NATIONAL QUALITY FORUM

National Quality Forum

Patient Reported Outcomes (PROs) in Performance Measurement

Draft Report for Comment

10/24/12

12/07/12 Modifications Based on Comments and Edits and TEP Discussion

Table of Contents

National Quality Forum
Patient Reported Outcomes (PROs) in Performance Measurement

INTRODUCTION

US Healthcare: Performance Improvement and Accountability

Widespread variation in the quality of healthcare in the United States is well documented.[1,2,3,4,5,6] Although many laudable examples can be identified across the country where safe, effective, affordable care and support services are consistently provided, serious gaps persist. Coupled with the need to constrain escalating costs of healthcare—threatening the livelihoods of individuals and families and the overall national economy—is the need to improve performance and hold providers accountable. The Patient Protection and Affordable Care Act of 2010 (hereafter, ACA) has several provisions targeting this challenge. One mandates creation of a National Quality Strategy (NQS) to serve as a blueprint to *improve the delivery of healthcare services, patient health outcomes, and population health.*[7] Released in March 2011 and updated yearly, the NQS identifies three overarching aims of better care, healthy people and communities, and affordable care; it also spells out six priority areas for collective action to drive toward a high-value health system (health and well-being, prevention and treatment of leading causes of mortality, person- and family-centered care, patient safety, effective communication and care coordination, and affordable care).[8,9]

Achieving Performance Improvement & Accountability through Patient-Reported Outcomes

Patient and family engagement is increasingly acknowledged as a key component of a comprehensive strategy, including performance improvement and accountability, to achieve a high quality, affordable health system. Emerging evidence affirms that patients who are engaged in their care tend to experience better outcomes[10] and choose less costly but effective interventions, such as physical therapy for low back pain, after participating in a process of shared decisionmaking.[11] Promising approaches to involve patients and their families at multiple levels are being implemented across the country; activities include serving on governance boards at hospitals and contributing to system and practice redesign to make care safer and more patient-centric.[12,13]

Historically, with the exception of collecting feedback on satisfaction or experience with care, patients remain an untapped resource in assessing the quality of health services and of long-term support services. Patients are a valuable and, arguably, the authoritative source of information on outcomes beyond experience with care. These include health-related quality of life, functional status, symptom and symptom burden, and health-related behaviors. For example, in the case of long-term support services for persons with disabilities, asking about valued outcomes such as increased communication and self-help skills and improved social interactions is crucial. Hence, two critical steps are to engage patients by building capacity and infrastructure to capture patient-reported outcomes routinely and then to use these data to develop performance measures to allow for accurate appraisals of quality and efficiency.

NQF Role in Promoting Accountability and Performance Improvement

Valid, reliable measures are foundational for evaluating and monitoring performance and fostering accountability. The National Quality Forum (NQF) is a voluntary consensus standards-setting organization as defined by the National Technology Transfer and Advancement Act.[14] In this role, NQF endorses performance measures as consensus standards to assess the quality of healthcare for use in accountability applications such as public reporting and payment as well as performance improvement. NQF is a neutral evaluator of performance measures, not a measure developer. NQF convenes diverse stakeholders to evaluate measures based on well-vetted criteria (available here).

The field of performance measurement is evolving to meet the demands of increased accountability to improve outcomes in both quality and costs. The direction for NQF-endorsed performance measures includes:

- driving toward higher performance reflected in more outcome measures rather than in basic processes such as performing an assessment;
- measuring disparities;
- shifting toward composite measures that summarize multiple aspects of care;
- harmonizing measures across sites and providers; and
- conducting measurement across longitudinal patient-focused episodes including outcome measures, process measures with direct evidence of impact on desired outcomes; appropriateness measures; and cost/resource use measures coupled with quality measures, including overuse.

Figure 1 depicts the relationship among structure, process, and outcome. For NQF endorsement, the hierarchical preference is for performance measures of health outcomes that are linked directly to evidence-based processes or structures or of outcomes of substantial importance with a plausible link to healthcare processes. Next in the preferred hierarchy are measures of intermediate outcomes and processes closely linked to desired outcomes. Measures of processes that are distal to desired outcomes (e.g., assess a patient clinical parameter) and those that are satisfied by a "checkbox" are considered to have the least impact on the goal of improving healthcare and health.

Figure 1. Structure-Process-Outcome

Patient-Reported Outcomes Tools & Performance Measures

Patient-reported outcomes (PROs) are defined as "any report of the status of a patient's (or person's) health condition, health behavior, or experience with healthcare that comes directly from the patient, without interpretation of the patient's response by a clinician or anyone else."[15] "PRO" has become an international term of art; the word "patient" is intended to be inclusive of all persons, including patients, families, caregivers, and consumers more broadly. It is intended as well to cover all persons receiving supportive services, such as those with disabilities. Key PRO domains include:

- Health-related quality of life including functional status;
- Symptoms and symptom burden;
- Experience with care; and
- Health-related behaviors.

Various tools (e.g., instruments, scales, single-item measures) that enable researchers, administrators, or others to assess patient-reported health status for physical, mental, and social well-being are referred to as PRO measures (PROMs). To include PROs more systematically as an essential component of assessing the quality of care or services provided, and as part of accountability programs such a value-based purchasing or public reporting, distinguishing between PROMs (i.e., tools) and aggregate-level performance measures is important.

A PRO-based performance measure (or PRO-PM) is based on PRO data aggregated for an entity deemed as accountable for the quality of care or services delivered. Such entities can include (but would not be limited to) supportive services providers, hospitals, physician practices, or accountable care organizations (ACOs). NQF endorses performance measures (PRO-PMs) for purposes of performance improvement and accountability; NQF does not endorse the tools (PROMs)to measure PROs. Table 1 illustrates the distinctions among PRO, PROM, and PRO-PM. Full definitions are in the glossary (see Appendix A).

Table 1. Distinctions among PRO, PROM, and PRO-PM: Two Examples

Concept	Patients With Clinical Depression	Persons with Intellectual or Developmental Disabilities
PRO (patient-reported outcome)	Symptom: depression	Functional Status-Role: employment
PROM (instrument, tool, single-item measure)	PHQ-9©, a standardized *tool* to assess depression	Single-item measure on National Core Indicators Consumer Survey: *Do you have a job in the community?*
PRO-PM (PRO-based performance measure)	Percentage of patients with diagnosis of major depression or dysthymia and initial PHQ-9 score >9 with a follow-up PHQ-9 score <5 at 6 months (NQF #0711)	The proportion of people with intellectual or developmental disabilities who have a job in the community

PRO-PMs Applications: Benefits and Challenges

Interest and appreciation of the value of using PROMs in performance measurement as part of the broader accountability and performance improvement landscape are mounting. To accelerate the adoption of PROMS to PRO-PMs that can be used for these purposes several underlying issues ~~two challenges~~ must be addressed, which will require collaborative and collective effort across multiple stakeholder groups including providers, consumers, purchasers, measure developers, researchers, and others. First, PROMs have not been widely adopted for clinical use outside research settings in the United States; for that reason, they may be unfamiliar to many health professionals, payers, and provider institutions. Therefore, awareness must be raised of the benefits of using PROMs and engaging patients in their care and the relationship to improved outcomes. Second, there are several method-related challenges such as ~~for~~ aggregating patient data on PROMs to measure performance at multiple levels of analysis (e.g., individual, group practice, organization) and use of proxy respondents. Therefore, more research is needed on best practices in this area.

To begin to address these complex issues, NQF, with funding from the Department of Health and Human Services, is conducting the *PROs in Performance Measurement* project. The project goals are to:

- Identify key characteristics for selecting PROMs to be used in PRO-PMs;
- Identify any unique considerations for evaluating PRO-PMs for NQF endorsement and use in accountability or performance improvement applications; and
- Lay out the pathway to move from PROMs to NQF-endorsed PRO-PMs.

NQF designed this project to bring together a diverse set of stakeholders (see Appendix B) who could facilitate the groundwork for developing, testing, endorsing and implementing PRO-PMs. ~~Those stakeholders included researchers, health professionals, performance measure developers, and consumer and purchaser representatives (see Appendix A).~~ Key steps in the project were to convene two workshops with an expert panel and to commission two papers. The papers focused on issues about methods and served as background for the workshops. The first paper focused on selecting PROMs for use in performance measurement and the second on the reliability and validity of PRO-PMs (papers available here).

National and international examples of successful experiences we can build on are encouraging. At the workshop, participants obtained valuable insights about approaches to data collection and aggregation and practical pointers about implementation (e.g., getting buy-in from providers). At the first workshop, colleagues from the Dartmouth Spine Institute and Massachusetts General Hospital presented information about their experiences using PROMs in patient care and performance improvement (available here). At the second workshop, representatives from the Centers for Medicare & Medicaid Services Health Outcomes Survey, England's National Health Service PROMs, and Sweden's national quality registers presented on their initiatives to report PRO-PMs publicly (available here). These discussions informed the recommendations found later in this report. Additionally, a large body of knowledge is available about experience with care measures as PRO-PMS (e.g., performance measures based on CAHPS®).

This report captures the insights from this effort to date and provides recommendations to move the field of performance measurement forward. The remaining sections of this report cover guiding principles, a detailed pathway from PROs to PRO-PMs, key implications and recommendations related to NQF endorsement criteria, and future directions.

GUIDING PRINCIPLES

During the first workshop, members of the Expert Panel discussed key characteristics for identifying PROMs most suitable for developing and testing PRO-PMs. They conceptualized these ideas as "guiding principles" for using PROMs in the context of performance measurement. They are not NQF endorsement criteria per se, but they serve as key constructs for recommendations on the pathway from PRO to PRO-PM ~~and related NQF endorsement criteria~~. PROM developers and PRO-PM measure stewards should also take these principles into account in preparing submissions and documentation for NQF consideration for endorsement.

The guiding principles, described below, place the patient front and foremost. They underpin the thinking that shaped the pathway from PROs to PRO-PMs discussed in the next section of this report. The word "patient" is often used as shorthand to comprise patients, families, caregivers, and consumers more broadly. We also use this term include persons receiving supportive services, such as those with disabilities. Moving forward, NQF must ensure that the emerging portfolio of PRO-PMs addresses a range of healthcare services that extend beyond the walls of a particular clinical setting of care.

The five guiding principles encompass the following: meeting technical psychometric standards; being person-centered; having meaning to individuals responding to PROMs; being actionable; and being implementable.

Psychometric Soundness

Workshop participants agreed on several psychometric properties as a baseline set of requirements to be considered in selecting PROMs for use in PRO-PMs. These are listed in Box 1 and are derived from the first commissioned paper. Appendix C provides the expanded explanations for these scientific properties of instruments or tools to measure PROs. The remaining three sets of principles below presume that the main elements of reliable, valid, responsive and feasible PROMs are adequately covered and demonstrated.

Box 1. Characteristics for Selecting PROMs Identified in Commissioned Paper

1. Conceptual and Measurement Model Documented
2. Reliability
   - 2a. *Internal consistency (multi-item scales)*
   - 2b. *Reproducibility (stability over time)*
3. Validity
   - *3a. Content Validity*
   - *3b. Construct and Criterion-related Validity*
   - *3c. Responsiveness*
4. Interpretability of Scores
5. Burden
6. Alternatives modes and methods of administration
7. Cultural and language adaptations
8. Electronic health record (EHR) capability

Person-Centered

"Person-centeredness" was the overarching theme that arose from the workshop discussions. In this context, using PROMs is viewed as an important step toward engaging patients, health professionals,

and other entities in creating a person-centered health system. Workshop participants also identified the opportunity for PROMs to facilitate shared decisionmaking (SDM), another strategy for engaging patients. SDM is defined as *a collaborative process that allows patients and their providers to make healthcare decisions together, taking into account the best scientific evidence available, as well as the patient's values and preferences*.[16] For SDM, clinicians and other healthcare staff can use the instrument, scale, or single-item measure (PROM) to engage patients in their own preferred self-management and goal attainment by identifying outcomes important to them and tracking change over time. An important caveat to this discussion is that not all patients want to engage in formal SDM activities. Therefore, although contributing to SDM efforts is desirable, not all PROMs need to enable SDM. Measures of decision quality, defined as *the match between the chosen option and the features that matter most to the informed patient*, fall under the PRO domain of experience with care and take into account whether patients are informed with the best available evidence and there is concordance between what matters most to them and the treatments they receive.[17,18]

Formatted: Not Highlight

As a final consideration of person-centeredness, as patients become more engaged in their care by providing systematic feedback on outcomes such as their functional or health status, the flow of information between clinicians and patients must be bi-directional. This may mean that health professionals interpret PROM information back to their patients; it may mean that mechanisms are established to give patients their own information directly (displayed in easy-to-understand ways). With steps such as these, respondents to PROMs can benefit from seeing results in a timely way, and this type of service can balance any perceived burdens they may feel about completing data-collection activities. Although these considerations may not affect NQF endorsement efforts directly, the Expert Panel wished to emphasize that having PRO-PMs that can be used in this manner is desirable.

Meaningful

Closely intertwined with person-centeredness is the concept of "meaningfulness." Meaningfulness encompasses the relevance and degree of importance of the concepts measured by the PROM from the perspective of patients, their families, and caregivers as well as clinicians and other health professionals who serve them. Among the concepts that PROMs would ideally capture are the following: the impact of health-related quality of life (including functional status); symptom and symptom burden; experience with care and satisfaction with the services; perceived utility of the services for achieving personal goals; or health-related behaviors. As suggested above, the focus comprises both "traditional" healthcare services broadly defined and supportive services for persons with disabilities.

Workshop participants debated how best to demonstrate ~~evidence~~ that stakeholders think a particular PROM is meaningful. The following framework, coined as the three "Cs," can serve as a starting point for thinking about how to operationalize this construct:

- Conceptual: ~~The first step is~~ engaging people in the dialogue about what matters most to them to define the concepts that PROs should cover. This upstream interaction is critical to meet a threshold consideration of what is being measured is important and meaningful to the individual.

- Contextual: ~~The second step is~~ learning how individuals use the information derived from either a PROM or a PRO-PM. Individuals here are defined very broadly to include not just patients (however construed for the application at hand) but also clinicians, other health professionals, administrators, and perhaps even policymakers. For example, does such information facilitate their participation managing their own healthcare? Does it help people to select a high-quality provider of health or supportive services? Do such data contribute to the discourse on larger social issues such as achieving high-quality care at acceptable costs?

- Consequential: ~~The third step is~~ determining what happens when PRO-PM information is used in accountability ~~programs~~ applications (e.g., public reporting, value-based purchasing) or performance improvement. ~~This step also needs to consider whether the PRO-PM is consequential to the individual or family member.~~ Performance data on PRO-PMs can have important consequences on the availability and receipt of quality health services, the type of services, and their responsiveness to individuals' needs.

~~Actionable~~Amenable to Change

~~"Actionability"~~ Amenable to change refers to evidence that the outcome of interest ~~(i.e., PRO)~~ is responsive to a specific health service or intervention and thus considered to actionable. This concept applies equally to PROs or other types of outcomes. ~~The guiding principle of actionability is that~~ The reasoning is that outcome performance measures (~~i.e.,~~ including PRO-PMs) intended for both accountability and improvement should be supported by evidence that the healthcare providers being evaluated can influence the person's short- or long-term outcomes. ~~The position held by the majority of workshop participants was that~~Therfore, without such evidence, ~~a PRO-PM~~ the performance measure would ~~is~~ not be considered a valid indicator of quality of care.

However, a unique aspect of PROs is that they require patient effort to provide the PROM data. As PROs start to be collected more routinely at the point of care and are embedded into workflows it becomes essential to ensure this information is of value to the patient and perceived as actionable from their vantage point. Analogous to the collection of a blood sample to measure glucose concentration over time for diabetes (e.g., HbA1c) results should be shared and appropriate intervention (or not) taken based on the best available evidence and informed by patient's preferences and treatment goals. When collecting individual level data through the use of PROMs special consideration must be given to the burden of data collection which ideally will be offset by the patient's assessment of actionability and meaningfulness.

~~From the workshop discussions emerged a spectrum of actionability for identifying PROs with high leverage to accelerate on the path to PRO-PMs. This continuum had three levels:~~

~~Highly actionable: evidence that the PRO is responsive to intervention as demonstrated in clinical studies and that the intervention has been implemented in practice. Initial efforts for developing PRO-PMs should be focused here.Moderately actionable: evidence of responsiveness to intervention in clinical studies but only limited experience with the intervention in practice.~~

~~Moderately actionable PROs can be used for accountability but with caution. This is the next tier for consideration of accountability and performance measurement.~~

~~Weakly or not actionable: evidence of responsiveness to intervention is weak in clinical studies and the intervention has not been implemented in practice. These PROs should not be considered for accountability or performance improvement purposes at this time (and thus not for NQF endorsement of PRO-PMs).~~

There was robust discussion among the workshop participants on this proposed guiding principle during which a ~~Some workshop participants offered a~~ strong counter argument was aired in respect to the idea that all PROs considered for purpose of accountability or performance measurement must be actionable. The~~ir~~ rationale was that some outcomes are worth measuring even though they may not, at this point in time, be amenable to change by providers. For example, there are some outcomes (e.g., time to recovery) that are meaningful to patients that may not currently be considered modifiable, but provide valuable information to ~~because~~ patients and help them, ~~and clinicians~~working in close relationship with their care provider, ~~can use them~~ to make informed decisions. Additionally, as in this case, if the outcome is deemed of high importance to patients the process of measuring and reporting could identify variation in performance and facilitate the spread of efective interventions. Although this disagreement was not resolved at the workshop, the point merits exploration.

Implementable

The guiding principle that a PROM should be "implementable" acknowledges that many diverse factors affect practical use of them in quality or accountability programs. Most of these factors relate to barriers to adopting such tools (PROMs) or collecting data and reporting on PRO-PMs in many practices, institutions, or other settings. Workshop participants raised many implementation issues. Although the examples were not exhaustive, workshop participants emphasized issues on the following list: administering PROMs in real-world situations; addressing literacy and health literacy of respondents; addressing cultural competency of clinicians and other service providers; dealing with the potential for unintended consequences related to patient selection; covering costs associated with using PROMS (especially those not available in the public domain); and adapting PROMs to computer-based platforms or other alternate formats.

PATHWAY FROM PRO TO NQF-ENDORSED PRO-PM

The pathway displayed in Figure 2, and described in detail below lays out the critical steps in developing a PRO-based performance measure suitable for endorsement by NQF. It begins with the conceptual basis for identifying a PRO for performance measurement; the pathway then proceeds through selecting a PROM and developing and testing a performance measure to achieving NQF endorsement of a PRO-PM and using the performance measure for accountability and performance improvement. This pathway ~~does not replace the existing NQF measure evaluation criteria; rather~~ ~~it~~ describes how a PROM may form the basis of a PRO-PM that NQF could eventually endorse based on the NQF criteria. ~~The existing NQF criteria *are* applicable to PRO-PMs as well as the PROM used in the performance measure.~~

~~Some recommendations for minor modifications to the NQF endorsement criteria to address the unique considerations of PRO-PMs are discussed in the next section.~~

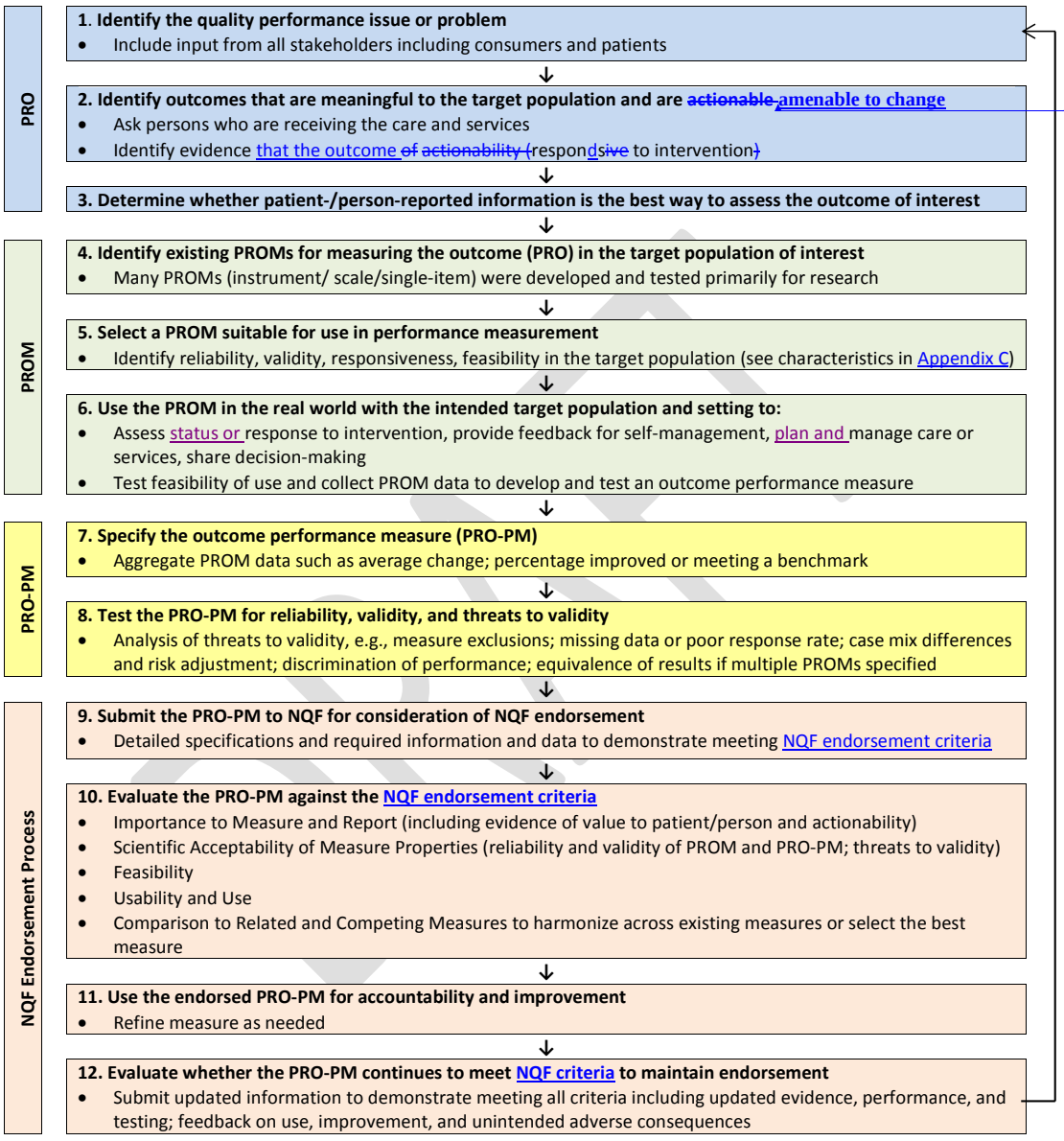
The quality performance measurement enterprise includes multiple stakeholders who collaborate to develop performance measures, including methodologists and statisticians, as well as those receiving care and services, those whose performance will be measured, and those who will use performance results. In this discussion, the reference to developers includes all the participants in developing performance measures, not just formal measure developer organizations.

Although NQF is involved in the last section of the pathway, the earlier steps have implications for whether a performance measure will be suitable for NQF endorsement. Thus, they are intended to serve as a guide and best practices to help ensure that PRO-PMs will meet NQF criteria. For example, steps 1 and 2 in the pathway indicate that patients (broadly defined as above) should be involved in identifying quality issues and outcomes that are meaningful to those receiving the care and supportive services. If patients are involved at those steps, then developers will have amassed the information needed to demonstrate that the outcome is of value to patients. In the context of using this pathway leading to an NQF-endorsed performance measure, step 2 also suggests that developers identify outcomes with evidence that the outcome is responsive to intervention.

Our first recommendation is stated in the box below. The steps shown in Figure 2 and described below are intended to help ensure that a proposed performance measure will meet NQF criteria for endorsement.

Recommendation 1.
Those developing PRO-PMs to be considered for NQF endorsement should follow the basic steps shown in the pathway in Figure 2. Doing so will help ensure that the eventual PRO-PM and its supporting documentation conform to NQF endorsement criteria.

299 Figure 2. Pathway from PRO to NQF-endorsed PRO-PM



Formatted: Font: Bold

Pathway Section Related to the PRO

The pathway begins with the conceptual basis for identifying a PRO for performance measurement.

1. Identify the quality performance issue or problem.

Before developers devote resources to performance measurement, they need a clear understanding of the quality performance issue or problem related to healthcare or supportive services for a target population. Such understanding will direct the focus and establish the need for a performance measure. Input from all stakeholders including the recipients of the care and services, providers whose performance will be measured, payers, purchasers, and policymakers is critical to identifying priorities for performance measurement.

2. Identify outcomes that are meaningful to the target population and are actionable by providers of care and services.

After developers articulate the quality performance issue, they should identify the specific outcomes that are valued and meaningful to the target population in the context of a specific healthcare or supportive service. That is, the people receiving the healthcare or supportive services should be asked for their input. At this stage, all relevant desired outcomes should be identified even if they might not be assessed through patient-reported data.

As noted previously, the Expert Panel ~~suggested~~ discussed focusing performance measures on outcomes that are ~~actionable, i.e.,~~ responsive to intervention by healthcare and service providers. The reason for this is twofold: 1) so patients are only asked to provide PROM data that is directly applicable to their care and treatment, and 2) so that providers' performance is measured on outcomes influenced by the care they provide. While there may be reasons to measure performance on important outcomes without such evidence, ~~Therefore,~~ outcomes with evidence that they are influenced by at least one structure, process, intervention, or service ~~are preferred~~should be considered as a starting point to garner broad-based support.

3. Determine whether patient- or person-reported information is the best way to assess the outcome of interest.

Patient- or person-reported data are not necessarily the best way to assess every desired outcome identified in the prior step. The domains of health-related quality of life including functional status, symptoms and symptom burden, and health-related behaviors are outcomes for which individuals receiving healthcare and services may be the best or only source of information. However, other meaningful outcomes such as survival (or mortality) and hospital readmission could be assessed using other data sources.

Pathway Section Related to the PROM

Given that one or more PROs are identified in the above steps, the pathway addresses the steps that organizations should take to select a PROM suitable to use in a performance measure.

4. Identify existing PROMs for measuring the outcome (PRO) in the target population of interest.

As ~~M~~many PROMs already exist, developers should use various strategies (e.g., literature searches, PROMIS, web searches, outreach to experts in the field) to search for and identify PROMs that measure the outcome of interest in the target population. PROMs that were developed years ago may not have benefited from patient input; therefore, including patients in selecting PROMs to be used in performance measures is important.

5. Select a PROM suitable for use in performance measurement.

The scientific (psychometric) characteristics that organizations should examine in selecting a PROM for performance measurement were summarized above and appear in detail in Appendix C. Of great importance is that PROMs be reliable, valid, and responsive in the target population. If no PROM for the target population seems to be suitable for use in a performance measure, then a developer or research group should test one or more PROMs in the target population or develop and test a wholly new PROM before a performance measure can be developed. The commissioned paper on methods issues related to PROMs is a resource on considerations for selecting PROMs (available here).

6. Use the PROM in the real world with the intended target population and in the intended setting.

The Expert Panel agreed that developers should collaborate with providers to use PROMs with the target population and in the settings for which performance measures are proposed before developing a PRO-PM. Many PROMs were developed for research studies and the resources and protocols for administering PROMs may not be realistic for broad scale implementation. ~~This~~ A real-world application will identify feasibility issues related to administration, data capture, and workflow to use the PROM to assess individuals' responses to healthcare or supportive services intervention, provide feedback for self-management, and (as desired) facilitate shared decisionmaking. At the first workshop, representatives from Dartmouth Spine Center and Partners Healthcare presented their experiences with using PROMs in clinical practice (available here).

Formatted: No underline, Font color: Auto

Formatted: No underline, Font color: Auto

Formatted: No underline, Font color: Auto

Formatted: No underline, Font color: Auto

Actual use of the PROM also generates the data needed to determine the best way to aggregate the PROM data in aperformance measure and test the PRO-PM for reliability and validity. Widespread implementation is not a prerequisite for NQF endorsement; however, testing for reliability and validity and addressing risk adjustment are required. The data for such testing could come from settings that have already implemented ~~T~~the PROM, ~~could be used in~~ a pilot study, or a broader demonstration. ~~or through more broad-based adoption. This step does not require a~~An endorsed performance measure focused on administering the PROM is not a necessary prerequisite and could divert resources and slow the endorsement of PRO-PMs. Performance measures focused on such assessments may not meet NQF criteria for endorsement and is discussed under recommendations realted to the NQF evalaution criteria.~~However, in some circumstances, adding steps for such a process measure may be considered and is discussed after the main pathway.~~

Pathway Section Related to the PRO-PM

After the developer has selected the PROM and collaborated with providers who used it in practice to generate sufficient data for testing, the pathway addresses how developers should specify and test a PRO-PM.

7. Specify the outcome performance measure (PRO-PM).

Developers specify how the outcome performance measure will be constructed. The metrics may be, for instance, an average change, the percentage of patients improved, or the percentage of respondents meeting a specific benchmark value. The performance measure needs to be fully specified including the specific PROM, guidance for administering it, and rules for scoring; it should also describe the target population and any exclusions, give time frames for PROM administration and performance measurement, and outline any needed risk adjustment procedures.

8. Test the PRO-PM for reliability, validity, and threats to validity.

Developers need to test the performance measure for reliability and validity. They explicitly need to address a variety of threats to validity or other technical issues. These include the need for risk adjustment or stratification and options for doing this, appropriateness of potential exclusions, and options for dealing with missing data. A further challenge is explaining the level of equivalence of results when multiple PROMs are used.

Testing the PRO-PM is distinct from testing the PROM. Using a PROM with sound psychometric properties is necessary but not sufficient to assure a reliable and valid PRO-PM. The commissioned paper on methods issues for PRO-PMs provides a resource on considerations and approaches to examining or demonstrating reliability and validity of the performance measure (available here).

Pathway Section Related to the NQF Endorsement Process

The last section of the pathway focuses on the NQF endorsement process.

9. Submit the PRO-PM to NQF for consideration of NQF endorsement.

The NQF endorsement process begins when developers submit a measure to NQF for consideration. Developers submit required information in NQF's standard form so that all the information needed to evaluate the measure is available to reviewers.

10. Evaluate the PRO-PM against the NQF Endorsement Criteria.

NQF evaluates measures against four main endorsement criteria listed here and described and discussed in more detail below.

1. Importance to Measure and Report
2. Scientific Acceptability of Measure Properties
3. Feasibility
4. Usability and Use

In addition, NQF has criteria and processes to address measure harmonization and selection of the best measure from among competing measures, which also would apply to PRO-PMs.

11. Use the endorsed PRO-PM for accountability and improvement.

Once endorsed, NQF expects the measure to be used for accountability and performance improvement applications. Implementation of the performance measure should facilitate the goal of improvement and allow for measuring and tracking performance. Use of the performance measure provides data on performance to be examined for intended and unintended consequences.

In the case of PRO-PMs initially endorsed with testing based on limited PROM data, implementation of the PRO-PM could be phased. The initial emphasis would be on collecting the PROM data to expand testing and refine the measure before reporting performance on the outcome.

Formatted: Not Highlight

12. Evaluate whether the PRO-PM continues to meet NQF Criteria to maintain endorsement.

NQF reviews each endorsed measure every three years to evaluate whether it continues to meet NQF criteria. In making its decision at this stage, NQF evaluates the measure on all criteria and considers information on actual use, improvement, and unintended adverse consequences. This information and ~~results of~~ the NQF endorsement maintenance decision also provide feedback to developers who are~~the beginning of the pathway and considerations for performance measurement~~ considering developing performance measures based on PROs.

~~Alternate Pathway~~

~~The main pathway depicted in Figure 2 and discussed above focuses on moving from a PRO to a PRO-PM. The core construct is an outcome that is meaningful to patients (broadly defined) and measured by a PROM that meets other desirable characteristics discussed in the guiding principles above. In some circumstances, however, beginning to measure performance related to the administration and data capture of the PROM itself may be considered before moving straight to using the PRO data themselves. Ultimately, however, the goal is to produce and use outcome performance measures.~~

~~The primary purposes of a "process performance measure" focused on administration~~ a PROM ~~are to facilitate use of the PROM as described in step 6 of the main pathway and to prepare the field for outcome performance measurement. Another potential reason for a process performance measure is that ptients and providers agree concern that although the PRO is valued, but may not think that it is not currently thought to be influenced by health care or other services currently can influence it – but could be in the future. However, in this case, the PRO may not even be a priority for performance measurement as indicated in step 2.~~

~~The alternate pathway entails developing, testing, endorsing, and implementing such a process measure before developing the outcome measure. Therefore, it has implications for time and resources. Some questions to consider before pursuing the additional steps related to a process performance measure include:~~

- ~~*Is another mechanism in place to facilitate use of a PROM?* If use of a PROM is achieved through other requirements, such as regulations, accreditation, or accepted guidelines, then a process performance measure may not appreciably influence the extent of use.~~
- ~~*Will the process performance measure yield the data needed to develop and test an outcome performance measure?* The process performance measure should be specified so that it requires that providers administer a specific PROM is administered at designated intervals and , with recording ofrecord the PROM value in the health record, — not merely checking off that it was administered. Alternatively, administering a PROM could be included in a more substantive process measure focused on an evidence-based intervention in response to a specific value of a the PROM could be constructed so that use of the PROM is required (e.g., a process measure that requires that a PROM on pain is administered to patients with cancer and pain treatment provided based on the response to the PROM).~~
- ~~*Does a credible plan exist to implement the process performance measure and collect data?* If the process performance measure is not implemented in a way that makes providers accountable for performance on using the PROM and capturing PROM data, it is less likely to affect adoption of the PROM and advance development of an outcome measure.~~

Formatted: No underline, Font color: Auto

Formatted: No underline, Font color: Auto

Formatted: No underline, Font color: Auto

Formatted: No underline, Font color: Auto

KEY IMPLICATIONS AND RECOMMENDATIONS RELATED TO NQF CRITERIA

Overview

The NQF endorsement criteria and guidance on evaluating all performance measures also apply to PRO-PMs. The four main endorsement criteria are: importance to measure and report, scientific acceptability of measure properties, feasibility, and usability and use. NQF committee members use the criteria to evaluate measures submitted for potential endorsement. When the performance measure meets the relevant criteria and NQF endorses a measure, it is considered suitable for purposes of accountability and performance improvement. Potential submitters (i.e., developers) need to be very familiar with the NQF criteria so as to be able assemble the required documentation as part of their submission.

The TEP's exploration of PRO-PMs highlighted some issues that also are relevant to other performance measures: PRO-PMs may, however, have ~~some special or even~~ unique aspects that warrant special consideration for measure evaluation. Some of the panel's recommendations consider therefore must be considered in the larger context of NQF endorsement criteria for all measures, specifically the evidence criterion. The TEP agreed that PRO-PMs should be held to the same criteria as other performance measures and recommended that NQF revise some criteria for all performance measures including PRO-PMs. **The same standards would be applied to PRO-PMS as any other outcome measures.**

Table 2 lists these factors~~,~~ in the context of the main NQF endorsement criteria. The left column provides an abbreviated description of each criterion. The middle column identifies special considerations that PRO-PMS bring to light, but they are not unique to PRO-PMs. Several unique aspects about PRO-PMs are identified in the right column~~; they may warrant some modifications to the~~

~~NQF criteria to ensure they are suitable for endorsement~~. This section provides recommendations and rationales for modifying the NQF criteria or guidance.

Table 2. NQF Endorsement Criteria and their Application to PRO-PMs

Abbreviated NQF Endorsement Criteria	Special Considerations For Evaluating PRO-PMs that are relevant to other performance measures	Unique Considerations for Evaluating PRO-PMs
Importance to Measure and Report a. High impact b. Opportunity for improvement c. Health outcome OR evidence-based process or structure of care	• Evidence supports that the outcome is responsive to intervention. • Evidence exception for performance measures focused soley on conducting an assessment (e.g., administering a PROM, lab test)	• Patients/persons must be involved in identifying PROs for performance measurement (person-centered; meaningful). • ~~Evidence supports that the PRO is responsive to intervention (actionable).~~
Scientific Acceptability of Measure Properties a. Reliability 1. precise specifications 2. reliability testing for either data elements or performance measure score b. Validity 1. specifications consistent with evidence 2. validity testing for either data elements or performance measure score 3. exclusions 4. risk adjustment 5. identify differences in performance 6. comparability of multiple data sources	• Data collection instruments (tools) should be identified (e.g., specific PROM instrument, scale, or single item) • If multiple data sources (i.e., PROMs, methods, modes, languages) are used, then comparability or equivalency of performance scores should be demonstrated. •	• Specifications should include standard methods, modes, languages of administration; whether (and how) proxy responses are allowed; standard sampling procedures; how missing data are handled; and calculation of response rates to be reported with the performance measure results. • Reliability and validity should be demonstrated for both the data (PROM) and the PRO-PM performance measure score. • Response rates can affect validity and should be addressed in testing. • Differences in individuals' PROM values related to PROM instruments or methods, modes, and languages of administration need to be analyzed and potentially included in risk adjustment.
Feasibility a. Data generated and used in care delivery	• The burdens of data collection, including those related to use of proprietary	• The burden to respondents (people providing the PROM data) should be minimized

Abbreviated NQF Endorsement Criteria	Special Considerations For Evaluating PRO-PMs that are relevant to other performance measures	Unique Considerations for Evaluating PRO-PMs
b. Electronic data c. Data collection strategy can be implemented	PROMs, are minimized and do not outweigh the benefit of performance measurement.	(e.g., availability and accessibility enhanced by multiple languages, methods, modes). • Infrastructure to collect PROM data and integrate into workflow and EHRs, as appropriate.
Usability and Use a. Accountability and transparency b. Improvement c. Benefits outweigh unintended negative consequences	• Adequate demonstration of the criteria specified above supports usability and ultimately the use of PRO-PM for accountability and performance improvement.	

Evidence that the PRO is of Value to the Target Population

Recommendation 2.
The NQF criterion or guidance for importance to measure and report should require evidence that the target population values the measured PRO and finds it meaningful.

Person-centeredness is a key principle for developing PRO-PMs. As shown in Figure 2, identifying outcomes of value to the target population is a critical early step in the pathway to endorse a PRO-PM. NQF's current criteria require evidence that the aspect of care being measured is of value to the patient when measures of experience with care are being evaluated. Experience with care is considered one type of patient-reported outcome; therefore, the requirement for having evidence of the value to the patient needs to be expanded to apply to all patient-reported outcomes.

Evidence that the Measured PRO is Responsive to Intervention

Recommendations 3-~~4~~5.
3. The NQF criterion and guidance regarding evidence should require identification of the causal pathway linking the relevant ~~PRO and healthcare~~ structures;, (processes, interventions, or services); intermediate outcomes; and outcomes ~~(i.e., process-outcome)~~.

4. NQF should consider applying the existing criterion and guidance regarding evidence for a process performance measure to health outcome performance measures, including PRO-PMs – i.e., *a systematic assessment and grading of the quantity, quality, and consistency of the body of empirical evidence* that at least one of the identified healthcare structures, processes, interventions, or services influences the ~~PRO~~outcome.

~~5. NQF should consider adopting this approach for all health outcome performance measures regardless of whether the data are self-reported by patients (or acceptable proxy respondents) or supplied by clinicians.~~

~~Actionability~~ Amenable to change was a key principle identified for developing PRO-PMs; however, the discussion and rationale extended to health outcome measures, in general. The Expert Panel suggested that evidence that the PRO or health outcome is responsive to intervention be required for NQF endorsement ~~of a~~ for all ~~PRO-PM~~outcome performance measures. This represents a departure from NQF's current NQF guidance regarding evidence for performance measures of health outcomes and will require further examination by the CSAC and Board and a plan and timeline for implementation if this approach is recommended by these bodies.

For health outcome measures, NQF requires only a rationale linking the outcome to at least one healthcare structure, process, intervention, or service; it does not require submitting and evaluating information on systematic reviews of the empirical body of evidence, as required for other types of performance measures. NQF's position on evidence for health outcomes is based on the following reasoning:

- Health outcomes such as survival, physical or cognitive function, relief of symptoms, or prevention of morbidity are the reasons for seeking care and the goal of providing care. Therefore, these outcomes are central to measuring the performance of those rendering healthcare or supportive services.
- Health outcomes are often integrative. As such, they may reflect the influence of multiple clinicians and care processes and therefore are based on multiple bodies of evidence. Submitting information on multiple bodies of evidence could be burdensome and a disincentive for submitting outcome performance measures for NQF endorsement.
- Measuring health outcomes to identify variability in performance is a key driver to identifying strategies for improvement, even for outcomes previously thought not to be modifiable such as central line-associated bloodstream infections.

As discussed under the guiding principles, these same rationale were identified for PROs. However, ~~T~~the current environment in which penalties may be associated with performance measure scores has increased concern about using outcome performance measures for accountability. To mitigate that concern to some extent, the Expert Panel suggested focusing performance measurement on PROs that are meaningful to patients *and* with evidence that they are responsive to intervention. England and Sweden are leaders in the area of measuring PROs for performance measurement and appear to have taken this approach. England measures and reports performance on PROMs focused on specific surgical procedures to ameliorate problems with function and symptoms-hip and knee replacement and varicose vein surgery (access reports here). Sweden measures and reports performance on PROMs related to surgical procedure outcomes and complications (access report here). Sweden also reports performance on PROMs for a few medical conditions such as functioning three months after a patient has suffered a stroke and improvement after patients have started biological drug therapy for rheumatoid arthritis.

The Expert Panel acknowledged the trade-offs to a condition-specific approach. First, it excludes much of the population receiving healthcare and supportive services. Second, even for a specific condition, limiting performance measurement to those who received only one possible intervention (e.g., surgery) does not provide a complete picture of performance related to the condition. A related question is whether to measure the PRO with generic or condition-specific PROMs. Condition-specific PROMs may be more responsive to change. However, generic measures offer more breadth, which is relevant, given that many patients have more than one condition. Using both generic and condition-specific PROMs affords the opportunity to better understand the benefits and drawbacks of both. These issues will need to be considered and revisited as the field gains experience with PRO-PMs.

Evidence Exception for a Performance Measure Focused on Administering a PROM

Recognizing the additional complexity of PRO-PMs, the TEP acknowledged that developing an outcome performance measure may not be immediately possible and that some flexibility to accept a performance measure focused on administering a PROM may be needed. However, an outcome measure is the goal and such a process performance measure should only be considered in an exceptional circumstance, and if the proposed process measure clearly specifies that data are collected and includes a credible plan to develop the outcome measure.

Recommendation 5

NQF should provide explicit guidance when a performance measure focused on an assessment, including administering a PROM, meets the exception for the evidence criterion.

In such exceptions, he following additional conditions should be required before it is considered for endorsement.

- The process performance measure is specified so that it requires providers to administer a specific PROM at designated intervals and record the PROM value in the health record, not merely check off that it was administered.
- The developer submits a credible plan to implement the process performance measure, collect data, and develop and test the outcome performance measure.

Formatted: Highlight

Another issue related to evidence is whether a performance measure focused on the process of administering a PROM could be considered for NQF endorsement. Typically, any assessment is quite distal to the outcome of interest. That is, there are multiple process steps between performing an assessment or collecting data and the outcome – i.e., review the data; interpret the data correctly; identify appropriate treatment options; discuss data, treatment options, and recommendaitons with patient; administer treatment. Assessment is necessary but not sufficient to influence outcomes; and the evidence generally will be focused on treatment rather than performing an assessment. NQF criteria and guidance indicate that outcomes and processes proximal to outcomes are preferred for performance measurement. However, NQF criteria do allow for an exception to the evidence criterion.

The primary purpose of a "process performance measure" focused on administrating a PROM is to facilitate use of the PROM to obtain the data needed to develop and test an outcome performance measure. Because developing, testing and endorsing such a process performance measure requires

considerable resources, it should only be considered in an exceptional circumstance and where there is a credible plan to develop the outcome measure.

Specification of the PRO-PM

Recommendation 6.
NQF should require measure specifications for PRO-PMs that include all the following: the specific PROM(s); standard methods, modes, and languages of administration; whether (and how) proxy responses are allowed; standard sampling procedures; the handling of missing data; and calculation of response rates to be reported with the performance measure results.

Performance measures used in accountability applications must be standardized. Therefore, developers must specify them in ways that will help to ensure consistent implementation across providers. Not unlike other performance measures, specifications should identify the data collection tool – i.e., the specific PROM(s) used to obtain the data for each patient (respondent). Specifications that are unique to PRO-PMs include standard methods, modes, and languages of administration, whether (and if so, how) proxy responses are allowed, standard sampling procedures, how missing data are handled; and how response rates are calculated and reported with the performance measure results.

Reliability and Validity of Both the PROM and the PRO-PM

Recommendations 7-8.
7. NQF should require testing for PRO-PMs that demonstrates the reliability of both the underlying PROM in the target population and the performance measure score.

8. NQF should require testing for PRO-PMs that demonstrate the validity of both the underlying PROM in the target population and the performance measure score. Empirical validity testing of the performance measure is preferred. If empirical validity testing of the performance measure is not possible, a systematic assessment of face validity should be accomplished with experts other than those who created the measure, including patients reporting on the PROM, and this assessment should specifically address the approach to aggregating the individual PROM values.

As already noted, NQF endorses performance measures; it does not endorse instruments or scales (i.e., the PROM) alone. However, the PROM values are the data used in the performance measure, so the psychometric soundness of the PROMs specified for use in the performance measures is crucial to the reliability and validity of the PRO-PM. The Expert Panel agreed that reliability and validity of the PROM is necessary but not sufficient to ensure reliability and validity of the PRO-PM; therefore, it recommended that testing for both the PROM and the PRO-PM are needed. Approaches to reliability and validity testing, risk adjustment, and analyses of potential threats to validity were discussed in a commissioned paper on methods issues related to PRO-PMs (available here).

NQF criteria currently allow for testing reliability and validity for either the critical data elements used in the performance measure or for the computed performance measure score. In the case of the PRO-PM, a critical data element is the PROM value.

PROMs have traditionally been developed for group comparisons in research rather than for decisions about individual patients or service recipients. In a research context, investigators usually assign subjects randomly to treatment and control groups; by contrast, in healthcare settings and systems, patients are not randomly assigned to provider of health or support services. The primary question is whether demonstrated reliability and validity of the PROM is sufficient in itself to assume reliability and validity of the performance measure. NQF can consider two approaches to deal with this issue.

1. Accept reliability and validity of the PROM in the target population as meeting NQF criteria for reliability and validity testing at the data element level as long as the additional issues related to threats to validity are tested and analyzed for the performance measure score (i.e., exclusions, risk adjustment, discriminating performance comparability if multiple PROMs are used).

2. Require reliability and validity testing of the computed performance measure score *in addition to* providing evidence of reliability and validity of the PROM in the target population. The related threats to validity must also be addressed (i.e., exclusions, risk adjustment, discriminating performance comparability if multiple PROMs are used).

The primary advantage of the first approach is that measure developers can expend fewer resources for measure testing. The primary disadvantage of the first approach is less confidence in the results of the performance measure. The advantages and disadvantages of the second approach are the opposite.

The Expert Panel agreed that the second approach is more appropriate in the context of performance measures that NQF endorses for purposes of accountability and performance improvement. Further, the impact on resources for additional testing is not substantial, given the need to ~~develop and test risk adjustment~~ address threats to validity, such as differences in case mix or use of multiple PROMs, with either approach. For example, developers could use t~~T~~he data needed ~~for the required testing and analysis related to the threats to validity (e.g.,~~ to develop~~ment~~ and test~~ing of~~ risk adjustment ~~and analysis of comparability if specified with multiple PROMs)~~ ~~could also be used~~ to conduct reliability testing of the performance measure such as a signal-to-noise analysis. Therefore, a requirement for reliability testing of the performance measure would not present an ~~undue~~ additional burden on developers.

Formatted: No underline, Font color: Auto

Formatted: No underline, Font color: Auto

Formatted: No underline, Font color: Auto

Formatted: No underline, Font color: Auto

Validity testing of the performance measure score would require additional data to test hypothesized relationships such as ~~data on~~correlation with another performance measure or ~~information to~~ compar~~e~~ison of performance scores for groups known to differ on quality. NQF criteria currently allow a systematic assessment of face validity of the performance measure score as an indicator of quality. Because developers can specify the performance measure to aggregate individual PROM values in various ways, the validity of results for indicating quality could differ as well. ~~there are a variety of ways that the individual values on the PROM could be aggregated, there could be differences in the validity of the results for indicating quality.~~ Ideally, developers would conduct empirical validity testing ~~would be conducted~~. If that is not possible, then they should evaluate face validity ~~should be evaluated~~ systematically with experts, including patients reporting on the PROM, other than those who created the measure.

Missing Data and Response Rates

Recommendations 9.
NQF should require analysis of missing data and response rates to demonstrate that potential problems in these areas do not bias the performance measure results.

Missing data is an important consideration when using PROM data for performance measurement. This issue encompasses missing responses on a multi-item scale; missing responses from eligible patients and its impact on potential response bias; missing information because of exclusions; and using proxies to mitigate potential missing responses. Systematic missing data affects validity. Processes must be in place to safeguard against these exclusions and biases, and more robust engagement strategies are needed over time to prevent or mitigate poor response rates. NQF criteria for validity currently address exclusions, and missing data is often an explicit or implicit exclusion. Because missing data are likely to be more prevalent with PRO-PMs than with performance measures based on clinical data, developers should address this problem explicitly in measure specifications and ~~in analysis~~testing the PRO-PM, which ~~and~~ will be evaluated by NQF~~ion of the PRO-PM~~.

Feasibility

Recommendation 10.
NQF's feasibility criterion should consider the burden to both individuals providing PROM data (patients, service recipients, respondents) and the providers whose performance is being measured. The electronic capture criteiorn needs to be modified to include PROM data, not just clinical data.

The general principles of feasibility for a performance measure apply to PRO-PMs. Burden of data collection usually applies to the healthcare or service provider whose performance is being measured; however, the unique issue that needs to be considered with PRO-PMs is the potential burden to the individuals who are providing the PROM data. Burdens to both individuals and the providers delivering health or support services will influence response rates, missing data, and ultimately the reliability and validity of a performance measure. Flexibility to decrease burden, such as collecting PROM data through tools developed in multiple languages and applying different methods and modes of administration, is desirable.

As with all performance measures, data collection and reporting for PRO-PMs may present a variety of costs to the providers whose performance is being measured. Such costs may involve expenditures on infrastructure such as computers and programming; they may, in some cases, entail paying licensing or other fees for proprietary instruments or measures. A potential difference between PRO-PMs and other performance measures regarding infrastructure is that, currently, PROMs are not widely in use and the needed information technology infrastructure is less advanced than that of electronic health records.

When considering burdens, ~~it is important to~~developers and NQF need to weigh them against benefits. Obtaining PROM data is not merely a process to collect data for performance measurement. R~~; r~~ather,

providers can use the PROM ~~is used~~ to assess patient status or response to intervention,~~;~~ plan and manage care or services, provide feedback for self-management,~~;~~ and engage patients in ~~shared decisionmaking~~SDM (as desired by patients). The benefits of performance measurement and reporting are widely accepted. As with other performance measures, the burden of data collection does not stop performance measurement; rather, it should serve as an impetus to find more efficient ways to collect PROM data and to use resources for performance measurement on PRO-PMs that meet NQF criteria.

Usability and Use

As with any NQF-endorsed measure, an NQF-endorsed PRO-PM is intended for use in both accountability and improvement applications. The primary indications of whether a performance measure can be applied for these purposes are whether it is in use and whether it is making a difference. At the time of initial NQF endorsement, of course, usability may be only theoretical. The performance measure may have a rationale and plans for use in accountability and improvement activities. On subsequent review for endorsement maintenance, however, NQF requires information on use and data on improvement. NQF also requests public comment on experiences with using the performance measure.

FUTURE DIRECTIONS

This project provided a forum for dialogue among numerous and diverse stakeholders to address difficult conceptual, methodological, and practical issues. The aim was to hasten the endorsement and ultimately the implementation of PRO-based performance measures for use in accountability programs and performance improvement initiatives. The guiding principles articulated above and the detailed pathway (Figure 2) of taking a PRO to a PRO-PM are intended to steer work in the field in ways that help to ensure a more person-centered approach. This report begins to lay a roadmap to get the nation there.

As a next step, PROMs that are at a state of readiness to address performance measure gap areas most meaningful to patients, such as functional status, could be taken down the recommended pathway to develop a PRO-PM and then through the NQF endorsement process. NQF anticipates incorporating PRO-PMs across the domains identified in this report into the broader measure endorsement agenda. PRO-PMs can be submitted for relevant condition-specific topic areas such as cardiovascular or pulmonary, as well as crosscutting areas such as functional status or care coordination.

Nevertheless, some pressing methods issues require further examination. The examples given here are high-priority needs to fill. First, identifying and evaluating best practices for using proxy respondents are important next steps; the goal is not to exclude from our assessments various disadvantaged populations, such as frail elders or children, who may be unable to respond to PROMs on their own. Second, PROs may be evaluated through different PROMs (tools); demonstrating the equivalency of the data from different PROMs warrants careful attention. Of particular concern is the trade-off between allowing implementers as much flexibility as possible without sacrificing validity and enhancing the

ability of users to do meaningful comparisons. Third, viable solutions are needed to overcome barriers to calibrating multiple individual-level PROMs (i.e., "disparate" data sources) to a standard scale. Finally, some considerations will arise as use of PROMs and PRO-PMs expands and evolves. These include the advisability and utility of calculating composite endpoints or combining PRO-PMs salient to a particular domain such as health-related quality of life or health-related behaviors. Having such a broad picture of the outcomes reflected in the PRO-PMs strongly appeals to consumers who want a complete picture of health and well-being.

Using information technology to enable the widespread collection and use of PRO-based performance measures requires further exploration to capitalize fully on existing and future infrastructure. Technology can increase response rates by allowing individuals or their proxy respondents to provide responses from home or elsewhere via telephone, computer tablet, or web-based PRO measurement systems. Technology permits scanning paper and pencil responses; this also allows for ~~quick~~ real-time scoring and giving feedback to respondents. Computers are an essential technology for real-time application of item response theory in computer adaptive testing, which allows more efficient administration of PROMs and calibration of multiple instruments to a standard scale.

Integrating PROMs into electronic health records (EHRs) can facilitate their use for patient-centered care management and also provide data for performance improvement, but implementers must take account of several factors. Data standards are needed before PROM data can be fully incorporated into EHRs. Formulating such standards requires making decisions about aspects of capturing PROM data such as the following: source of the information (e.g., self or proxy); specific PROM instrument; method and mode of data collection; PROM value or response; and dates on which information was captured and scores were computed. In addition, how PROM data might be used in clinical practice needs to be clearly specified. These features include how best to display results and when and how alerts should appear.

Incorporating data provided by patients into the health record may increase their sense of ownership of the record; doing so may also raise demands for extracting information and for providing data. This is an opportune time to include PROMs in EHRs and leverage the resources being directed to adoption of EHRs through the Medicare EHR Incentive Program referred to as "Meaningful Use." Nevertheless, some PROMs, such as those focused on people's experience with care, may not be appropriate to include in EHRs because current tools and approaches are based on the premise of anonymity.

In closing, the path forward toward NQF endorsement of PRO-based performance measures (PRO-PMs) is promising. This project has built on many years of exemplary work in the field of patient-reported outcomes. It now lays out concrete steps to move measurement and use of such data to the forefront of accountability and performance improvement.

REFERENCES

1. Institute of Medicine, *To Err Is Human: Building a Safer Health System*, Washington, DC: National Academy Press; 1999.

2. Institute of Medicine, *Crossing the Quality Chasm: A New Health System for the 21st Century*, Washington, DC: National Academy Press; 2001.

3. Fisher ES, Wennberg DE, Stukel TA, et al., The Implications of Regional Variations in Medicare Spending (Parts 1 & 2), *Ann Intern Med*, 2003;138(4):273-298.

4. McGlynn EA, Asch SM, Adams J, et al., The Quality of Health Care Delivered to Adults in the United States, *NEJM*, 2003;348(26):2635-2645.

5. The Commonwealth Fund Commission on a High Performance Health System, *Why Not the Best? Results from the National Scorecard on U.S. Health System Performance, 2011*, New York: The Commonwealth Fund; 2011.

6. Agency for Healthcare Research and Quality (AHRQ), *National Healthcare Disparities and Quality Reports*, Rockville, MD: AHRQ; 2010. Available at www.ahrq.gov/qual/qrdr10.htm. Last accessed October 2012.

7. Patient Protection and Affordable Care Act, PL No. 111-148, Sec. 3011 (2010).

8. Department of Health and Human Services (HHS), *Report to Congress: National Strategy for Quality Improvement in Health Care*, Washington, DC: HHS; 2011. Available at www.healthcare.gov/law/resources/reports/nationalqualitystrategy032011.pdf. Last accessed October 2012.

9. Department of Health and Human Services (HHS), *Attachment to the Annual Report to Congress: National Strategy for Quality Improvement in Health Care: Agency-Specific Quality Strategic Plans*, Washington, DC: HHS; 2012. http://www.ahrq.gov/workingforquality/nqs/nqsplans.pdf. Last accessed October 2012.

10. Remmers C, Hibbard J, Mosen DM, et al., Is Patient Activation Associated with Future Health Outcomes and Healthcare Utilization among Patients with Diabetes? *J Ambul Care Manage*, 2009;32(4):320-7.

11. Weinstein JN, Clay K, Morgan TS, Informed Patient Choice: Patient-Centered Valuing of Surgical Risks and Benefits, *Health Affairs*, 2007;26(30):726-730.

12. Agency for Healthcare Research and Quality (AHRQ), *Guide to Patient and Family Engagement Environmental Scan Report*, Rockville, MD: AHRQ; 2012. Available at http://www.ahrq.gov/qual/ptfamilyscan/ptfamilyscan.pdf. Last accessed October 2012.

13. Conway J, Johnson B, Edgman-Levitan S, et al., *Partnering with Patients and Families to Design a Patient- and Family-Centered Health Care System: A Roadmap for the Future: A Work in Progress*,

765 Bethesda, MD: Institute for Family-Centered Care; 2006. Available at
766 http://www.ipfcc.org/pdf/Roadmap.pdf. Last accessed October 2012.
767
768 14. National Technology Transfer and Advancement Act of 1995, PL 104-113, 15 USC 3701 (1996).
769
770 15. U.S. Food and Drug Administration, Guidance for Industry, Patient-Reported Outcome Measures:
771 Use in Medical Product Development to Support Labeling Claims, *Fed Regist.* 2009;74(35):65132-
772 133. Available at
773 http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UC
774 M193282.pdf. Last accessed October 2012.
775
776 16. Informed Medical Decisions Foundation, *What is Shared Decision Making?* Available at
777 http://informedmedicaldecisions.org/what-is-shared-decision-making/. Last accessed October 2012.
778
779 17. Elwyn G ,O'Connor A ,Stacey D, et al. Developing a quality criteria framework for patient decision
780 aids: online international Delphi consensus process, *BMJ,* 2006;333:417.
781
782 18. Sepuch KR, Stacey D, Clay CF, et al. Decision quality instrument for treatment of hip and knee
783 osteoarthritis: a psychometric evaluation, *BMC Musculoskelet Disord,* 2011;12:149.

APPENDICES

Appendix A—Glossary

Health-related quality of life (HRQL):

Health-related behaviors:

Patient-reported outcome (PRO): The concept of any report of the status of a patient's health condition that comes directly from the patient, without interpretation of the patient's response by a clinician or anyone else. PRO domains included in this project encompass:

- health-related quality of life including functional status;
- symptom and symptom burden;
- experience with care; and
- health-related behaviors.

PRO measure (PROM): Instrument, scale, or single-item measure used to assess the PRO concept as perceived by the patient, obtained by directly asking the patient to self-report (e.g., PHQ-9).

Performance measure: Numeric quantification of healthcare quality for a designated accountable healthcare entity, such as hospital, health plan, nursing home, clinician, etc.

PRO-based performance measure (PRO-PM): A performance measure that is based on PROM data aggregated for an accountable healthcare entity (e.g., percentage of patients in an accountable care organization whose depression score as measured by the PHQ-9 improved).

Appendix B—Expert Panel Roster

***Patricia Brennan, RN, PhD, FAAN, FACMI (Co-Chair)**
University of Wisconsin-Madison, Madison, WI

***Joyce Dubow, MUP (Co-Chair)**
AARP, Washington, DC

Richard Bankowitz, MBA, MD, FACP
Premier healthcare alliance, Washington, DC

Ethan Basch, MD, MSc
University of North Carolina at Chapel Hill, Chapel Hill, NC

Jim Bellows, PhD, MPH
Kaiser Permanente, Oakland, CA

Laurie Burke, RN
Food and Drug Administration, Silver Spring, MD

Jennifer Eames-Huff
Pacific Business Group on Health, San Francisco, CA

***Stephan Fihn, MD, MPH**
Veterans Health Administration, Seattle, Washington

Floyd Fowler, PhD
Informed Medical Decision Making Foundation, Boston, MA

Lori Frank, PhD
Patient Centered Outcomes Research Institute, Washington, DC

Theodore Ganiats, MD
University of California San Diego Health System, La Jolla, CA

***Kate Goodrich, MD**
Centers for Medicare & Medicaid Services, Washington, DC

Judith Hibbard, DrPH
University of Oregon, Eugene, OR

Dennis Kaldenberg, PhD
Press Ganey Associates, South Bend, IN

Irene Katzan, MD, MS
Cleveland Clinic, Cleveland, OH

***Lewis Kazis, Sc.D**
Boston University School of Public Health, Boston, MA

Uma Kotagal, MSc
Cincinnati Children's Hospital Medical Center, Cincinnati, OH

Kevin Larsen, MD, FACP
Office of the National Coordinator for Health Information Technology, Washington, DC

Kathleen Lohr, PhD, MPhil, MA
RTI International, Chapel Hill, NC

Elizabeth Mort, MD
Massachusetts General Hospital, Boston, MA

Charles Moseley, Ed.D.
National Association of State Directors of Developmental Disabilities Services, Alexandria, VA

***Eugene Nelson, DSc, MPH**
Dartmouth- Hitchcock Medical Center, Lebanon, NH

Kenneth Ottenbacher, PhD, OTR
The University of Texas Medical Branch at Galveston, Galveston, TX

***Greg Pawlson, MD, MPH**
BlueCross BlueShield Association, Washington, DC

Eleanor Perfetto, PhD
Pfizer, Washington, DC

Collette Pitzen, BSN, RN, CPHQ
Minnesota Community Measurement, Minneapolis, MN

Cheryl Powell
Centers for Medicaid & Medicare Services, Baltimore, MD

David Radley, PhD
Institute for Healthcare Improvement, Cambridge, MA

Ted Rooney, RN, MPH
Maine Quality Counts, Manchester, ME

Debra Saliba, MD, MPH
UCLA Borun Center, Los Angeles VA Medical Center, The RAND Corporation, Los Angeles, CA

Marcel Salive, MD, MPH
National Institutes of Health, Rockville, MD

Barbara Summers, PhD, RN, FAAN
University of Texas-MD Anderson Cancer Center, Houston, TX

Kalahn Taylor-Clark, PhD, MPH, BA
National Partnership for Women & Families, Washington, DC

Mary Tinetti, MD
Yale New Haven Health System, New Haven, CT

Phyllis Torda, MA
National Committee for Quality Assurance, Washington, DC

John Wasson, MD
Dartmouth Medical School, Lebanon, NH

Robert Weech-Maldonado, PhD
University of Alabama at Birmingham, Birmingham, AL

Linda Wilkinson, MBA
Dartmouth Hitchcock Medical Center, Lebanon, NH

Albert Wu, MD, MPH
Johns Hopkins Health System, Baltimore, MD

" * " Indicates a member of the Planning Committee

Project Staff

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

Karen Adams, PhD, MT
Vice President, National Priorities

Heidi Bossley, MSN, MBA
Vice President, Performance Measures

Karen Pace, PhD, MSN
Senior Director, Performance Measures

Gene Cunningham, MS
Project Manager, Performance Measures

Evan M. Williamson, MPH, MS
Project Analyst, Performance Measures

Appendix C—Characteristics for Selecting PROMs

Table 4[1]. Important characteristics and best practices to evaluate and select PROs for use in performance measures[279,284]

	Characteristic	Specific issues to address for performance measures	Example: The Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC)[354] for use in hip arthroplasty
1.	**Conceptual and Measurement Model**		
	A PRO measure should have documentation defining and describing the concept(s) included and the intended population(s) for use.	• Target PRO concept should be a high priority for the healthcare system and patients. Patient engagement should define what is an important concept to patients. • Target PRO concept must be actionable in response to the healthcare intervention.	• Factorial validity of the physical function and pain subscales has been inadequate.[355]
	There should be documentation of how the concept(s) are organized into a measurement model, including evidence for the dimensionality of the measure, how items relate to each measured concept, and the relationship among concepts.		
2.	**Reliability**		
	The degree to which an instrument is free from random error.		
2a.	***Internal consistency*** *(multi-item scales)*	Classical Test Theory (CTT): ▪ reliability estimate ≥ 0.70 for group-level purposes ▪ reliability estimate ≥ 0.90 for individual-level purposes Item Response Theory: • item information curves that demonstrate precision [181] • a formula can be applied to estimate CTT reliability	• Cronbach alphas for the three subscales range from 0.86 to 0.98.[356-358]
2b.	***Reproducibility*** *(stability over time)* ▪ type of test-retest estimate depends on the response scale (dichotomous, nominal ordinal, interval, ratio)		• Test-retest reliability has been adequate for the pain and physical function subscales, but less adequate for the stiffness subscale.[358]
3.	**Validity**		

[1] This table is adapted from recommendations contained within a report from the Scientific Advisory Committee of the Medical Outcomes Trust and a report submitted to the PCORI Methodology Committee. The recommendations from these sources have been adapted to enhance relevance to PRO selection for performance measurement.

	Characteristic	Specific issues to address for performance measures	Example: The Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC)[354] for use in hip arthroplasty
	The degree to which the instrument reflects what it is supposed to measure.	• There are a limited number of PRO instruments that have been validated for performance measurement. • PRO instruments should include questions that are patient-centered.	
3a.	***Content Validity***		
	The extent to which a measure samples a representative range of the content.		
	A PRO measure should have evidence supporting its content validity, including evidence that patients and/or experts consider the content of the PRO measure relevant and comprehensive for the concept, population, and aim of the measurement application.		• Development involved expert clinician input, and survey input from patients,[359] as well as a review of existing measures.
	Documentation of qualitative and/or quantitative methods used to solicit and confirm attributes (i.e., concepts measured by the items) of the PRO relevant to the measurement application.		
	Documentation of the characteristics of participants included in the evaluation (e.g., race/ethnicity, culture, age, socio-economic status, literacy).		
	Documentation of sources from which items were derived, modified, and prioritized during the PRO measure development process.		
	Justification for the recall period for the measurement application.		
3b.	***Construct and Criterion-related Validity***		
	A PRO measure should have evidence supporting its construct validity, including: • documentation of empirical findings that support predefined hypotheses on the expected associations among measures similar or dissimilar to the measured PRO • documentation of empirical findings that support predefined hypotheses of the expected differences in scores between "known" groups		• Patient ratings of satisfaction with arthroplasty were correlated with WOMAC scores in the expected direction.[22,360,361]
	A PRO measure should have evidence that shows the extent to which scores of the instrument are related to a criterion measure.		
3c.	***Responsiveness***		
	A PRO measure for use in longitudinal initiatives should have evidence of responsiveness, including empirical evidence of changes in scores consistent with predefined hypotheses regarding changes in the target population.	• If a PRO measure has cross-sectional data that provides sufficient evidence in regard to the reliability (internal consistency), content validity, and construct validity but has no data yet on responsiveness	• Demonstrates adequate responsiveness and ability to detect change in response to clinical intervention.[362]

	Characteristic	Specific issues to address for performance measures	Example: The Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC)[354] for use in hip arthroplasty
		over time (i.e., ability of a PRO measure to detect changes in the construct being measured over time), would you accept use of the PRO measure to provide valid data over time in a longitudinal study if no other PRO measure was available?	
		• Important to emphasize responsiveness because there is an expectation of consequences. Need to be able to demonstrate responsiveness if action is to be taken.	
		• PRO must be sensitive to detect change in response to the specific healthcare intervention	
4.	**Interpretability of Scores**		
	A PRO measure should have documentation to support interpretation of scores, including: • what low and high scores represent for the measured concept • representative mean(s) and standard deviation(s) in the reference population • guidance on the minimally important difference in scores between groups and/or over time that can be considered meaningful from the patient and/or clinical perspective	• If different PROs are used, it is important to establish a link or cross-walk between them. • Because the criteria for assessing clinically important change in individuals does not directly translate to evaluating clinically important group differences, [327] a useful strategy is to calculate the proportion of patients who experience a clinically significant change[271,327]	• Availability of population-based, age- and gender-normative values[363] • Availability of minimal clinically important improvement values[364] • Can be translated into a utility score for use in economic and accountability evaluations[365]
5.	**Burden**		
	The time, effort, and other demands on the respondent and the administrator.	• In a busy clinic setting, PRO assessment should be as brief as possible, and reporting should be done in real-time. • Patient engagement should inform what constitutes "burden."	• Short form available[366] • Average time to complete mobile phone WOMAC = 4.8 minutes[367]
6.	**Alternatives modes and methods of administration**	• The use of multiple modes and methods can be useful for diverse populations. However, there should be evidence	• Validated mobile phone and touchscreen based platforms[368,369]

	Characteristic	Specific issues to address for performance measures	Example: The Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC)[354] for use in hip arthroplasty
		regarding their equivalence.	
7.	**Cultural and language adaptations**	• The mode, method and question wording must yield equivalent estimates of PRO measures.	• Available in over 65 languages[370]
8.	**Electronic health records (EHR)**	Critical features: ▪ interoperability ▪ automated, real-time measurement and reporting ▪ sophisticated analytic capacities	▪ Electronic data capture may allow for integration within EHR[367]