NATIONAL QUALITY FORUM

Driving measurable health improvements together

Building a Roadmap From Patient-Reported Outcome Measures to Patient-Reported Outcome Performance Measures

TECHNICAL GUIDANCE – FINAL DRAFT
NOVEMBER 30, 2021

This report is funded by the Centers for Medicare & Medicaid Services under contract HHSM-500-2017-00060I –75FCMC20F0003.

Contents

Executive Summary

The Building a Roadmap From Patient-Reported Outcome Measures to Patient-Reported Outcome Performance Measures initiative (henceforth referred to as Building a Roadmap) aims to provide guidance to measure developers on how to select high quality patient-reported outcome measures (PROMs) for use in performance measures and how to develop a digital patient-reported outcome performance measure (PRO-PM) that can be used in accountability programs. This work is funded by the Centers for Medicare & Medicaid Services (CMS) and led by the National Quality Forum (NQF). NQF and CMS are using this opportunity to build upon their long-standing commitment to advance the use of patient-reported outcomes (PROs) and to support CMS' commitment to Meaningful Measures and an all-digital measurement portfolio by 2025.[1]

This Technical Guidance Report (i.e., *the Roadmap*) is the third and final report from the project's first year. It builds on two previous reports from the Building a Roadmap initiative: The Environmental Scan Report describes the current state of using PROMs as the data collection instruments for performance measures, while the Interim Report identifies the attributes of high quality PROMs for use in digital PRO-PMs. The Roadmap builds upon these reports by providing guidance to measure developers on developing a digital PRO-PM that is fully tested and ready for submission to NQF's endorsement process. It addresses a gap in the published literature and aligns with CMS' priorities on digital quality measures and PRO-PMs.[1]

NQF convened a multistakeholder Technical Expert Panel (TEP) that met eight times between January and September 2021 to guide the creation of the Roadmap. The TEP identified four stages in the Roadmap that outline the process of developing a PRO-PM:

- **Stage 1: Definition of Measurement Goals**
- **Stage 2: Exploration and Assessment of PROMs**
- **Stage 3: Development and Testing of the PRO-PM**
- **Stage 4: Finalization and Implementation of the PRO-PM**

The TEP also identified 16 tasks that occur within the four stages. The TEP designed the Roadmap with the understanding that the measure developer can flexibly move these tasks within a stage or to a different stage to accommodate individual and organizational style and preference. This flexibility is referenced throughout the report.

Many of the tasks are highly technical and require specific knowledge or training. For example, one task includes developing and testing a risk adjustment model, which requires an advanced understanding of statistics and public health. The Roadmap is not intended as a tutorial to train the reader on how to become a measure developer, nor is it intended as a replacement for key measure development resources (e.g., the CMS Measures Management System Blueprint [henceforth referred to as *the CMS Blueprint*], NQF's Measure Developer Guidebook, or NQF's Measure Evaluation Criteria). Instead, it is intended as a companion to use alongside these resources and as a tool that provides guidance on developing a digital PRO-PM to people with the appropriate training (e.g., psychometrics, informatics, and/or statistics), even if they are new to the field of measure development.

Introduction

Diverse healthcare stakeholders increasingly view PRO-PMs as an important opportunity to ensure that the patient's voice is captured in assessments of quality.[2] Policymakers, payers, and healthcare providers are beginning to use PRO-PMs to inform clinical decision making, improve quality of care, modify provider payment, and evaluate the value of medical technologies.[2]

PROs represent the measurement of a patient's health and behavioral condition, or experience with the healthcare system, directly from the patient without interpretation of the patient's response by a clinician.[3] PROMs represent the tools and instruments that are used to collect the data (e.g., the Patient Health Questionnaire 9 [PHQ-9]).[3] Depending on the measurement concept, PROMs can be used to collect data for PRO-PMs that aggregate patient-reported data to assess the quality of an accountable healthcare entity (e.g., a hospital, health plan, or clinician).[3] In order to properly employ PROMs for these purposes, measure developers must test numerous criteria of the resultant PRO-PM, including its reliability, validity, and feasibility.

CMS views patient-reported measures and digital measures as important components of its measurement strategy.[4] CMS prioritizes outcome measures (i.e., a measure that focuses on the health status of a patient, or the change in health status, that results from healthcare) and patient-reported measures and has established the goal of transforming all measures to digital by 2025.[4,5] Although PRO-PMs make up a small percentage of the overall measures used in CMS' value-based purchasing (VBP) programs and alternative payment models (APMs), CMS sees the broader development and use of these performance measures as an important part of its evolving initiatives to incentivize high quality care. In 2010, the Affordable Care Act (ACA) was passed, and CMS adopted new VBP programs that shifted towards improving and rewarding value rather than volume.[6] CMS designed the VBP programs to ensure that healthcare was more person centered by creating care that focuses on patients' preferences and desired outcomes.[6] Several VBP programs apply to various provider settings, such as hospitals and outpatient centers.[6] In addition to VBP programs, CMS utilizes APMs to incentivize eligible participants to provide high quality and cost-efficient care.[6] VBP programs and APMs are likely to interact during the shift toward improving value, given that incentives linked to APMs may be similar to VBP programs for some providers.[6]

Despite increasing support for PRO-PMs, their development and use are still emerging. Significant challenges hamper the broad adoption of these measures across healthcare. PROMs, the tools on which PRO-PMs are based, have not yet become standard practice in clinical use, and some healthcare professionals do not yet possess a clear understanding of their benefits. This lack of widespread adoption creates multiple challenges, including a lack of patient awareness of the importance of PROMs and insufficient data for efficient testing of PRO-PMs. There is also a lack of thorough, accessible technical guidance that measure developers can use to identify PROMs for use in digital PRO-PMs and to develop high-impact outcome measures based on patient-reported data.[7]

Goals and Objectives

The Roadmap is the third and final report from the first year of the Building a Roadmap initiative. It builds on two previous reports from this initiative: The Environmental Scan Report describes the current state of using PROMs as the data collection instruments for performance measures, while the Interim Report identifies the attributes of high quality PROMs for use in digital PRO-PMs. (The Interim Report and the Roadmap are designed to be used together.)

The goal of the Roadmap is to provide guidance on developing digital PRO-PMs, from the identification of a measure concept and the establishment of a stakeholder advisory group to the preparation for submission for NQF endorsement. The report is intended for novice and advanced measure developers alike. While its guidance is generally applicable to all PRO-PMs, it specifically focuses on digital PRO-PMs that are intended for use in CMS' VBP programs and APMs and can be calculated and transmitted via electronic health record (EHR) systems and other health information technology (IT) systems.

The Roadmap provides an overview that a new developer can quickly read in order to gain a broad understanding of the work involved in developing a digital PRO-PM that is:

- aligned with best and promising practices;
- appropriate for regulatory purposes;
- usable by public and private payers;
- appropriately adjusted for risk; and
- able to attribute fair and accurate linkages between a health outcome and the entity that has control over it.

The Roadmap does not address every question or scenario related to the development of digital PRO-PMs; instead, it is a high-level guide that helps measure developers to understand a complex process, ask educated questions, and contribute to creating a more robust database of digital PRO-PMs. The Roadmap is not intended as a replacement or substitute for existing measure development resources (e.g., the CMS Blueprint, NQF's Measure Developer Guidebook, or NQF's Measure Evaluation Criteria), but rather as a companion to use alongside these resources.

Unless a fact or recommendation is explicitly attributed to a specific source, information in the Roadmap was gathered from the TEP and synthesized by NQF. The Building a Roadmap initiative does not recommend any PROM, nor does it identify any PROM as being "high quality." NQF does not currently endorse, recommend, rank, or prioritize PROMs. This report includes specific PROMs only as examples.

Background

The Roadmap was developed through a multistep process that began with convening a multistakeholder TEP over the course of eight web meetings in 2021. The broad experience of the TEP helped to ensure diverse viewpoints throughout the process. The perspectives of the TEP included patients, patient advocacy groups, health IT professionals, clinicians, health systems, payers, purchasers, and researchers. Due to the technical nature of the Roadmap, NQF intentionally accepted numerous TEP members who are either individually involved in the measure development process or work at organizations that actively develop measures. Many TEP members have contributed to the development of PROMs, some of which are freely available for use and some of which generate licensing fees. Other TEP members are directly involved with either the development or stewardship of PRO-PMs or other outcome measures. The work of these individuals and organizations was transparently acknowledged during the disclosure of interest process that occurred during the TEP's first web meeting, which was open to the public.

TERMINOLOGY

The definitions for PRO, PROM, and PRO-PM used in the Roadmap (Table 1) align with the definitions used in the CMS Blueprint. Because the terminology regarding PRO-PMs is highly technical, the report includes a glossary of key terms (Appendix A: Glossary of Terms).

Table 1: Distinctions Among PROs, PROMs, and PRO-PMs

Concept	Definition	Example
Patient-Reported Outcome (PRO)	What gets measured. The status of a patient's health condition or health behavior that comes directly from the patient (i.e., outcome data)[3]	Symptom: depression
Patient-Reported Outcome Measure (PROM)	How PROs are measured. The tools/instruments used to collect data[3]	Patient Health Questionnaire 9 (PHQ-9)©, a standardized tool to assess depression
Patient-Reported Outcome Performance Measure (PRO-PM)	How PROs are calculated. A way to aggregate the information from patients into a reliable, valid measure of performance[3]	Percentage of patients with a diagnosis of major depression or dysthymia and an initial PHQ-9 score >9 with a follow-up PHQ-9 score <5 at 6 months (NQF #0711)

The Roadmap discusses two different concepts that have similar terminologies, so it is important that readers understand these differences:

- **Attributes and the Attribute Grid:** Throughout the Building a Roadmap initiative, the word *attribute* is used to describe certain characteristics or traits of a high quality PROM that make it suitable for use in a digital PRO-PM. These attributes are collected in a table (i.e., the Attribute Grid) that is designed to facilitate a side-by-side comparison of different PROMs.
- **Attribution:** A process used in quality measurement that aims to assign accountability for a patient's outcomes to a clinician, groups of clinicians, or a facility.[8]

Throughout the Roadmap, **the term *measure developer* refers to the person/team who is developing the PRO-PM**. The term *PROM developer* refers to the person/team who developed the PROM. Although the same person or organization will occasionally act as both the PROM developer and the measure developer, this is an unusual scenario that is not addressed in the Roadmap.

ENVIRONMENTAL SCAN FINDINGS

The TEP's initial responsibility was to advise NQF on the creation of an Environmental Scan Report that assessed the current state of PRO-PM development. Key findings from the Environmental Scan Report included the following:

- **A limited number of NQF-endorsed PRO-PMs exist:** At the time of this publication, only 29 PRO-PMs were actively endorsed by NQF, compared to approximately 200 NQF-endorsed process measures and 320 outcome measures.[9] While PRO-PMs are still an emerging field of measurement, different challenges contribute to their development, including resource limitations (e.g., finances, time, and staff) and a lack of clear guidance.[9]

- **Digital PRO-PMs align with CMS' priorities of reduced burden and elevated patient voices:** The CMS Meaningful Measures 2.0 initiative acknowledges the importance of easily collecting and using patient-reported data across digital systems, the transformation of measures to fully digital (i.e., allowing for data entry, storage, integration, calculation, and reporting to be conducted by EHRs and other health IT systems) by 2025, and the reduction of measurement burden on clinicians and patients.[5] These priorities can be addressed, in part, through the use of digital PRO-PMs, which are intended to collect data that are generated during the normal course of clinical care.[10]
- **PRO-PMs can be designed to use data from one PROM or many different PROMs:** While most NQF-endorsed PRO-PMs collect data from a single PROM that is defined in the measure specifications, there can be benefits to developing performance measures that allow for the use of different PROMs to measure the quality concept of interest. There are advantages and disadvantages to both approaches.
    - **A one-to-one relationship between the PROM and PRO-PM:**
        - The measure developer can select one PROM that is singularly well suited to collect data for the PRO-PM.
        - The measure developer can tailor the PRO-PM specification to the data structure, scoring, and cut points (i.e., markers that indicate the need to screen for a diagnosis or provide treatment) of a single PROM, potentially resulting in a more straightforward development process.
        - The measure steward's maintenance of a single-PROM performance measure is typically less burdensome.
    - **A many-to-one relationship between the PROM and PRO-PM:**
        - Clinical settings have increased autonomy to implement the PROM that best meets their population and business needs (e.g., PROMs that are culturally sensitive and translated into languages that are most relevant to the patient population or PROMs that are free versus those with a licensing cost).
        - The measure developer has the flexibility to create a PRO-PM that collects data from different instruments that measure the same underlying domain (e.g., health-related quality of life [HRQoL], symptoms and symptom burden, or health functioning); however, a crosswalk must exist or be developed to ensure the different PROMs consistently assess quality.
- **Integrated health IT systems are needed for widespread use of digital PRO-PMs:** The widespread use of PROMs and PRO-PMs will require improved integration with EHRs and other health IT systems. This can be achieved through a combination of interoperability standards (e.g., Fast Healthcare Interoperability Resources [FHIR]) and coding schemes (e.g., Logical Observation Identifiers Names and Codes [LOINC]).

INTERIM REPORT RECOMMENDATIONS

Upon the completion of the Environmental Scan Report, the TEP guided the development of the project's second report: the Interim Report.

Attributes of a High Quality PROM and the Attribute Grid

The Interim Report defines the attributes of high quality PROMs that are appropriate for use in a digital PRO-PM. These attributes are described in the Interim Report and presented in the form of an Attribute Grid that measure developers can use to analyze and compare different PROMs that measure similar outcomes (Appendix B). The attributes identified in the Interim Report are listed below:

- Covers desired PROs from patient and/or caregiver perspective
- Outcome measured in PROM is result of care for which relevant clinical quality is being measured
- Interpretable scores, defined and actionable cut points or targets, and anchors and/or defined meaningful change
- Clear conceptual and measurement models
- Psychometric soundness: Reliability
- Psychometric soundness: Validity
- Psychometric soundness: Responsiveness
- Usability/Feasibility of use: Low burden (e.g., length, time/effort to complete) and feasibility
- Usability/Feasibility of use: Fits with standard of care and related workflows (e.g., actionable, incorporated, and discussed at point of care)
- Usability/Feasibility of use: Cultural appropriateness, language, and translations with culturally appropriate items
- Usability/Feasibility of use: Availability of standardized clinical terminology and codes
- Usability/Feasibility of use: Guidance on standardized data collection (including modes and methods)

The attributes in the Interim Report are not intended to be prescriptive: The Attribute Grid does not generate a score or a pass/fail determination, and the TEP opted against defining any "must-have" attributes. The Attribute Grid is intended to allow measure developers to add additional attributes that are pertinent to a PRO-PM being developed.

Prioritized PRO-PM Domains

The Interim Report also addressed the five domains, or areas of patient-reported health status, of PRO-PMs (HRQoL, functional status, symptoms and symptom burden, health behaviors, and patient experience).[11] While all five domains represent important dimensions of performance, the Interim Report described three primary drivers for the Roadmap's focus on HRQoL, functional status, and symptoms and symptom burden. The three drivers are summarized below:

1. **Representation of each domain in currently endorsed NQF measures:** Approximately one-third of NQF-endorsed PRO-PMs fall within the patient experience domain, while HRQoL and symptom-based PRO-PMs are significantly underrepresented.[9]
2. **Assessment of healthcare entity performance:** The Building a Roadmap initiative focuses on areas in which the clinical performance of healthcare entities most directly influences outcomes.[5] These areas are most effectively captured by the domains of HRQoL, functional status, and symptoms and symptom burden. The patient behavior domain centers more on the actions and behaviors of patients than the clinical performance of the entities being assessed, while patient experience typically measures how well a healthcare entity serves patients (e.g., communication or ease of scheduling).
3. **Data collection methodology and suitability for digital measurement**: PROMs for patient experience measures, such as the Consumer Assessment of Healthcare Providers and Systems (CAHPS), typically utilize methodologies that recommend external partners to collect and analyze patient-reported data using mixed mode data collection (e.g., mail, web, telephone, and personal interviews).[12] These methodologies are different than data collection and analysis for the other domains and typically allow the healthcare entity to have less control over PROM selection and data collection workflows. Additionally, this domain may not be as readily applicable to digital measurement.

This prioritization is not intended to diminish the importance of the patient behavior or patient experience domains, and many of the tasks identified in the Roadmap are applicable to these domains.

NQF'S HISTORIC WORK WITH PROS

The Building a Roadmap initiative adds to NQF's body of work on PROs from the past decade, including reports that established clear terminology and provided a pathway from PROs to NQF-endorsed PRO-PMs. This body of work includes, but is not limited to, the following reports:

- 2012: Methodological Issues in the Selection, Administration, and Use of Patient-Reported Outcomes in Performance Measurement in Health Care Settings (henceforth referred to as *Methodological Issues*; this report was updated in 2015 by its authors.)[11]
- 2012: PRO-Based Performance Measures for Healthcare Accountable Entities
- 2013: Patient-Reported Outcomes in Performance Measurement
- 2020: Patient-Reported Outcomes: Best Practices on Selection and Data Collection (henceforth referred to as *PRO Best Practices*)

Roadmap for the Development of a Digital PRO-PM

OVERVIEW AND VISUALIZATION OF THE ROADMAP

The Roadmap provides guidance on the development of PRO-PMs, particularly those that are digital, targeted for NQF endorsement, and suitable for use in CMS' VBP programs and APMs. The Roadmap is organized into four stages. A series of 16 tasks occur within the four stages, and the measure developer should address each task during the development process.

- **Stage 1: Definition of Measurement Goals**
- **Stage 2: Exploration and Assessment of PROMs**
- **Stage 3: Development and Testing of the PRO-PM**
- **Stage 4: Finalization and Implementation of the PRO-PM**

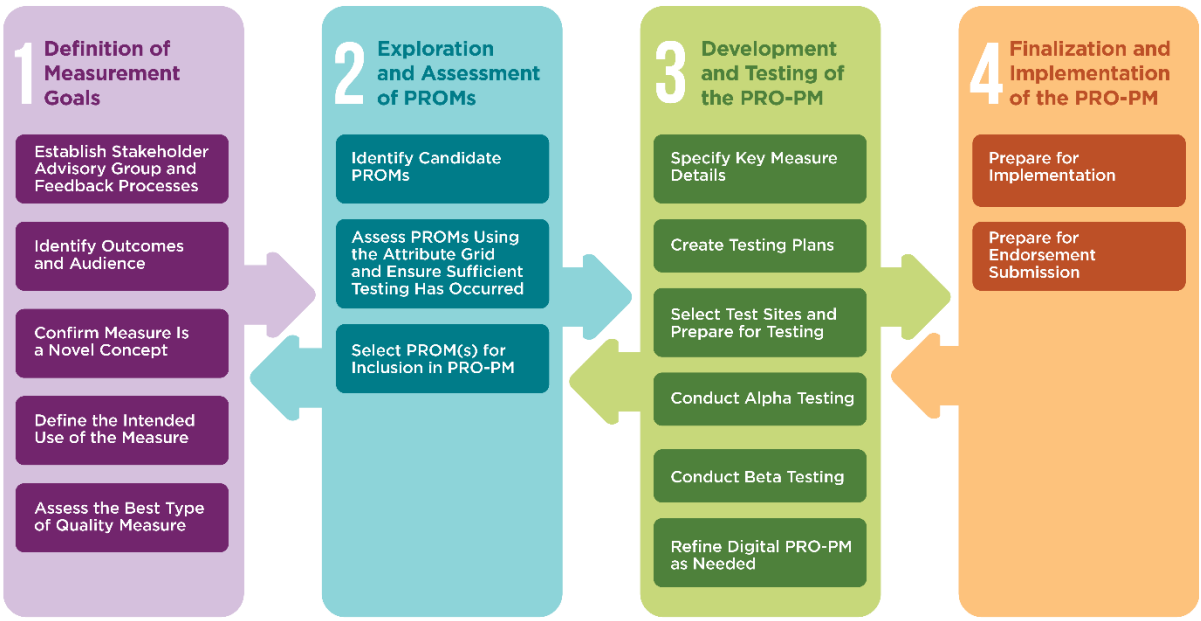

Figure 1. PRO-PM Roadmap: Each column contains one stage, and the bidirectional arrows indicate that tasks can move freely and be iterated across stages.

In the same way that an actual roadmap offers many routes between two locations, the digital PRO-PM Roadmap recognizes that there are many different ways to get from identifying the need for a measure to submitting a fully tested digital PRO-PM for NQF endorsement. While the four stages generally occur in the sequence listed above, the 16 tasks are not entirely linear and are not bound to any specific stage. Different measure developers and organizations may follow different processes when developing digital PRO-PMs; therefore, the Roadmap is designed to allow tasks to move across stages based on individual and organizational style, preference, and need.

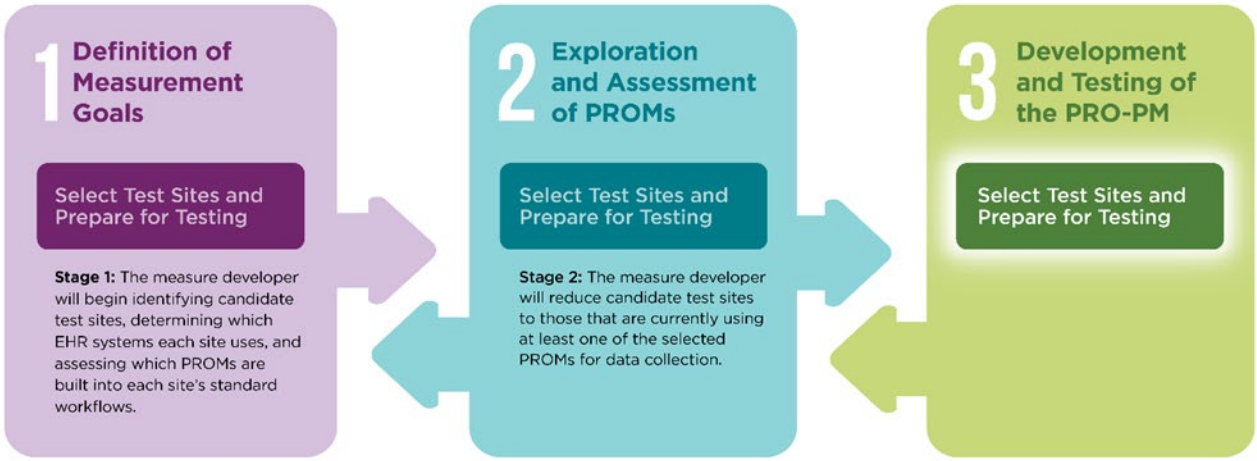

Figure 2: Example of how one task, "Select Test Sites and Prepare for Testing," can be addressed across multiple stages.

For example, attribution is addressed in the stage 3 task titled "Specify Key Measure Details"; however, measure developers on the TEP expressed different preferences on when this task should occur, with some preferring to design the attribution model as part of the preliminary work that occurs during stage 1. Because these variations occur across different people and organizations, the Roadmap is not intended to be prescriptive as to when each task must be performed, and it acknowledges that significant flexibility exists within these tasks. **However, every task is an important part of the digital PRO-PM development process and should be addressed by the measure developer. If a task is not relevant to the PRO-PM being developed, the measure developer should thoroughly and transparently explain the rationale within the measure's documentation.**

In addition to the Roadmap's flexibility on when a given task should be addressed, the TEP recognizes that many tasks related to digital PRO-PM development are iterative and can cross multiple stages. For example, beta-testing a digital PRO-PM is an expensive and complex process that occurs in stage 3, but developers should begin planning for this task during the earliest stages and expand upon it regularly throughout the life of the project.

The visualization of the Roadmap (Figure 1) illustrates the TEP's support for grouping the entire digital PRO-PM development process into four stages, with 16 tasks that can shift across stages. This reflects the flexibility that measure developers may take when developing a PRO-PM.

The stages and tasks are described in detail in the following sections of the Roadmap, and this high-level summary may assist readers in understanding what to expect during each stage.

- **Stage 1: Definition of Measurement Goals.** The measure developer assembles a stakeholder advisory group that is composed of multiple expert perspectives to help identify the measure's outcomes and audience, ensure the measure is novel, determine its intended use, and ascertain

that a PRO-PM is the proper measurement approach. Intentional and strategic thought during stage 1 can help establish a clear plan for the entire measure development life cycle. Successful execution of the tasks within this stage can influence the success of the entire Roadmap up to preparation for NQF endorsement in the final stage.

- **Stage 2: Exploration and Assessment of PROMs.** This stage of the Roadmap includes the identification, assessment, and selection of the PROM(s) that will collect data for the PRO-PM. The Interim Report describes, in detail, the process and criteria for assessing the attributes of a PROM and determining whether it is appropriate for use in a PRO-PM. The Interim Report, along with the Attribute Grid it contains, can help a measure developer to select high quality PROM(s) for use with the performance measure. The Interim Report is designed to be used during stage 2 of the Roadmap.
- **Stage 3: Development and Testing of the PRO-PM.** The third stage of the Roadmap addresses the specification of key details about the measure, including risk adjustment and attribution; the selection of test sites; and the creation and execution of testing plans, including alpha and beta testing. Depending on the measure developer's experience, preference, or organizational policies, the tasks within this stage may vary. However, every task in stage 3 must be considered within the development process. If the measure developer and the key advisors determine that a task is not appropriate for the PRO-PM (e.g., if stratification by clinical or social risk is recommended instead of risk adjustment), the measure developer should thoroughly document these decisions for future reference.
- **Stage 4: Finalization and Implementation of the PRO-PM.** During this stage, the measure developer creates the implementation documentation and prepares for NQF's Consensus Development Process (CDP) depending on whether the PRO-PM will be submitted for NQF endorsement. The preparation for tasks in stage 4 will likely begin early in the development process, even though these tasks are not completed until late in the process. Careful planning and preparation for these tasks will improve the likelihood of successfully testing the PRO-PM and submitting it for NQF endorsement.

It is critical for readers of this report to understand that **the Roadmap is a guide, not a guarantee** of endorsement. Many factors influence whether a measure is endorsed by NQF. The authors of this report encourage measure developers to carefully review and follow the official NQF endorsement processes.

STAGE 1: DEFINITION OF MEASUREMENT GOALS

As with any major project, planning and preparation at the early stages of measure development can positively influence the entire development life cycle. A measure developer should complete these preliminary tasks prior to beginning the development of a PRO-PM. This early work helps to ensure the measure will assist in improving the desired outcomes. Although the Roadmap is generally intended to be flexible, the measure developer should address the following five tasks before moving beyond stage 1. While the measure developer may decide to move other tasks into stage 1, particularly those that will benefit from early planning or iterations, the five tasks listed below should be completed early in the development life cycle.

Establish Stakeholder Advisory Group and Feedback Processes

The measure developer must establish one or more stakeholder advisory groups. The creation of stakeholder advisory groups is a critical early step in developing PRO-PMs because it ensures key perspectives are represented throughout the measure development life cycle. The stakeholder advisory group offers expert feedback and guidance to the measure developer at key information gathering and

decision making points. A responsibility of the stakeholder advisory group is to ensure that the new measure will provide value to a broad range of stakeholders.

Although the Roadmap generally refers to a single measure advisory group, measure developers can choose to have multiple stakeholder advisory groups depending on the feedback needed. A single stakeholder advisory group allows the measure developer to gather diverse viewpoints and seek consensus on complex issues, while multiple groups allow the measure developer to gain more targeted feedback. In situations where multiple stakeholder advisory groups are appropriate, the measure developer should consider scheduling a mixture of separate and combined advisory group meetings to maximize opportunities for transparency, diverse feedback, and consensus.

The composition of the stakeholder advisory group will depend on the measure. The measure developer should first identify which perspectives are mandatory and which are optional, then select individuals to represent these perspectives. **The stakeholder advisory group must include representation from patients, patient advocacy groups, caregivers, and/or consumer groups: These are the people whose perspectives drive the performance measure.** Depending upon the intended use of the PRO-PM, a balanced stakeholder advisory group should also include clinicians, administrative staff (e.g., staff who will potentially administer or assist with PROM completion), payers, health IT staff, EHR vendors, data scientists, policy experts who represent the entities that will be affected by the measure, and any other stakeholder that will be directly affected by the PROM completion and collection process and/or the PRO-PM itself.

In addition to establishing the stakeholder advisory group, the measure developer should clearly define processes to ensure stakeholders can submit feedback in a timely manner throughout the development life cycle and that other interested parties can provide input at key milestones (e.g., public commenting periods). The type and complexity of the proposed measure will influence the involvement of the stakeholder advisory group. More complex measures will likely require more input from the stakeholder advisory group than less complex measures.

Identify Outcomes and Audience

The measure developer, with guidance from the stakeholder advisory group, should identify a consensus definition for the concept that is to be measured, including the desired outcomes that will be monitored and the primary audience of users. Examples of outcomes include functional status after a procedure, evaluation of symptom remission, or assessment of pain control throughout the care continuum. The primary audience of the measure includes whom the measure is intended to assess (e.g., the clinicians who performed a procedure). The measure developer should conduct an environmental scan and utilize empirical data, literature, and/or guidelines to determine and support the measure concept and target population. It is also important that the identified target population for patient-reported measures values the measured outcome and finds it meaningful. The measure developer should gather feedback from the stakeholder advisory group on the proposed concept and target population. Individual patients or focus groups can also be used to determine the value and meaningfulness of a proposed measure.

Confirm Measure Is a Novel Concept

Before moving forward with the measure concept, the measure developer should scan existing measure databases (e.g., NQF's Quality Positioning System [QPS], the CMS Measure Inventory Tool [CMIT], and CMS' Pre-Rulemaking webpage that includes the Measures Under Consideration [MUC] list) to identify and assess any competing or related measures and ensure that the definition, outcome, and audience are all novel or significantly improve upon existing measures. By conducting an environmental scan of measures, the measure developer can avoid duplicating work already being done and adding

unnecessary burden on measured entities. If a similar measure(s) already exists, the measure developer should evaluate whether the concept is unique enough to justify another measure. If the proposed measure concept appears to conflict with an existing measure, the measure developer should work with the existing measure's steward to harmonize the measures as much as possible.

Define the Intended Use of the Measure

The intended use of the measure identifies where or how this measure will be used once completed. Examples of intended use can include quality improvement purposes; comparison of provider performance; or inclusion within accountability programs, such as VBP programs or APMs. The measure developer should finalize the intended use of the measure before moving beyond stage 1. Development and testing processes will vary depending on the intended use of the measure, with requirements being more stringent for measures intended for accountability programs than for quality improvement programs. Determining the intended use ahead of time will help to ensure that the resulting measure is appropriate and likely to be successfully implemented.

Assess the Best Type of Quality Measure

Once the measure concept, target population, and intended use are all determined, the measure developer should evaluate which measure type is best suited to achieve the identified goals. The measure developer should ensure that the goals of the measure are best achieved by a digital PRO-PM, as opposed to another type of measure (i.e., an outcome, process, or structural measure). A PRO-PM is a performance measure that is based on PROs assessed through data that are typically collected via a PROM and then aggregated for an accountable healthcare entity.[13] PRO-PMs are best suited for measure concepts that seek to gather information from patients through questionnaires. If the measure is focused on structured clinical data (e.g., a lab value, blood pressure result, or weight), a PRO-PM is most likely not appropriate.

Digital quality measures (dQMs) are intended to automatically pull data that are generated during the normal course of clinical care. dQMs allow for more efficient capture, retrieval, and sharing of patient data.[10] PRO-PMs are well suited to be dQMs because PROMs often have structured data fields and may need to gather information from multiple data sources (e.g., PROM data and claims data). However, developing PRO-PMs, and specifically digital PRO-PMs, can be complex, time consuming, and expensive. All measure types should be carefully considered prior to deciding on a digital PRO-PM as the correct type of quality measure.

Once the measure developer has confirmed that the measure is unique and should proceed as a digital PRO-PM, stage 2 of the Roadmap can begin.

STAGE 2: EXPLORATION AND ASSESSMENT OF PROMS

Once the initial strategic work is completed, the measure developer can begin an objective and impartial assessment and selection process to determine which PROM(s) are the most appropriate data collection tools for the PRO-PM. Because hundreds of PROMs already exist, including instruments that are specific to a disease or condition as well as those that are designed for general use, readers should note that the Roadmap does not address the rare scenario in which the roles of PROM developer and measure developer overlap, and a single team creates and tests a new PROM before developing the PRO-PM.

The Interim Report and the Roadmap are designed to be used together, and the Interim Report provides extensive detail on the tasks that occur during Stage 2 of the Roadmap.

Identify Candidate PROMs

The measure developer must first identify at least one PROM that is a candidate for use in the digital PRO-PM. PROMs measure one or more of the five patient-reported health status domains: HRQoL, functional status, symptoms and symptom burden, health behaviors, and patient experience.[11] The measure developer can consult several resources to identify candidate PROMs that might be suitable data collection tools for the desired PRO-PM.

- The International Consortium for Health Outcomes (ICHOM) Standard Sets provide lists of recommended PROMs that have been vetted by diverse expert panels.
- Many professional societies have convened working groups of experts to review and recommend PROMs that are relevant within that specialty or discipline.
- Articles from peer-reviewed journals can provide insights into which PROMs are most widely used within a discipline, and which novel PROMs are offering noteworthy contributions.
- The descriptions or specifications of PRO-PMs typically list which PROM(s) are used to collect data for the PRO-PM; websites such as QPS and CMIT include these details if they are available.
- Some payment models in the CMS Quality Payment Program (QPP), such as APMs, identify required PROMs.

Assess PROMs Using the Attribute Grid and Ensure Sufficient Testing Has Occurred

Once the measure developer has identified candidate PROMs, the Attribute Grid (Appendix B; described in detail in the Interim Report) can help to compare PROMs and determine which are most suitable for use in the digital PRO-PM being developed. The Attribute Grid provides a structure for performing a side-by-side comparison of the benefits and limitations of each PROM by prompting the measure developer to review the literature and gather information that indicates whether a PROM is high quality for use in a digital PRO-PM (i.e., its ability to collect data—including patient responses, scores, or threshold data—that are appropriate for use in a digital PRO-PM that is suitable for inclusion in payer assessments, such as a CMS VBP program or APM). Patients, caregivers, advocates, and consumers from the stakeholder advisory group will be important when assessing the first attribute (whether the PROM addresses the desired outcomes from the patient and/or caregiver perspective), while clinician advisors should be closely engaged when assessing the second attribute. However, each of the 12 attributes are important to consider and the stakeholder advisory group may have meaningful input during this process, particularly if the group includes health IT professionals, data scientists, or other uniquely relevant professional perspectives.

The attributes of a high quality PROM that are well suited for use in a digital PRO-PM are as follows:

- Covers desired PROs from patient and/or caregiver perspective
- Outcome measured in PROM is the result of care for which relevant clinical quality is being measured
- Interpretable scores, defined and actionable cut points or targets, and anchors and/or defined meaningful change
- Clear conceptual and measurement models
- Psychometric soundness: Reliability
- Psychometric soundness: Validity
- Psychometric soundness: Responsiveness
- Usability/Feasibility of use: Low burden (e.g., length, time/effort to complete) and feasibility
- Usability/Feasibility of use: Fits with standard of care and related workflows (e.g., actionable; incorporated and discussed at point of care)

- Usability/Feasibility of use: Cultural appropriateness, language, and translations with culturally appropriate items
- Usability/Feasibility of use: Availability of standardized clinical terminology and codes
- Usability/Feasibility of use: Guidance on standardized data collection (including modes and methods)

The TEP strongly encourages the measure developer to consider each of the 12 attributes in the Attribute Grid when selecting a PROM. However, the TEP opted against utilizing a scoring mechanism or identifying any must-have attributes, as these could vary depending on the specific need and intent of the PROM. As such, the measure developer has the flexibility to prioritize the 12 attributes based on the needs of the PRO-PM that is being developed. The measure developer may also identify additional attributes that are relevant to the performance measure and add new rows to accommodate these attributes.

Some of the attributes will have quantifiable results, and some will be subjective. For example, the attribute of "Psychometric Soundness: Reliability" should always include a numeric result. While a measure developer can quantifiably measure the number of languages in which a PROM is translated, the cultural appropriateness of those translations may contain more subjectivity. However, PROM developers have access to well-developed processes for establishing the linguistic and cultural appropriateness of each PROM translation, including pilot testing/cognitive debriefing with patients and clinicians, so even potentially subjective results should be assessed against established scientific methods. The subjective attributes allow the measure developer to think holistically about a performance measure and consider the perspectives of multiple stakeholders, particularly patients and caregivers, when assessing PROMs. The quantifiable attributes are more structured, and there is generally guidance available in literature on how to interpret the results, such as the 2012 NQF Methodological Issues Report, which recommends that PROMs should have a reliability estimate of greater than or equal to 0.70 for group-level purposes and greater than or equal to 0.90 for individual-level purposes.[14] However, the Attribute Grid itself does not specify which results should be considered acceptable. The measure developer should utilize professional judgement and expertise in selecting PROMs while analyzing both the quantifiable and subjective attributes, knowing that the pool of instruments and literature is constantly expanding. The measure developer should refer to existing tools and recommendations when selecting PROMs, such as literature from ICHOM or the Consensus-Based Standards for the Selection of Health Measurement Instruments (COSMIN) as well as previous NQF reports, including the 2012 Methodological Issues Report.

Many elements of testing—including reliability, validity, responsiveness, and mode of administration—are assessed using the Attribute Grid. However, a PRO-PM could require a PROM to undergo additional testing. During this task, the measure developer should also examine additional evidence supporting the PROM's suitability for use in the PRO-PM, such as the determination of whether the PROM was tested with the population in the setting (e.g., inpatient surgery, ambulatory surgery center) that the PRO-PM will be implemented. **Because PRO-PM developers typically do not have the resources to independently test the reliability and validity of each PROM, developers should consider seeking PROMs in which the existing reliability and validity tests were performed in settings and with populations that align with the entities being measured by the PRO-PM.** If the original testing of the PROM does not match the current intended use, the PRO-PM developer should contact the owner of the PROM to determine whether additional information is available.

Select PROM(s) for Inclusion in PRO-PM

Based on the analysis of the Attribute Grid results, the measure developer should determine which PROM(s) best meet the needs of the PRO-PM and whether the PRO-PM will use patient-reported data from a single PROM or from several different PROMs. While choosing a single PROM is likely to result in a simpler development process that focuses on one data structure and scoring mechanism without the need for a crosswalk, it also limits the use to that PROM. This can create issues in care settings if the PROM has licensing fees, does not have translations suitable for the patient population, requires the creation of new workflows, or is not culturally appropriate for the population. Due to these limitations, VBP programs and APMs are more likely to favor PRO-PMs that accept data from different PROMs. However, accepting data from different PROMs makes the development process more complex and may make comparisons between measured entities more difficult. PRO-PM-based outcome measures that use multiple PROMs need to ensure that all PROMs selected have comparable cut points or scores that allow the reporting of outcome rates across entities.

STAGE 3: DEVELOPMENT AND TESTING OF THE PRO-PM

The measure developer can begin stage 3, Development and Testing of the PRO-PM, after assessing and selecting the PROM(s) and confirming that defined instrument parameters and psychometric properties are sufficient for the proposed measurement concept. However, some developers and organizations will prefer to address certain tasks (e.g., attribution or risk adjustment) at different points in the development life cycle. Regardless of where these tasks are initiated, they must be iteratively updated as new information is gained throughout the development process. While the timing of the tasks in stage 3 is flexible, all tasks should be considered during the development process. The measure developer and the key stakeholders might determine that certain tasks are unnecessary for a specific PRO-PM. If this occurs, the measure developer should thoroughly and transparently document these decisions in the PRO-PM's supporting documentation, in part to address questions that will arise during the endorsement process.

Specify Key Measure Details

The technical specifications for a PRO-PM identify important details about the measure, such as the population (i.e., the patients whose PROM data will be aggregated for the measure), the inclusion and exclusion criteria, and the data sources (e.g., PROMs and claims data). In other words, the specification identifies the information and data that are needed to generate the measure result. Measure specifications provide technical instructions on how the measure should be collected and used consistently, reliably, and effectively.[15] Developing measure specifications is an iterative process, and measure developers must include a variety of stakeholder perspectives.[15] The measure developer should engage the stakeholder advisory group in creating and refining the measure specifications and may consider conducting public commenting once the initial draft of the measure specifications is complete in order to acquire additional feedback before investing in the testing process.

Creating full measure technical specifications includes defining the data source (including all PROMs that can collect data for the PRO-PM as well as any other required data source, such as claims data), specifying the code systems, constructing the data protocol, and documenting the measures. Elements that are typically part of technical specifications include the following:

- Measure name/title
- Description
- Population
- Numerator and denominator statements

- Inclusion and exclusion criteria and exceptions
- Data sources
- Key terms
- Data elements, codes, and code systems
- Unit of measurement of analysis
- Sampling methodology
- Risk adjustment
- Time intervals
- The calculation algorithm[15]

Specify Information for Related PROMs

Technical specifications for PRO-PMs require information about the PROM(s) that are used as data collection instruments. While the work from stage 2 should contribute some of this information, other details may be needed, such as acceptable modes of PROM administration, instructions for recording multiple responses to a single PROM question, and the definition of a completed PROM.

Specify Information for Digital PRO-PMs

Digital PRO-PMs may include a narrative measure specification (i.e., a human-readable document) along with specifications in machine-readable formats that have the potential to be updated remotely in real time, much in the way an app on a smartphone can be updated.[15] Measure specifications should use industry-accepted technical specifications (e.g., health quality measure format [HQMF], the Quality Data Model [QDM], and Clinical Quality Language [CQL]), and value sets should be vetted through the National Library of Medicine's (NLM) Value Set Authority Center (VSAC). BONNIE, a tool that allows developers to test and verify the behavior of their digital measure logic using a constructed patient test deck, should be used to confirm that the measure logic works as expected and that value sets are included in the VSAC. The growing support for FHIR, a Health Level Seven (HL7) standard that defines how different IT systems can securely exchange clinical and administrative healthcare information in a timely way, will enhance the use of dQMs and streamline the specification process.[16]

Determine Appropriate Attribution Model

The measure developer must answer an important question during the digital PRO-PM development process: Who should be responsible for the patient outcomes? Attribution is used in quality measurement to assign accountability for a patient's outcomes to a clinician, group of clinicians, facility, or accountable care organizations (ACOs) and health plans.[8] An *attribution model* is a set of rules that defines an accountable unit for a patient's healthcare outcomes; using an attribution model helps determine and specify who is responsible and how the measure can be implemented.[8]

NQF's Attribution: Principles and Approaches Report provides an Attribution Model Selection Guide as a useful resource for determining the appropriate model selection. The report references six guiding principles that acknowledge the challenges in implementing an attribution model:

1. Attribution models should be fair and accurate.
2. Attribution models are an essential part of development, implementation, and policy and program design.
3. Considered choices among available data are fundamental in the design of an attribution model.
4. Attribution models should be reviewed and updated regularly.
5. Attribution models should be transparent and consistently applied.
6. Attribution models should align with the goals and purpose of the program.[8]

During this task, pertinent advisory stakeholders should be engaged on how an entity (e.g., a clinician or a health plan) will be held accountable by the PRO-PM.

Develop and Test Risk Adjustment Model

Risk adjustment refers to statistical methods used to account for patient-, community-, health plan-, or facility-level risk factors when computing outcome performance measures and resource use measures.[17] Patients bring these characteristics into healthcare, and they are outside of the clinician's control. Risk-adjusting has traditionally accounted for differences in health status and clinical factors (e.g., comorbidities or severity of illness) that are present at the start of care, which has been widely accepted and implemented within measure development.[18,19] The idea of incorporating social or functional status risk factors has been debated due to potential unintended consequences, such as masking disparities and institutionalizing differing standards of performance.[20] A measure developer should determine how to approach these characteristics when developing a digital PRO-PM, as they may affect a patient's health outcome. Measures that are risk-adjusted at the facility level allow the healthcare system or hospital to identify improvement opportunities, although risk adjustment can also mask certain patients with poor outcomes. By controlling for appropriate factors, however, risk adjustment can promote a fair comparison among different healthcare entities.

Risk stratification, defined as the division of a population or resource services into similar groups of data that enable analysis of specific subgroups, can be used as an alternative to risk adjustment to identify healthcare disparities.[20] NQF endorsement requires an evidence-based risk adjustment strategy that has demonstrated adequate discrimination and calibration and is based on patient factors that both influence the measured outcome and are present at the start of care. Differences in individuals' responses related to instruments or methods, modes, and languages of administration need to be analyzed and potentially included in risk adjustment.[21] If risk adjustment is not appropriate for the PRO-PM or unrealistic within the development timeline, the measure developer must document and provide rationale and evidence to support the lack of risk adjustment and/or provide the intended plan for future risk adjustment as more PRO-PM data become available after implementation and use in the field.[21,22]

Create Testing Plan(s)

An important aspect of developing measures is the creation of a comprehensive and accurate testing plan that closely follows NQF's measure evaluation criteria. To ensure practical and logistical success of the digital PRO-PM, the measure developer should identify the objectives and goals of the testing process and ensure these are reflected in the testing plan. Data requirements should be addressed in the testing plan, including preliminary strategies for addressing issues that are likely to arise due to issues with data availability, data accuracy, insufficient sample sizes, or other common challenges.

Testing plans can vary depending on measure type and the complexity of the measure. Generally, two types of testing are conducted (alpha and beta) and separate testing plans are created for each. Alpha testing determines how feasible it will be to test and ultimately implement the draft specifications, so the testing plan for alpha testing is usually prepared early in the measure development life cycle. A testing plan for beta testing is usually created later in the project since beta testing should not occur until the measure specifications are almost final. A beta-testing plan should explain the planned analyses for determining scientific acceptability, feasibility, and usability and use. Beta-testing plans generally include the following:

- Name(s) of measure(s)
- Type of testing

- Study objective(s)
- Timeline for testing and report completion
- Data collection methodology
- Description of test population, including number and distribution of test sites/data sets, when available
- Description of data elements for collection
- Sampling methods, if applicable
- If using multiple sites or data sets, a description of strategy to recruit measured entities/obtain test data sets
- Analysis methods planned and description of test statistics to support assessment
- Description and forms documenting patient confidentiality and description of Institutional Review Board (IRB) compliance approval or steps to obtain data use agreements (if necessary)[4]

Select Test Sites and Prepare for Testing

Field testing is a resource-intensive and expensive process that is critical to the success of the digital PRO-PM. The measure developer should coordinate field testing of the digital PRO-PM early in the development process, given its inherent complexity, and should address many details during this task, including the following:

- Identify and confirm test sites to ensure testing occurs in the appropriate setting
    - Select at least two test sites with different EHRs
    - Confirm selected PROM(s) are/can be implemented in potential test sites
- Define workflows, train test site staff, provide resources (e.g., an implementation guide) and ensure health IT (e.g., EHRs, patient portals) functions as expected
- Finalize testing plan with necessary details (e.g., how, when, and where testing will occur; what patients are included in testing; what data are captured; and how to analyze and harmonize data from different test sites.)
- Document how testing results will be assessed and how issues will be resolved
- Prepare any legal documentation and communication, including IRB submissions or data use agreements (DUAs) required under the Health Insurance Portability and Accountability Act (HIPAA)

Conduct Alpha Testing

Alpha testing is an initial assessment that allows the measure developer to determine the feasibility of developing the measure concept. According to NQF's criteria for evaluation, feasibility is the "extent to which the specifications, including measure logic, require data that are readily available or could be captured without undue burden and can be implemented for performance measurement."[21] Examples include clinical data that are routinely generated and used during care delivery (e.g., blood pressure readings or lab test results) and data available in EHRs or other electronic sources. Feasibility should be demonstrated by a data collection strategy that can be reasonably implemented.[22] For PRO-PMs, measure developers must demonstrate that the burden to respondents (i.e., people providing the data) is minimized (e.g., availability and accessibility enhanced by multiple languages, methods, and modes), and the infrastructure to collect instrument-level data is integrated into workflows and EHRs.[21] For digital measures, preliminary feasibility assessments (consisting of both the data model and how various EHR systems map and store the data elements) need to confirm the desired information is available within an EHR or can be added, ideally in a structured format.[4] Measure developers can use NQF's Feasibility Scorecard to rate the feasibility of a measure's data elements using four domains: data availability, data accuracy, data standard, and workflow. The measure developer should perform testing to demonstrate that the measure logic will work (i.e., BONNIE testing). The current state of PROM data

collection can create feasibility challenges, but the measure developer can mitigate these risks by using the Attribute Grid to identify high quality PROMs with LOINC codes and an evidence base that supports successful data collection in the relevant setting from the specified patient population.

Conduct Beta Testing

Beta testing is generally much more extensive than alpha testing. It is used to gather additional information on feasibility and to assess the scientific acceptability and usability of a measure.

Assess Feasibility and Usability

Feasibility should be assessed again once the measure specifications are updated and ready for beta testing to confirm that the finalized data elements can be pulled from the selected data source. Usability should also be evaluated during beta testing. Usability refers to the extent to which potential audiences are able to interpret and understand performance results. The measure developer should consider unintended consequences and a plan to assess whether they outweigh the evidence of improving healthcare quality.[22]

Evaluate PRO-PM for Scientific Acceptability

The scientific acceptability of measure properties is one of the criteria for evaluation that NQF Standing Committees assess when considering a PRO-PM for endorsement.[21] These criteria determine whether the PRO-PM, as specified, produces consistent (i.e., reliable) and credible (i.e., valid) results.[21] **The measure developer must not conflate the reliability and validity of the PRO-PM with the reliability and validity of the chosen PROM(s) because these are completely separate processes.**

Reliability comprises two subcriteria, as described in NQF's Measure Evaluation Criteria:

- The measure must be well defined and precisely specified so that it can be implemented consistently. For PRO-PMs, specifications should include standard methods, modes, and 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.[21]
- The measure's data elements are able to be reproduced, repeating the same or similar results at a high proportion of the time when assessed in the same population in the same time period and/or that the measure score is accurate.[21]
    - PRO-PMs must demonstrate reliability at both the performance measure level (e.g., the PRO-PM) and the data element level (e.g., the PROM and claims-based data elements).[21,22]
    - All digital measures must be tested using the latest accepted technical specifications. Reliance on data from structured data fields is best. If unstructured data are used, reliability and validity for those unstructured data elements must be empirically demonstrated via the data element level.

Validity comprises six subcriteria, as described in NQF's Measure Evaluation Criteria:

- The measure data elements are correct and/or the measure score correctly reflects the quality of care provided, adequately identifying differences in quality.
    - For PRO-PMs, validity must be demonstrated for the data element level as well as for the computed performance score.
    - All digital measures must be tested using the latest accepted technical specifications. Reliance on data from structured data fields is best. If unstructured data are used,

reliability and validity for those unstructured data elements must be empirically demonstrated via the data element level.

- Exclusions are supported by the clinical evidence and are of sufficient frequency to warrant inclusion in the specifications of the measure. If patient preference (e.g., informed decision making) is a basis for exclusion, there must be evidence that the exclusion affects performance on the measure; in such cases, the measure must be specified so that the information about patient preference and the effect on the measure is transparent (e.g., numerator category computed separately, denominator exclusion category computed separately).
- An evidence-based risk adjustment strategy is specified and based on patient factors (including clinical and social risk factors) that influence the measured outcome and are present at the start of care, and it has either demonstrated adequate discrimination and calibration or there is a rationale and/or data to support no risk adjustment or stratification.
- Data analysis demonstrates that methods for scoring and analysis of the specified measure allow for identification of statistically significant and practically/clinically meaningful differences in performance.
- If multiple data sources/methods are specified, there is demonstration that they produce comparable results.
- Analyses identify the extent and distribution of missing data (or nonresponse) and demonstrate that performance results are not biased due to systematic missing data (or differences between responders and nonresponders) and how the specified handling of missing data minimizes bias.[21]

As with many of the tasks in stage 3, evaluation for scientific acceptability is a complex process. The measure developer should begin planning for reliability and validity testing as early as possible in the development life cycle and should recognize the iterative nature of this task.

Refine Digital PRO-PM As Needed

Measure developers should use the testing results to determine whether the measure is ready for implementation. If results show the measure is both reliable and valid, and the benefits outweigh any potential negative consequences, the measure developer should move forward with the measure and begin the implementation process. If the results show that there is low reliability and/or validity or the stakeholder advisory group expresses major concerns about the measure, the measure developer should evaluate what improvements are needed to strengthen the measure in order to prepare it for implementation. The measure developer should also create a mechanism for gathering feedback from those being measured and to consider this feedback when changes are made to the measure.[21]

STAGE 4: FINALIZATION AND IMPLEMENTATION OF THE PRO-PM

The final stage of the Roadmap addresses preparing for implementation of the measure and preparation for NQF endorsement. As with the other stages, the measure developer may address these tasks earlier in the development process based on preference and organizational policies. In fact, the success of the tasks in stage 4 heavily depends on the planning that is recommended during stage 1, as it can be difficult to remedy a shortcoming at the end of the development life cycle if it was overlooked earlier in the process.

Prepare for Implementation

A systematic approach to implementation can minimize unique challenges from varied clinical settings and contribute to the success of implementation across diverse clinical environments. The measure developer should treat the creation of the implementation guidance as an iterative process that

incorporates information from every stage of the PRO-PM development life cycle, including context on data collection via the PROM(s) and lessons learned during testing. It should contain specific examples that are applicable in the relevant settings and with the targeted populations. Implementation guides must provide additional information outside of the dQM technical specification that elucidates the measure developer's intent for each cohort definition and how to ensure the reliability of the information sought in local data systems.

The implementation guidance is a resource document that prepares implementers to put the new measure into practice. By educating implementers on suggested resources and other services that might support measure implementation, the measure developer should collaborate with representatives from testing sites as well as the advisory stakeholders to prepare a guide that will help the entities adapt to the new measure and, ultimately, facilitate improved patient outcomes.

Prepare for Endorsement Submission

Preparing for endorsement is specific to the measure steward or developer. As referenced within stage 1 of the report, the measure developer should determine whether they intend to submit the PRO-PM for NQF endorsement prior to moving forward with the early stages of development. An important consideration throughout the development process is whether testing of the digital PRO-PM is robust enough to support endorsement.

While all NQF measure evaluation criteria are mentioned in the Roadmap, they are not extensively addressed in the Roadmap. Extensive information on both the endorsement process and measure evaluation criteria is available on the NQF website, including the following pages that may be particularly useful to measure developers:

- The Submitting Standards page
- A description of the CDP
- The Measure Evaluation Criteria
- The Measure Developer Guidebook

As mentioned previously in the Roadmap, the CMS Measures Management System Blueprint is also a valuable source of guidance, and the measure developer should closely review this document.

In addition to the tasks outlined in the Roadmap, the measure developer will benefit from documenting information throughout the PRO-PM development process, including rationale for key decisions, lessons learned throughout the development life cycle, and a proposed plan to maintain and update the PRO-PM in the future.

Conclusion

Digital PRO-PMs are an important component of the future of healthcare that will elevate the patient's voice in the assessment of healthcare quality. However, because only a few dozen PRO-PMs currently exist, it is critical that measure developers (particularly those developers who are at the beginning of their careers) receive resources to help them better understand and navigate the development process. When combined with the CMS Blueprint, NQF's Measure Developer Guidebook, NQF's Measure Evaluation Criteria, and the Attribute Grid in the Interim Report, this Roadmap is a resource that can help measure developers navigate the development of digital PRO-PMs. By familiarizing themselves with the four stages of digital PRO-PM development (i.e., Definition of Measurement Goals, Exploration and Assessment of PROMs, Development and Testing of the PRO-PM, and Finalization and Implementation

of the PRO-PM) and the 16 iterative tasks that exist within those stages, measure developers can gain an overview of the PRO-PM development process in a relatively short period of time. The Roadmap is not a textbook that will answer every question related to the development of digital PRO-PMs, but it is a primer that will help measure developers understand a complicated process, ask informed and insightful questions, and ultimately aid in building a more robust database of digital PRO-PMs.

References

1 Centers for Medicare & Medicaid Services (CMS). Meaningful Measures 2.0: Moving from Measure Reduction to Modernization. https://www.cms.gov/meaningful-measures-20-moving-measure-reduction-modernization. Published March 31, 2021. Last accessed August 2021.

2 Discern Health. Patient-Reported Outcome Performance Measures: Current Environment And Next Steps. *Pharmaceutical Research and Manufacturers of America*. December 2017:23.

3 CMS. *Supplemental Materials to MMS Blueprint: Patient Reported Outcome Measures*.; 2021:6. https://www.cms.gov/files/document/blueprint-patient-reported-outcome-measures.pdf.

4 CMS. *CMS Measures Management System Blueprint*. CMS; 2021:112. https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/MMS/Downloads/Blueprint.pdf. Last accessed October 2021.

5 Meaningful Measures 2.0: Moving from measure reduction to modernization | CMS. https://www.cms.gov/meaningful-measures-20-moving-measure-reduction-modernization. Last accessed April 2021.

6 National Quality Forum (NQF). Building a Roadmap From Patient-Reported Outcome Measures to Patient-Reported Outcome Performance Measures Environmental Scan. May 2021.

7 Weldring T, Smith SMS. Patient-Reported Outcomes (PROs) and Patient-Reported Outcome Measures (PROMs). *Health Serv Insights*. 2013;6:61-68.

8 NQF. NQF: Attribution - Principles and Approaches. https://www.qualityforum.org/Publications/2016/12/Attribution_-_Principles_and_Approaches.aspx. Published 2016. Last accessed June 2021.

9 NQF. *NQF: Quality Positioning System*.; 2021. https://www.qualityforum.org/QPS.

10 CMS. *Understanding Clinical Quality Measures: How CMS Is Modernizing Its Approach to Digital Measurement*.; 2020. https://www.youtube.com/watch?v=fsChOWzM2GI&feature=youtu.be. Last accessed December 2020.

11 Cella D, Hahn EA, Jensen SE, et al. *Patient-Reported Outcomes in Performance Measurement*. Vol 97. RTI Press Research Triangle Park, NC; 2015.

12 Agency for Healthcare Research and Quality (AHRQ). Frequently Asked Questions About CAHPS. https://www.ahrq.gov/cahps/faq/index.html. Last accessed June 2021.

13 NQF. *Patient Reported Outcomes (PROs) in Performance Measurement*.; 2013. https://www.qualityforum.org/WorkArea/linkit.aspx?LinkIdentifier=id&ItemID=72537.

14 Cella D, Hahn EA, Jensen SE, et al. Methodological issues in the selection, administration and use of patient-reported outcomes In performance measurement in health care settings. September 2012. http://www.qualityforum.org/WorkArea/linkit.aspx?LinkIdentifier=id&ItemID=72156.

15 CMS. Specification | CMS. https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/MMS/MSP-Specification. Published 2021. Last accessed August 2021.

16 Office of the National Coordinator for Health Information Technology (ONC). *What Is FHIR?* Office of the National Coordinator for Health Information Technology (ONC); 2019. https://www.healthit.gov/sites/default/files/2019-08/ONCFHIRFSWhatIsFHIR.pdf. Last accessed October 2021.

17 Lane-Fall MB, Neuman MD. Outcomes measures and risk adjustment. *Int Anesthesiol Clin*. 2013;51(4):10.1097/AIA.0b013e3182a70a52.

18 Franks P, Fiscella K. Effect of patient socioeconomic status on physician profiles for prevention, disease management, and diagnostic testing costs. *Med Care*. 2002;40(8):717-724.

19 Blum AB, Egorova NN, Sosunov EA, et al. Impact of socioeconomic status measures on hospital profiling in New York City. *Circ Cardiovasc Qual Outcomes*. 2014;7(3):391-397.

20 NQF. *Developing and Testing Risk Adjustment Models for Social and Functional Status-Related Risk Within Healthcare Performance Measurement*.; 2021. https://www.qualityforum.org/WorkArea/linkit.aspx?LinkIdentifier=id&ItemID=95504.

21 NQF. *Measure Evaluation Criteria and Guidance for Evaluating Measures for Endorsement*.; 2019. https://www.qualityforum.org/WorkArea/linkit.aspx?LinkIdentifier=id&ItemID=92804.

22 NQF. NQF: Measure evaluation criteria. https://www.qualityforum.org/Measuring_Performance/Submitting_Standards/Measure_Evaluation_Criteria.aspx. Last accessed April 2021.

23 CMS. Alternative payment models (APMs) overview - QPP. CMS.gov. https://qpp.cms.gov/apms/overview. Last accessed April 2021.

24 McGlothlin AE, Lewis RJ. Minimal clinically important difference: Defining what really matters to patients. *JAMA*. 2014;312(13):1342-1343.

25 NQF. *Patient-Reported Outcomes: Best Practices on Selection and Data Collection - Final Technical Report*.; 2020. https://www.qualityforum.org/Publications/2020/09/Patient-Reported_Outcomes__Best_Practices_on_Selection_and_Data_Collection_-_Final_Technical_Report.aspx. Last accessed April 2021.

26 Deutsch A, Smith L, Gage B, et al. *Patient-Reported Outcomes in Performance Measurement: Commissioned Paper on PRO-Based Performance Measures for Healthcare Accountable Entities*.; 2012. http://www.qualityforum.org/WorkArea/linkit.aspx?LinkIdentifier=id&ItemID=72157.

27 Polascik BA, Hidaka C, Thompson MC, et al. Crosswalks between knee and hip arthroplasty short forms: HOOS/KOOS JR and Oxford. *JBJS*. 2020;102(11):983-990.

28 NQF. *NQF Health Information Technology Glossary: A Guide to HIT Jargon*.; 2015. https://www.qualityforum.org/Measuring_Performance/HIT_Glossary.aspx.. Last accessed April 2021.

29 NQF. *NQF: Measuring Performance*. https://www.qualityforum.org/Measuring_Performance/Measuring_Performance.aspx. Last accessed February 2021.

30 Wright JD. International Encyclopedia of the Social & Behavioral Sciences. In: *2015*. 2 ed. Amsterdam: Elsevier; 2015:26.

31 CMS. Value-based programs. CMS.gov. https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/Value-Based-Programs/Value-Based-Programs. Published January 6, 2020. Last accessed April 2021.

Appendices

APPENDIX A: GLOSSARY OF TERMS

Alternative Payment Models (APMs)
A payment approach that gives added incentive payments to provide high quality and cost-efficient care. APMs can apply to a specific clinical condition, care episode, or population. [23]

Anchors
Anchor-based methods are one of three types of methods used to determine minimal clinically important difference (MCID). A numerical scale for an outcome is "anchored" to a subjective and independent assessment of improvement. For example, a response of "a little better" to a question about how the patient feels post-treatment can be anchored to a numeric outcome.[24]

Attribute
A characteristic or trait of a PROM. Past National Quality Forum (NQF) reports have used *attribute* and *characteristic* synonymously.[13,25] Throughout the Roadmap, *attributes* primarily refer to the characteristics that make a PROM suitable for use in a PRO-PM.

Attribute Grid
A table designed to provide a systematic method to perform a side-by-side comparison of PROMs on the basis of meaningful PROM attributes.[25]

Attribution
A process used in quality measurement that aims to assign accountability for a patient's outcomes to a clinician, groups of clinicians, or a facility.[8]

Burden
Burden refers to the time, effort, or other demands placed on respondents or those administering the PROM. This can include the number and complexity of items and the literacy level needed to understand and complete the measure.[26]

Crosswalk
A concordance table to convert scores from one scale to the other and vice versa.[27] Crosswalks can allow harmonization of PROMs that measure similar outcomes (e.g., HRQoL after a knee replacement surgery), which may facilitate multicenter collaboration or allow sites to switch PROMs without loss of historic comparison data.[27]

Cut Points
Clinically meaningful thresholds of a score change within a PROM that is often associated with either improvement in patient outcome or indication of need for treatment.[24]

Digital Quality Measures (dQMs)
Digital quality measures originate from sources of health information that are captured and can be transmitted electronically and via interoperable systems.[5] These measures utilize data that are generated during the normal course of clinical care. Other types of dQMs include information generated from medical devices, such as ventilators and digitized information from patient portals or other modules.[10]

Electronic Clinical Quality Measures (eCQMs)

These are the most recognizable of the digital measures and are specified for use in the Medicare and Medicaid EHR Incentive Programs. Eligible professionals, eligible hospitals, and critical access hospitals are required to submit eCQM data from certified EHR technology to help measure and track the quality of healthcare services provided within the healthcare system. These measures use data associated with providers' ability to deliver high quality care or related to long-term goals for quality healthcare.[28]

Interpretability

The degree to which the meaning of the scores can be easily understood by any group requiring use of the scores. A PRO measure should have documentation to support interpretation of scores, including the meaning of low and high scores and guidance on the minimally important difference in scores between groups and/or over time.[13]

Logical Observation Identifiers, Names, and Codes (LOINC)

LOINC is a database and universal standard for identifying medical laboratory observations. It was developed in 1994 and is maintained by the Regenstrief Institute, a U.S. nonprofit medical research organization. LOINC was created in response to the demand for an electronic database for clinical care and management and is publicly available at no cost.[28]

Minimal Clinically Important Difference (MCID)

This is the smallest improvement needed after treatment that would be considered worthwhile from the patient's perspective.[24] MCID can be calculated using three different methods: consensus or delphi method, which depends on consensus of an expert panel; anchors (described above); and a distribution-based method, which relies on the statistical analysis of the distribution of outcome scores.[24]

Patient-Reported Outcome (PRO)

1. The measurement of a patient's health and behavioral condition, or experience with the healthcare system, directly from the patient without interpretation of the patient's response by a clinician[3]
2. Any information on the outcomes of healthcare obtained directly from patients without modification by clinicians or other healthcare professionals[13]

Patient-Reported Outcome Measure (PROM)

1. The tools and instruments that are used to collect PRO data.[3] Depending on the measurement concept, PROMs can be used to collect data for PRO-PMs that aggregate patient-reported data to assess the quality of an accountable-healthcare entity (e.g., a hospital, health plan, or clinician).[3]
2. Any standardized or structured questionnaire regarding the status of a patient's health condition, health behavior, or experience with healthcare that comes directly from the patient (i.e., a PRO). The use of a structured, standardized tool, such as a PROM, will yield quantitative data that enable comparison of patient groups or providers.[13]

Patient-Reported Outcome Performance Measure (PRO-PM)

A performance measure that is based on PROs assessed through data often collected through a PROM and then aggregated for an accountable-healthcare entity.[13]

Performance Measures (PMs)

These are standards that can be used to measure and quantify healthcare processes, outcomes, patient perceptions, organizational structure, and/or systems that are associated with the ability to provide high quality care.[29]

Psychometric Soundness

How consistently and accurately an assessment measures what it purports to measure.[26] Validity and reliability are key aspects to attaining psychometric soundness. Psychometrics is a scientific discipline concerned with how psychological constructs (e.g., intelligence, neuroticism, or depression) can be optimally related to observables (e.g., outcomes of psychological tests, genetic profiles, and neuroscientific information).[30]

Value-Based Purchasing (VBP) Program

Value-based programs reward healthcare providers with incentive payments or penalties for the quality of care they give to people with Medicare. These programs are part of CMS' larger quality strategy to reform how healthcare is delivered and paid for.[31]

APPENDIX B: ATTRIBUTES OF HIGH QUALITY PROMS FOR USE IN PRO-PMS

Originally, this was Appendix C from the Building a Roadmap Interim Report. For samples of a completed Attribute Grid, please see Appendices D and E in the Interim Report.

This appendix shows the Attribute Grid with columns for four PROMs that could be compared side by side. (Any number of PROMs can be compared in the grid by adding or removing columns.)

Attribute	PROM 1	PROM 2	PROM 3	PROM 4
Covers desired PROs from patient and/or caregiver perspective	*	*	*	*
Outcome measured in PROM is the result of care for which relevant clinical quality is being measured	*	*	*	*
Interpretable scores, defined and actionable cut points or targets, and anchors and/or defined meaningful change	*	*	*	*
Clear conceptual and measurement models	*	*	*	*
Psychometric Soundness: Reliability	*	*	*	*
Psychometric Soundness: Validity	*	*	*	*

Attribute	PROM 1	PROM 2	PROM 3	PROM 4
Psychometric Soundness: Responsiveness	*	*	*	*
Usability/Feasibility of Use: Low burden (e.g., length, time/effort to complete) and feasibility	*	*	*	*
Usability/Feasibility of Use: Fits with standard of care and related workflows (e.g., actionable, incorporated, and discussed at point of care)	*	*	*	*
Usability/Feasibility of Use: Cultural appropriateness Language Translated with culturally appropriate items	*	*	*	*
Usability/Feasibility of Use: Availability of standardized clinical terminology and codes	*	*	*	*
Usability/Feasibility of Use: Guidance on standardized data collection (including modes and methods)	*	*	*	*

APPENDIX C: COMMITTEE MEMBERS, FEDERAL LIAISONS, AND NQF STAFF

Committee Members

Catherine MacLean, MD, PhD *(Co-Chair)*
Chief Value Medical Officer, Rheumatologist, Hospital for Special Surgery
New York, New York

Sam Simon, PhD *(Co-Chair)*
Director, Health Program Improvement, Mathematica
Cambridge, Massachusetts

David Andrews, PhD
Patient Advisor
Aiken, South Carolina

Katherine Ast, MSW, LCSW
Director, Quality & Research, American Academy of Hospice and Palliative Medicine
Glenview, Illinois

Rachel Brodie, BA
Senior Director, Measurement & Accountability, Pacific Business Group on Health
San Francisco, California

Zahid Butt, MD, FACG
President and Chief Executive Officer (CEO), Medisolv, Inc.
Columbia, Maryland

Collette Cole, BSN, RN, CPHQ
Clinical Measurement Developer, Minnesota Community Measurement
Minneapolis, Minnesota

Paula Farrell, BSN, RN, CPHQ, LSSGB
Associate Project Director, Clinical Quality Measurement, Department of Quality Measurement, The Joint Commission
Oakbrook Terrace, Illinois

Mark Friedberg, MD, MPP
Senior Vice President, Performance Measurement & Improvement, Blue Cross Blue Shield of Massachusetts
Boston, Massachusetts

Debbie Gipson, MD, MS
Professor of Pediatrics, University of Michigan
Ann Arbor, Michigan

Ben Hamlin, MPH
Senior Research Informaticist, Research & Performance Measurement, National Committee for Quality Assurance
Washington, District of Columbia

Janel Hanmer, MD, PhD
Medical Director of Patient Reported Outcomes, University of Pittsburgh Medical Center
Pittsburgh, Pennsylvania

Helen Haskell, MA
Founder and President, Mothers Against Medical Error
Columbia, South Carolina

Brian Hurley, MD, MBA, DFASAM
Director of Addiction Medicine, Los Angeles County Department of Health Services
Los Angeles, California

Christine Izui, MS
Health Policy Principle, Center for Transforming Health, MITRE Corporation
McLean, Virginia

Laura Jantos, LFHIMSS
Patient Advocate, Healthcare Technology/Digital Health
Mercer Island, Washington

Kirk Munsch
Patient Advocate, Rare Patient Voice
Colorado Springs, Colorado

Deborah Paone, DrPH, MHA
Performance Evaluation Lead and Policy Consultant, SNP Alliance
Washington, District of Columbia

Brenna Rabel, MPH
Lead Health Research Scientist, Battelle Memorial Institute
Arlington, Virginia

Nan Rothrock, PhD, MA
Clinical Psychologist, Feinberg School of Medicine, Northwestern University
Chicago, Illinois

Mike Sacca, AS
Vice President, Healthcare Quality and Data Analytics, RELI Group
Douglassville, Pennsylvania

Rachel Sisodia, MD
Senior Medical Director for Specialty Care and Patient Reported Outcomes, Massachusetts General Brigham Health
Boston, Massachusetts

John Spertus, MD, MPH, FACC, FAHA
Clinical Director of Cardiovascular Education and Outcomes Research, Mid America Heart Institute
Kansas City, Missouri

Ruth Wetta, PhD, MSN, MPH, RN
Lead Clinical Researcher, Population Health, Cerner Corporation
Kansas City, Missouri

Albert Wu, MD, MPH, FACP
Professor, Department of Health Policy and Management and Medicine, Johns Hopkins Bloomberg School of Public Health
Baltimore, Maryland

Federal Liaisons

Joel Andress
Centers for Medicare & Medicaid Services (CMS), Center for Clinical Standards and Quality

Janis Grady, RHIT, FAC-COR III
CMS, Center for Clinical Standards and Quality

Sandra Mitchell, PhD, CRNP, FAAN
National Institutes of Health (NIH), National Cancer Institute

Ashley Wilder Smith, PhD, MPH
NIH, National Cancer Institute

Rhona Limcango, PhD
Agency for Healthcare Research and Quality

Meghan McHugh, PhD, MPH
Substance Abuse and Mental Health Services Administration

Clifford A. Smith, PhD, ABPP-Cn
Veterans Affairs, Office of Mental Health and Suicide Prevention

David Au, MD, MS
U.S. Department of Veterans Affairs (VA)

Girma Alemu, MD, MPH
Office of Planning, Analysis, and Evaluation, Health Resources and Services Administration

NQF Staff

Kathleen Giblin, RN
Senior Vice President, Quality Innovation

Alejandra Herr, MPH
Senior Managing Director

Maha Taylor, MHA, PMP
Managing Director

Chuck Amos, MBA
Senior Director

Poonam Bal, MHSA
Senior Director

Teresa Brown, MHA, MA
Senior Manager

Deidra Smith, MBA, PMP
Senior Project Manager

Evelyn Thomas, MPH, CHES
Senior Analyst

Jhamiel Prince, MA
Analyst

Juanita Rogers, MS, CHES
Analyst

CMS Staff

LaWanda Burwell
Task Order COR

Gequincia Polk
IDIQ COR

Helen Dollar-Maples
Acting Director, DPMS/QMVIG/CCSQ

APPENDIX D: PUBLIC COMMENTS AND RESPONSES

The draft Technical Guidance Report for the Building a Roadmap From Patient-Reported Outcome Measures to Patient-Reported Outcome Performance Measures initiative was posted on the National Quality Forum (NQF) project webpage for public and NQF member comment from September 2, 2021, through September 24, 2021. Six prompts were offered to guide public commenters on key areas of interest. Eight comments from four organizations are grouped below by prompt, and the responses from NQF and the TEP are included beneath each comment. Unless otherwise noted, public comments are presented as they were received by NQF and have not been edited, except for minor updates to spacing, spelling, and punctuation.

What general recommendations, comments, or feedback do you have for the report?

America's Health Insurance Plan (AHIP)

COMMENT

AHIP appreciates the opportunity to comment on this report. Health insurance providers are committed to providing quality healthcare and believe measurement is an essential element in ensuring we meet this goal. The report notes that PROMs are not yet standard practice and that some clinicians and payers possess limited understanding of their benefits. Health insurance providers agree with the benefits of PRO-PMs, but the implementation burden can be high on all parties, including patients. For example, clinicians need to change workflows to implement PROMs, patients may experience survey fatigue, and collecting and aggregating data can be costly for payers. AHIP supports efforts to make measurement more person-centered and ensure we assess what truly matters to patients, but minimizing the implementation burden will be key to increasing adoption of PRO-PMs. This report could explore the impact of increased interoperability and the possibility of leveraging ONC's efforts around TEFCA and the Cures Act Final Rule as well as CMS's Interoperability and Patient Access Final Rule to increase data sharing. To facilitate greater use of PROs, we suggest developers create digital patient-reported outcome measures and leverage emerging technology such as APIs to alleviate the data collection burden. Incentivizing the collection of patient-reported outcomes would allow greater understanding of outcomes from the consumer's point of view and advancement of person-centered care.

Misaligned measures have been a source of extra burden on the healthcare system. One reason measures are not aligned is the varying ability to calculate a measure. For example, private payers do not have access to the clinical data necessary to calculate current measures that use paper medical records or EHRs as data sources. The transition to dQMs presents a new opportunity to ensure all stakeholders can use the same measure, thus reducing variation in measure specifications and the use of related measures by different payers. We support exploring opportunities to further align measures across payers. We suggest leveraging the work of the Core Quality Measures Collaborative (CQMC) to provide a framework for greater alignment of measurement. In fact, the CQMC created a workgroup to discuss many of the issues raised here and to seek consensus on the path forward for advancing digital measurement. For example, digital measurement provides new potential for further alignment on not just the measures themselves but the full measurement model from the collection of data elements through reporting and calculation of the measures and back to the providers for near real-time feedback and clinical decision support. This may be particularly important for adoption of digital PRO-PMs. Increased alignment would lead to more payers and providers using the same PRO-PM and decrease the burden of implementation.

RESPONSE

Thank you for your response. Some of your suggestions, such as burden and workflows related to PROMs, were addressed in the September 2020 NQF report titled *Patient-Reported Outcomes: Best*

Practices on Selection and Data Collection. Others were identified in the Environmental Scan Report for the Building a Roadmap initiative, published in May 2021. We have added additional language to the Technical Guidance Report to address some of your suggestions. We also agree with your recommendation to leverage the work of the Core Quality Measures Collaborative. We aim to expand on this alignment and emerging developments in the rapidly expanding knowledgebase of digital measurement in future versions of this report.

American Academy of Physical Medicine and Rehabilitation (AAPM&R)

COMMENT

On behalf of the over 9,000 members, The American Academy of Physical Medicine and Rehabilitation thanks NQF for the opportunity to comment on "Building a Roadmap From Patient-Reported Outcome Measures to Patient-Reported Outcome Performance Measures." We appreciate NQF's efforts, in conjunction with the Roadmap Technical Expert Panel, to harness multiple inputs and put some framework around PRO-PM measure development; however, we did not find the current report to provide the in-depth, thorough guidance needed to differentiate standard measure development (of PRO-PMs or ANY type of measure) to a digital format. This report merely grazed the surface of key steps in measure development and once again summarized existing reports and guidance versus providing specificity in tasks needed to differentiate development and implementation of digital metrics. For example, from page 16: "PRO-PMs must demonstrate reliability at both the performance measure level (i.e., the PRO-PM) and the data element level (i.e., the PROMs)." This is a generic statement and is a requirement of any measure that goes through the NQF Consensus Development Process. We would have loved to see expanded discussion and direction on different testing parameters to ensure reliability of data for digital collection and reporting as well as expanded discussion on importance of using standardized data from EHRs and registries. This is just one example from a report that does not meet the stated objective of providing technical guidance in moving PROMs to PRO-PMs for digital translation. We hope to see and collaborate with NQF on future work on this important topic.

RESPONSE

Thank you for your comment. We worked to find an appropriate balance between high-level guidance and detailed instructions, knowing that the report is intended to help measure developers—particularly junior staff who have not developed a PRO-PM—understand the overarching process of developing a digital PRO-PM. In this spirit, the report identifies existing reference materials for measure developers, including the CMS Measures Management System Blueprint and NQF documents on Measure Evaluation Criteria and the Consensus Development Process (CDP). While we believe we found an appropriate balance in the Technical Guidance Report, NQF and the TEP hope to create future versions of the report that expand upon digital PRO-PM development and incorporate suggestions and lessons learned from measure developers who have used the Technical Guidance Report while developing a digital PRO-PM.

American College of Physicians (ACP)

COMMENT

The American College of Physicians (ACP) appreciates the opportunity to comment on NQF's Building a Roadmap From Patient-Reported Outcome Measures to Patient-Reported Outcome-Based Performance Measures Technical Guidance - Part 2.

ACP recognizes the importance of patient-reported outcome-based performance measures (PRO-PMs) and applauds the work of the Committee in getting us closer to the goal. However, this requires the development of methodologically sound PRO-PMs that have been shown to have an impact on patient outcomes and to provide actionable information to physicians. These principles for PRO-PMs are

necessary to achieve patient-centered care. We have some comments and concerns for the Committee's consideration.

1. The ACP values the Committee's identification of the four stages in the process of developing a PRO-PM and the 15 tasks that fall within each of those stages. We agree that the process needs to be fluid enough to allow for variation in measure development processes or organizational differences. However, we also agree with the Committee's perspective that if a measure developer believes that a task is not relevant to the PRO-PM being developed, the measure developer should document the rationale for future reference. The ACP would also take that one step further and recommend that a thorough rationale be provided and made publicly available and transparent within the overall documentation for the measure.

2. In stage 1, the ACP also supports the Committee's emphasis on ensuring that key stakeholders "who represent the entities that will be affected by the measure, and any other stakeholder that will be directly affected by the PROM completion and collection process and/or the performance measure" are identified. We support the recommendation that they offer their guidance throughout the development process.

3. The ACP concurs that too often developers and endorsers confuse reliability and validity of the PRO-PM with the reliability and validity of the PROM. We recommend that this key point be expanded upon to prevent this from happening in the future.

4. The emphasis on feasibility of performance measures is one of prime importance to ACP and its members, and we are pleased to see it noted in the roadmap. As we highlighted in ACP's Recommending Caution in Patient-Reported Outcome–Based Performance Measurement paper (Ann Intern Med. doi:10.7326/M19-3603), the seamless collection of PROM data is not yet part of the routine practice of medical care in the United States. This relates to many different factors, including the burden associated with setting up systems to collect the data, the lack of engagement of patients in reporting outcomes, and limitations on clinical utility or ability to act on results, among others.

The ACP would recommend that feasibility across a variety of different practices and settings be assessed as part of the field testing described in stage 4.

5. Finally, we support this roadmap as a primer for developers contemplating developing PRO-PMs, as it includes a number of the important considerations in the development process. However, the comment that "measure developers can clearly understand the PRO-PM development process in less than one hour" with this roadmap seems inconsistent with other statements in the document and is oversimplifying a rather complicated process.

RESPONSE

Thank you for your support of the report. We agree with your suggestions and have edited the language in the report accordingly.

National Committee for Quality Assurance (NCQA)

COMMENT

While NCQA commends NQF for taking on this important topic, we feel that the Technical Report does not fully accomplish its stated goals of increasing understanding about moving from PROMs to PRO-PMs and paving the way for needed digitalization of PRO-PMs. It did not address the need for data sharing, quality, and validation nor take full advantage of recent progress in the area of PRO-PM digitalization.

NCQA also thinks it would have been helpful if NQF through this project had explored some "big picture" issues about the role of PRO-PMs in the "quality ecosystem" and how they can enhance patient-centered care and health equity. It would have been very useful if NQF had tried to tackle important and thorny issues regarding levels of accountability and the use of PRO-PMs for QI versus accountability. A more visionary, detailed, and up-to-date exploration of PRO-PMs and their role in quality improvement is needed to assure development of robust, digital patient-centered, culturally competent PRO-PMs capable of driving high quality care that matters to patients.

RESPONSE

Thank you for your comment. We have edited the report with additional text related to the area of digital PRO-PMs, and we hope to expand on this information in future versions of the report as new information emerges. While we agree with other portions of your comment, such as the need for guidance on "big picture" issues related to PRO-PMs, the scope of this project does not permit a thorough exploration of these topics beyond what was presented in this report and the Environmental Scan and Interim Reports. We encourage NCQA to remain engaged with NQF and continue to seek new opportunities to expand the knowledge around PRO-PMs.

Because there is not one "correct" linear process to measure development, the report refers to a series of stages and tasks. How clear is this structure, and how can it be improved?

National Committee for Quality Assurance (NCQA)

COMMENT

NCQA agrees in general with the stages and tasks as explicated in the report. We might suggest that some would be done in a different order, but we understand and agree with the idea that the tasks are iterative, and the process is sometimes nonlinear. NCQA thinks this could be improved if there were more detail. The report presents a superficial description of the process and does not add much insight into the complexities of measure development, particularly PRO-PMs and digital measures.

RESPONSE

Thank you for your comment. We aimed to find an appropriate balance between high-level guidance and detailed instructions and provide a guidance document targeted to measure developers (particularly junior staff who do not have experience with PRO-PMs) that will help them to understand the overarching development process and identify existing reference materials. We do agree that more detail is necessary in some places, and we have added additional detail to several tasks in the report, particularly within stage 3.

How can the guidance on digital measurement be improved in the report?

America's Health Insurance Plans (AHIP)

COMMENT

AHIP supports aligning data needed for measurement with the requirements of the Interoperability and Patient Access rule. This would support measure alignment across payers, allow comparisons across providers, prevent competing or duplicative requirements, and allow the incorporation of new data types into measures. However, there are still technical and infrastructure challenges that must be addressed to facilitate a transition to digital measurement. The healthcare system lacks the infrastructure to successfully and efficiently report and exchange electronic data. This report could explore where there are gaps in the content and technical standards needed to exchange digital PRO-PMs. These standards as well as implementation guides must be fully developed and sufficiently tested

for successful implementation of truly interoperable sharing and transparency. Mature standards should be a precursor to implementation. Ongoing work by HL7, such as the C-CDA and Da Vinci Implementation Guides, can lay the groundwork for better exchange of the data needed. Additionally, there is still a significant gap in EHR use by clinicians and PAC/LTC providers and [an] even bigger gap in their ability to exchange data with plans. Providers with minimal or no EHR use may not be able to report using dQMs. Plans have expressed difficulty getting provider buy-in on current digital approaches such as NCQA's ECDS measure reporting. Clinicians who are meaningfully using EHRs may be reluctant or unwilling to share information contained in those records with public or private sector payers.

This report could also explore potential ideal attributes of a digital PRO-PM. For example, they should be compatible with any source data system that implements standard interoperability requirements. If separate data systems require separate dQMs, we will only reinforce the current fragmented and siloed approaches to medical data and performance measurement. We agree that dQMs should exist separately from digital data sources and be tested and maintained independently from the data source systems. The adoption of dQMs should not require that providers or payers adopt specific EHR systems or join certain registries or health information exchanges to calculate dQMs. Digital measures should also have the flexibility to be deployed by different stakeholders depending on the program and measure needs and specifications.

RESPONSE

Thank you for your comment. We agree that there are unresolved infrastructure challenges and gaps that hinder the development, testing, and use of digital PRO-PMs. We aim to create future versions of this Technical Guidance Report that address the rapidly evolving field of digital measurement. While your suggestion for ideal attributes of digital PRO-PMs is not within the current scope, we will consider it for future versions of this report as well as for future initiatives.

National Committee for Quality Assurance (NCQA)

COMMENT

NCQA notes that there was minimal guidance about navigating the exploding digital landscape in healthcare data interoperability in general and digital quality measurement specifically. There was essentially one paragraph in the entire report that dealt exclusively with digitalization. There was no sophisticated discussion directed toward key issues such as how digitalized PROMs interact with digitalized PRO-PMs or how interoperability standards must be addressed, for example. It would be useful to have examples of current best practices in clinical entities using PRO-PMs that could illustrate the state of art on digitalized PRO-PMs and include information on clinical training in use and interpretation.

RESPONSE

Thank you for your comment. NQF and the TEP have added information about digital performance measurement to this report. Additionally, NQF aims to create future drafts of the Technical Guidance Report that capture and build upon the rapidly expanding realm of digital measurement and interoperability.

What improvements would make the report more useful to new measure developers?

National Committee for Quality Assurance (NCQA)

COMMENT

NCQA would like to commend and thank NQF and CMS for convening an experienced and knowledgeable multistakeholder Technical Expert Panel to tackle critical questions about PRO-PMs: how

to assess PROMs for use in performance measures and how to develop digital PRO-PMs for use in accountability programs. We believe that the importance of PRO-PMs, which bring the voice of the patient into quality measurement, is unfortunately matched by the paucity of appropriate measures, and we agree with NQF's efforts to produce a report that is useful to new measure developers.

However, NQCA thinks that the report falls short of providing the guidance and "Roadmap" that has the requisite information to encourage and increase new PRO-PM measure development. Much of the information provided is similar to currently available information. The refinements added do not really address advancements in measurement science; digitalization standards; interoperability; and data aggregation, sharing, and validation that will inform development of new, digital PRO-PMs. Rather, the TEP appeared to focus more on PROM psychometric attributes that have been understood for years and a high-level description of some aspects of performance measure development in general that are also already in wide use and are not specific to either PRO-PM development or measure digitalization.

The examples of PROMs used to explain PROM attributes were helpful to illustrate how to evaluate PROMs using these attributes. However, this exercise did not address challenges in the digitalization of either a PROM or a PRO-PM.

Other important issues were not discussed in detail. Patient and provider acceptability of both PROMs and PRO-PMs is critical to their use for quality improvement or accountability. There was little information about characteristics of PROMs or PRO-PMs that make them acceptable and meaningful to patients and non-burdensome to providers. A PRO-PM that measures the delivery of care that matters to patients and aligns with patients' goals for their health might have much more meaning for both patients and clinicians than multiple questions about minor improvements in a symptom or function that may seem repetitive to a patient and of unclear use to a clinician.

NCQA notes that the attributes of high quality PROMs did mention the importance of translations into various languages and testing in diverse populations. The point was also made that translations may not reflect all desired domains and that testing was necessary. Although NCQA thinks the report could have benefited from a more detailed discussion of any health equity implications of PROMs and PRO-PMs, we agree that key issues were raised.

RESPONSE

Thank you for your comment. We have added information to the report to address some of the advancements in digital measurement. Additionally, we aim to develop future versions of the report that not only address the rapidly evolving state of digital measurement but also capture the most important high-level information that new measure developers need when working on a digital PRO-PM. We agree with the importance of other issues that are not addressed in detail in this report, such as patient and provider acceptability of both PROMs and PRO-PMs. The NQF report titled *Patient-Reported Outcomes: Best Practices on Selection and Data Collection*, published in September of 2020, discusses many of these concerns.