, MSN, MBA
Vice President, Performance Measures

Karen Pace, PhD, MSN
Senior Director, Performance Measures

Gene Cunningham, MS
Project Manager, Performance Measures