NATIONAL QUALITY FORUM
Driving measurable health improvements together

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

TECHNICAL GUIDANCE - DRAFT 2

September 14, 2022

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

Contents

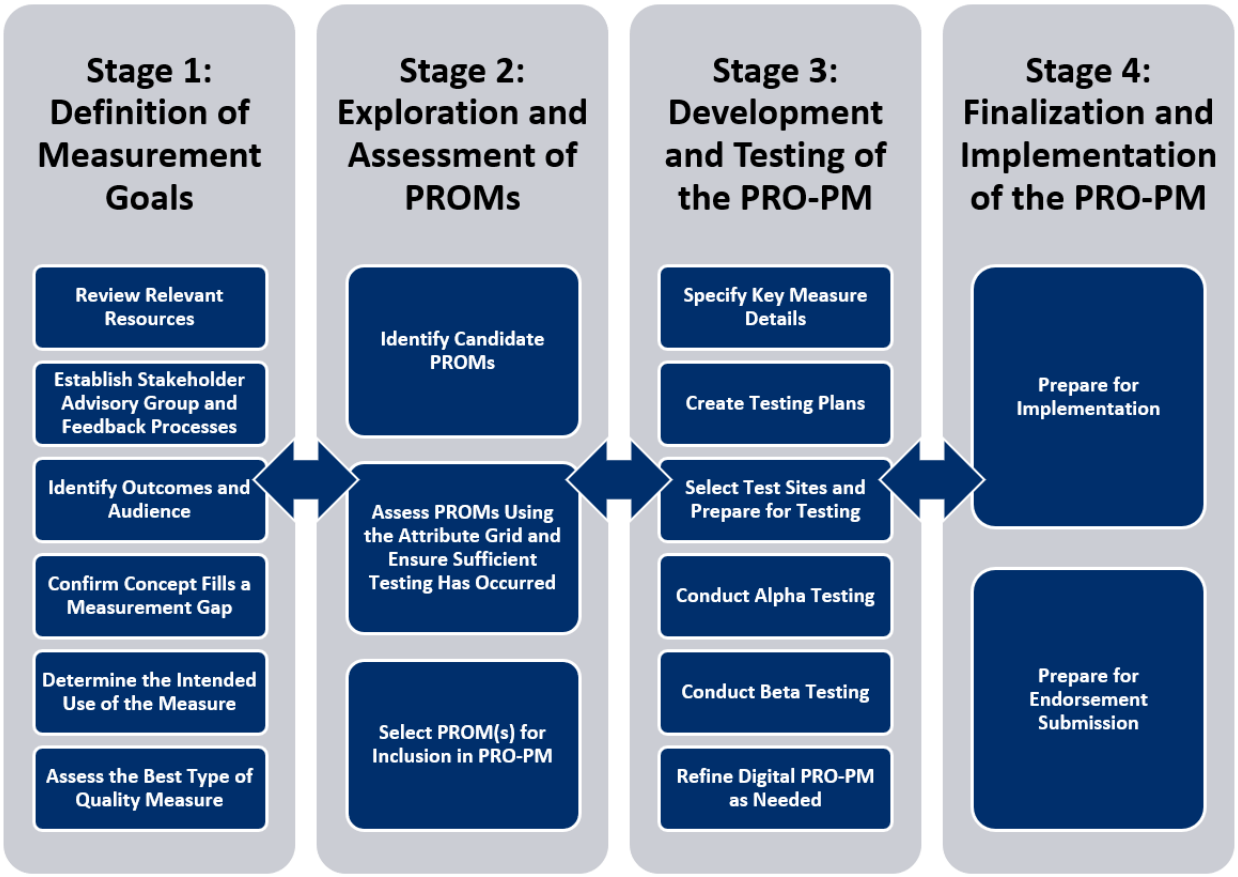
Executive Summary

The Building a Roadmap From Patient-Reported Outcome Measures to Patient-Reported Outcome Performance Measures initiative (henceforth referred to as *Building a Roadmap*) provides guidance to measure developers on the development of digital patient-reported outcome performance measures (PRO-PMs). This work is funded by the Centers for Medicare & Medicaid Services (CMS) and led by the National Quality Forum (NQF), both of whom have a long-standing shared commitment to advance the use of patient-reported outcomes (PROs) and digital measurement.

CMS identifies the importance of digital quality measurement across its priority initiatives (e.g., the Digital Quality Measures [dQM] Strategic Roadmap and the Meaningful Measures Framework). 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 while reducing measurement burden. However, PRO-PMs compose less than 7 percent of all NQF-endorsed quality measures. To increasingly incorporate the patient perspective, measure developers need meaningful resources to help them navigate the development of digital PRO-PMs.

This Technical Guidance Report (henceforth referred to as the *Roadmap*) is a resource for measure developers who are developing digital PRO-PMs. It compiles existing resources on measure development, endorsement, and interoperability and frames them through the lens of patient-reported performance measurement. NQF convened a Technical Expert Panel (TEP) (Appendix A) to inform the Roadmap, which consists of four stages composed of 17 tasks:

- **Stage 1: Definition of Measurement Goals.** Measure developers set goals for the measure concept and the development of a PRO-PM.
- **Stage 2: Exploration and Assessment of Patient-Reported Outcome Measures.** Measure developers identify, assess, and select the patient-reported outcome measure(s) (PROM[s]) that will collect data for the PRO-PM.
- **Stage 3: Development and Testing of the PRO-PM.** Measure developers specify the details of the measure, select test sites, and create and execute testing plans.
- **Stage 4: Finalization and Implementation of the PRO-PM.** Measure developers prepare for implementation and the Consensus Development Process (CDP).



The Roadmap is designed so that measure developers can complete recommended tasks within or across stages, ensuring consistent but flexible guidance to individuals and organizations that approach measure development differently.

This Roadmap is an updated version of the original Roadmap published by NQF in November 2021. NQF updated the Roadmap based on a series of key informant interviews (KIIs) with measurement and interoperability experts and guidance from the TEP. Improvements include the following:

- A new stage 1 task titled "Review Relevant Resources," that guides measure developers to a set of source-of-truth resources. The addition of this task ensures the Roadmap will stay relevant and meaningful by linking to permanent webpages (when possible) for general guidance on measure development, NQF endorsement, and interoperability. NQF removed detailed information that mirrored source-of-truth resources.
- Tailored information on elements of development, testing, and endorsement that are unique to digital PRO-PMs.
- Additional examples of how patients can contribute to digital PRO-PM development.
- New information on digital measurement, including references to the Electronic Clinical Quality Improvement (eCQI) Resource Center page on dQMs and an example of how machine-readable and narrative specifications differ.

The Roadmap is not a textbook that will answer every question related to the development of digital PRO-PMs, but rather a primer that will help measure developers understand a complicated process, ask informed questions, and navigate the steps required to develop a high quality PRO-PM. The Roadmap is

a catalyst for building a more robust database of digital PRO-PMs, which will ultimately elevate patient voices and prioritize outcomes that matter to patients.

Introduction

Digital PRO-PMs offer an important opportunity to ensure that patient voices are captured in assessments of quality while supporting learning health systems.[1–3] Policymakers, payers, and healthcare providers are increasingly using PRO-PMs to inform clinical decision making, improve quality of care, modify provider payment, and evaluate the value of medical technologies.[1] This growth aligns with the CMS Meaningful Measures Framework, which includes Meaningful Measures 2.0 (e.g., prioritizing outcome measures and PROs), and the CMS National Quality Strategy goals (e.g., foster engagement between individuals and their care teams, embrace the digital age, and increase alignment of performance metrics across stakeholders).[3–5]

Despite this growth, PRO-PMs make up less than 7 percent of all NQF-endorsed quality measures, and only one PRO-PM was included on CMS' list of Merit-Based Incentive Payment System (MIPS) quality measures for 2022.[6,7] Challenges hamper the broad adoption of these measures. 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 understand their benefits. Patients lack awareness about the benefits of PROMs and PRO-PMs, such as the ability to compare outcomes across different clinicians, hospitals, and health plans. PRO-PMs are complex measures, and measure developers lack thorough yet accessible technical guidance to develop high quality performance measures based on patient-reported data.[8,9]

Goals and Objectives

The Roadmap aims to provide measure developers with thorough yet accessible guidance on developing digital PRO-PMs, from identifying a measure concept to preparing to submit the PRO-PM for NQF endorsement review. The Roadmap is intended for novice and advanced measure developers alike. Although its guidance is generally applicable to all PRO-PMs, it specifically focuses on digital PRO-PMs that are intended for use in CMS' value-based purchasing (VBP) programs and alternative payment models (APMs) and can be calculated and transmitted across interoperable health information technology (IT) systems. dQMs are measures that automatically pull data generated during the normal course of clinical care, and thus allow for more efficient capture, retrieval, and sharing of patient data.[2,10] PRO-PMs are well suited to be dQMs because PROMs often have structured data fields and PRO-PMs may require information from multiple data sources (e.g., PROM data, electronic health record [EHR] data, and claims data).

The Roadmap describes how to develop a digital PRO-PM that is:

- meaningful to patients;
- aligned with best and promising practices;
- appropriate for regulatory purposes;
- usable by public and private payers; and
- consistent with standards of scientific acceptability.

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 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, but rather a companion to use alongside these resources (Appendix B).

Background

Although PRO-PMs make up a small percentage of the measures used in CMS' VBP programs and 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. Following the passage of the Affordable Care Act (ACA) in 2010, CMS adopted new VBP programs that shifted towards improving and rewarding value rather than volume.[11] CMS designed the VBP programs to ensure that healthcare was more person centered by creating care that focuses on patients' preferences and desired outcomes.[11] Several VBP programs apply to various provider settings, such as hospitals and outpatient centers.[11] In addition to VBP programs, CMS utilizes APMs to incentivize eligible participants to provide high quality and cost-efficient care.[11] 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.[11]

CMS has a long-standing commitment to PRO-PMs and has partnered with NQF for more than 10 years to advance the knowledge of patient-centered performance measurement through a series of foundational reports. This body of work includes, but is not limited to, the following:

- 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.)[12]
- 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*)

The Roadmap is the final report in the Building a Roadmap initiative, which is the most recent PRO-PM collaboration between CMS and NQF and includes four distinct reports:

1. The **Environmental Scan Report** identifies and summarizes existing information relevant to the use of high quality PROMs as the basis for PRO-PMs in accountability programs. It is the first report in the Building a Roadmap initiative and provides background for the subsequent publications. The report was originally published in spring 2021 and updated in June 2022.
2. The **Interim Report** identifies and describes 12 attributes of PROMs that are suitable data collection instruments for high quality digital PRO-PMs. It provides guidance to measure developers on selecting PROMs for use in a PRO-PM and is best read alongside the Roadmap.
3. The **Developer Feedback Report** summarizes improvement opportunities for the Roadmap from measure developers and health IT experts.
4. The **Technical Guidance Report** describes a series of four stages and 17 tasks that measure developers can follow when developing and testing digital PRO-PMs that are suitable for CMS VBP programs and APMs. The Roadmap was originally published in autumn 2021 and updated in in autumn 2022.

Methodology

NQF began development of the Roadmap by convening a multistakeholder TEP that met 14 times from 2021 to 2022. The TEP comprised diverse viewpoints that represented patients, patient advocates, measure developers, health IT professionals, clinicians, health systems, payers, purchasers, and researchers. Due to the technical nature of the Roadmap, NQF intentionally included multiple PROM developers and PRO-PM developers on the TEP. These individuals acknowledged their work during the disclosure of interest portion of the TEP's first web meeting, which was open to the public. In addition to the TEP, liaisons from federal agencies provided guidance on PRO-PMs from the federal perspective. Lastly, 12 experts participated in KIIs that identified improvement opportunities for the Roadmap.

Unless a fact or recommendation is explicitly attributed to a specific source, information in the Roadmap comes from the TEP and the KIIs, 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.

Terminology

The definitions for PRO, PROM, and PRO-PM used in the Roadmap (Table 1) align with the definitions used in the CMS Measures Management System Blueprint (henceforth referred to as the *CMS Blueprint*).[9] Because the terminology regarding PRO-PMs is highly technical, the report includes a glossary of key terms (Appendix C).

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)[9]	Symptoms of depression
Patient-Reported Outcome Measure (PROM)	How PROs are measured. The tools/instruments used to collect data[9]	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[9]	Percentage of patients with a diagnosis of major depression or dysthymia and an initial PHQ-9 score greater than 9 with a follow-up PHQ-9 score less than 5 at six months (NQF #0711)

The Roadmap focuses on digital PRO-PMs, which are a type of dQM. The Roadmap also discusses electronic clinical quality measures (eCQMs).

- **Digital quality measures:** The definition of dQMs is evolving but the draft definition is "Quality measures, organized as self-contained measure specifications and code packages, that use one or more sources of health information that is captured and can be transmitted electronically via interoperable systems. Data sources for dQMs may include administrative systems, electronically submitted clinical assessment data, case management systems, electronic health

records, laboratory systems, prescription drug monitoring programs, instruments (for example, medical devices and wearable devices), patient portals or applications (for example, for collection of patient-generated data such as a home blood pressure monitor, or patient-reported health data), health information exchanges, or registries, and other sources."[13]

- **Electronic clinical quality measures:** eCQMs are clinical quality measures expressed and formatted to use data from EHRs and/or health IT systems to measure healthcare quality, ideally data captured in structured form during the process of patient care. For the measured entity to report an eCQM from an EHR, eCQM developers format the Health Quality Measure Format using the Quality Data Model to define the data elements and Clinical Quality Language to express the logic needed to evaluate a provider or organization's performance.[14]

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 is suitable for use in a digital PRO-PM. These attributes are collected in a table (i.e., the Attribute Grid, Appendix D) 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.[15]

Throughout the Roadmap, **the term *measure developer* refers to the person/team that is developing the PRO-PM**. The term *PROM developer* refers to the person/team that developed the PROM. The scenario in which a single person or organization acts as both the PROM developer and the measure developer is not addressed in the Roadmap.

Environmental Scan Findings

The Environmental Scan summarized information relevant to the use of high quality PROMs as the basis for digital PRO-PMs in accountability programs and provided background for the other documents in the Building a Roadmap initiative. Key findings from the Environmental Scan Report included the following:

- **A limited number of NQF-endorsed PRO-PMs exist:** At the time of the Roadmap's publication, only 31 PRO-PMs were actively endorsed by NQF.[7] PRO-PMs make up less than 7 percent of all NQF-endorsed measures.[7] Additionally, the gap between the number of NQF-endorsed PRO-PMs and the limited number currently used for accountability programs constrain CMS' ability to elevate the patient voice. As described in the *Interim Report Recommendations* section of this report, the Roadmap focuses on PRO-PMs that measure health-related quality of life (HRQoL), functional status, and symptoms and symptom burden, in part due to their small representation among the total number of PRO-PMs. 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.[7]
- **PRO-PMs can use data from one PROM or many different PROMs:** While most NQF-endorsed PRO-PMs collect data from a single PROM, there can be benefits to developing performance measures that allow for the use of different PROMs to measure the quality concept of interest. Advantages and disadvantages to both approaches are discussed in the stage 2 task titled "Select PROMs for Inclusion in the PRO-PM."

- **Digital PRO-PMs align with CMS' priorities of reduced burden and elevated patient voices:** CMS' Meaningful Measures 2.0 acknowledges the importance of 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), and the reduction of measurement burden on clinicians and patients.[3] These priorities can be addressed, in part, through the use of digital PRO-PMs, which are intended to collect data generated during the normal course of clinical care.[10]
- **Data collection methods are an important consideration in development of PRO-PMs:** Stakeholders agree that modes of PROM administration and methods of data collection are important to consider when developing PRO-PMs. NQF's PRO Best Practices Report focused on the importance of using multiple modes to maximize patient participation and response.[16] Patient-level factors may also impact PROMs.

Environmental Scan Findings on Interoperability

The Environmental Scan Report also identified **the importance of interoperability to the ongoing work to advance data standardization and data sharing, as well as opportunities for measure developers to contribute to national priority setting on advancing interoperability for specific data classes and data elements**. The widespread use of PROMs and PRO-PMs will require improved integration of data collected from patients with data from EHRs and other health IT and data exchange systems that support quality measure calculation and public reporting. This can be achieved through the widespread use of interoperable data standards (e.g., Fast Healthcare Interoperability Resources [FHIR]) and standardized coding schemes (e.g., Logical Observation Identifiers Names and Codes [LOINC]).

The U.S. is advancing data interoperability in multiple ways that will lower the burden associated with collecting and sharing data used for quality measurement. CMS recently released its dQM Strategic Roadmap (executive summary and full report), which sets forth how interoperability can be leveraged to advance and reduce the burden of quality measurement. The development and widespread support for FHIR have allowed it to emerge as the leading standard for interoperable data. FHIR is a Health Level Seven International (HL7) standard that defines how healthcare information can be standardized for exchange between different IT systems regardless of how it is stored in those systems.[17] The Office of the National Coordinator for Health Information Technology (ONC) has also advanced the use of FHIR through policy. The 21st Century Cures Act Final Rule defined a minimum interoperable set of data classes and specific data elements: the United States Core Data for Interoperability (USCDI). The rule requires as part of certification criteria that certain healthcare providers maintain access to the full scope of data defined in the USCDI Version 1 through standards-based application programming interfaces (APIs) for certain uses, such as providing patients access to their own data. These requirements will build upon the investments that providers and their vendors make to map their data to FHIR resources and make those data accessible in the FHIR standard.

ONC will expand the scope of available data over time through a USCDI Standards Version Advancement Process (SVAP). While the availability of standardized data relevant to PROMs and PRO-PMs will likely be minimal initially, this will evolve as USCDI is updated and ONC adopts new versions for certification requirements. For example, Version 3 standards, which have not yet been proposed for incorporation into certification criteria, include a new class of data called Health Status/Assessment, which includes functional status data elements. These additions were supported by the quality measurement

community seeking to advance digital PRO-PMs. In addition, ONC announced in 2021 that it is developing a set of standards to support a number of federal partners, including CMS and other agencies, implementing quality measurement and related use cases.[18] This initiative will support the development of harmonized data sets for FHIR-based quality reporting. ONC is committed to engaging private as well as public stakeholders in this process.

ONC's current structured process for updating USCDI and ONC, as well as other federal agencies' new work on USCDI+, creates opportunities for measure developers to advance interoperability for the measures they are developing by identifying and advocating for inclusion in both national core data sets and data elements as well as FHIR specifications. Information on how to submit data classes and elements for inclusion in future USCDI updates can be found in the USCDI ONC New Data Element and Class (ONDEC) Submission System.

In summary, advances in interoperability will ease data sharing, reduce clinical and administrative burden, create the infrastructure and specificity of standardized data and shared data formats needed to aggregate data across measured entities, and allow for the use of shared data in centralized measure score calculations. Measure developers must stay abreast of these advances as they design and implement digital PRO-PMs and can contribute to setting priorities for advancing the interoperable data needed for measurement.

Interim Report Recommendations

The Interim Report describes the attributes of high quality PROMs. The TEP and NQF recommend that measure developers use the Interim Report alongside the Roadmap. The goal of the Interim Report is to guide measure developers in selecting high quality PROMs that will be most conducive to the development and use of a digital PRO-PM. Key components of this report include the attributes of a high quality PROM, an Attribute Grid to facilitate analysis of PROMs, and prioritized PRO-PM domains.

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 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 D). The attributes identified in the Interim Report are as follows:

- 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)

Although the TEP recognizes the attributes in the Interim Report as being important indicators of high quality PROMs, the Attribute Grid is 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 assess any subset of attributes important to them and add additional attributes that are pertinent to a PRO-PM in development.

Prioritized PRO-PM Domains

PRO-PMs can be grouped within five domains, or areas of patient-reported health status: HRQoL, functional status, symptoms and symptom burden, health behaviors, and patient experience.[12] 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:

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 underrepresented.[7]
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.[19] 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 within the patient experience domain, such as the Consumer Assessment of Healthcare Providers and Systems (CAHPS), utilize methodologies that significantly differ from other domains. To help prevent retaliation against patients who report poor experiences with a provider or health system, patient experience measures typically depend on external partners to confidentially collect and analyze patient-reported data using mixed modes and methods of data collection (e.g., self-administered mail and/or web questionnaires, with telephonic questionnaires for patients who do not respond).[20] As a result of these methodological differences, the patient experience domain may not be as applicable to digital PRO-PMs as other domains.

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.

Roadmap for the Development of a Digital PRO-PM

Overview and Visualization of the Roadmap

The Roadmap provides guidance on the development of digital PRO-PMs that are intended for NQF endorsement and suitable for use in CMS' VBP programs and APMs. The Roadmap is organized into four stages of development:

- 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

A series of 17 tasks occur within the four stages, and the measure developer should address each task during the development process. 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 17 tasks that can shift across stages. This reflects the flexibility that measure developers may take when developing a PRO-PM.

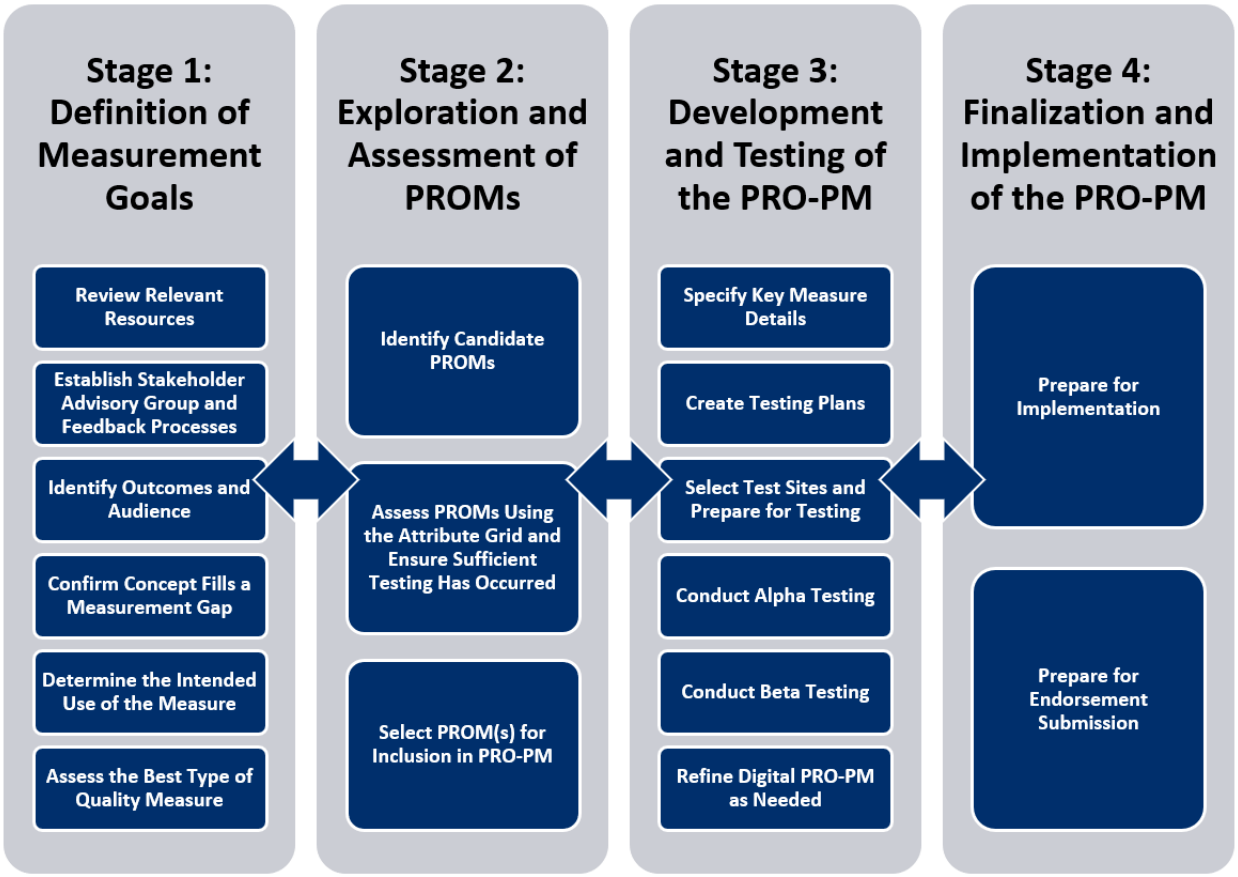


Figure 1. PRO-PM Roadmap: Each column contains one stage, and the bidirectional arrows indicate that tasks can move freely and be iterated across stages.

In the same way that an actual roadmap offers many routes between two locations, the digital PRO-PM Roadmap recognizes that there are many routes to get from identifying the need for a measure to submitting a fully tested digital PRO-PM for NQF endorsement review. While the four stages generally

occur in the sequence listed above, the 17 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. 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 designing 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 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, the TEP recognizes that many tasks related to digital PRO-PM development are iterative and may repeat across multiple stages. For example, 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 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 becomes familiar with key reference documents, then assembles a stakeholder advisory group that is composed of multiple expert perspectives. The stakeholder advisory group helps to 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. The tasks within stage 1 establish a clear plan for the entire measure development life cycle. Execution of the tasks within stage 1 can influence the success of the entire digital PRO-PM development life cycle.
- **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 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 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, 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 CDP depending on whether the PRO-PM will be submitted for NQF endorsement review. The preparation for

tasks in stage 4 will likely begin early in the development process, even though many of 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 review.

It is critical for readers of this report to understand that **the Roadmap is a guide for development, 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, early planning and preparation can improve the entire measure development life cycle. A measure developer should complete these tasks before beginning the development of a digital PRO-PM. This early work helps to ensure the measure will assist in improving the desired outcomes. While the measure developer may decide to move other tasks into stage 1, particularly those that will benefit from early planning or iterations, the six tasks listed below should be completed before moving beyond stage 1.

Review Relevant Resources

The TEP designed the Roadmap as a high-level guide that measure developers can use alongside source-of-truth resources on quality measures, PRO-PMs, and interoperability. Developers who are working on digital PRO-PMs should familiarize themselves with these resources, which are developed and maintained by organizations with expertise on specific aspects of measure development. Developers should reference these resources throughout the four stages of digital PRO-PM development. The Roadmap contains a comprehensive list of recommended resources to aid with the development of digital PRO-PMs (Appendix B).

Establish Stakeholder Advisory Group and Feedback Processes

The creation of the stakeholder advisory group 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 important 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, or subgroups within a main group, 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.

Measure developers might consider creating a patient-focused subgroup to address issues that are pertinent to patients, families, and caregivers. A patient-focused subgroup likely spends less time on technical facets of measure development and more time helping the measure developer ensure the measure is meaningful and valuable to patients. Tasks for which measure developers might engage a

patient-focused subgroup include the identification of outcomes and audiences (stage 1, task 3), the exploration and assessment of PROMs (stage 2), and the selection of test sites (stage 3, task 3). If the measure developer creates a patient subgroup, it should regularly engage with the full stakeholder advisory group.

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 PRO-PM's measurement goals, a balanced stakeholder advisory group might also include clinicians, administrative staff (e.g., staff who will 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 other stakeholders who will be directly affected by the PROM completion and collection process and/or the PRO-PM. Measure developers should consider including PROM developer(s) in the stakeholder advisory group, although this might not be possible until PROM(s) are selected during stage 2.

Health equity is the attainment of the highest level of health for all people, and it warrants consideration throughout the measure development process.[21] This consideration begins with the stakeholder advisory group. During this task, measure developers can determine other stakeholder perspectives that are necessary to add to the group to ensure the PRO-PM is equitable. Depending on the measure's outcomes and audience, members of the stakeholder advisory group could include patients and/or clinicians from rural settings, non-English speakers who are in the measure's target population, people who have physical or cognitive disabilities, patients and/or clinicians with conditions or diagnoses targeted by the measure (e.g., people with depression or who had hip replacement surgery), and other representatives from vulnerable or underserved populations.

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, with complex measures requiring more input.

Identify Outcomes and Audience

The measure developer, with guidance from the stakeholder advisory group, should identify a consensus definition for the measure concept (i.e., a high-level recommendation that includes a clear description of what should be measured). The concept should include 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 use empirical data, literature, and/or guidelines to determine and support the measure concept and target population.

PRO-PMs must be meaningful to patients and/or caregivers. While NQF does not have prescriptive requirements for demonstrating face validity, measure developers need to provide evidence that the

measure is meaningful to its target population. The measure developer has multiple opportunities to gather this feedback. The stakeholder advisory group and any patient-focused subgroups might participate in these activities, and the measure developer may also choose to engage members of the target population through surveys, focus groups, or other information-gathering activities. These activities help the measure developer to empirically demonstrate that the measure concept is meaningful to patients.

Confirm Concept Fills a Measurement Gap

A new PRO-PM must address a measurement gap. It might measure an important issue with significant variation that is not currently measured, or it might be an improvement to an existing measure. A measure developer can determine whether a measurement gap exists by scanning 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 similar measures and ensure the definition, outcome, and audience are either new or significantly improve upon existing measures. By conducting an environmental scan of measures, the measure developer can avoid duplicating existing work and adding unnecessary burden on measured entities. If a similar measure(s) already exists, the measure developer can consult with the stakeholder advisory group as well as the steward of the existing measure to determine whether the concept justifies another measure.

Determine the Intended Use of the Measure

Different measures serve different purposes. Some performance measures are used solely for quality improvement, while others might be used for accreditation, comparison of provider performance, or payment in APMs or VBP programs. Due to the time, effort, and cost required for developing a digital PRO-PM, the measure developer and the stakeholder advisory group should determine how the measure will be used before moving beyond stage 1. Development and testing processes will vary depending on this determination, and requirements are more stringent for accountability than for quality improvement.

Table 2: Distinctions in Testing PROMs and PRO-PMs

Distinction	PROM	PRO-PM
Who is responsible for testing?	The PROM developer is responsible for testing a PROM.	The measure developer is responsible for testing a PRO-PM.
Where is this discussed in the report?	PROM testing is discussed further in stage 2.	PRO-PM testing is discussed further in stage 3.

Distinction	PROM	PRO-PM
What additional resources exist?	• The 2013 NQF Committee Final Report, Patient-Reported Outcomes in Performance Measurement • The 2015 report by David Cella et al, also titled Patient-Reported Outcomes in Performance Measurement • The 2021 Building a Roadmap Interim Report on high quality PROMs for use in PRO-PMs	• NQF Measure Evaluation Criteria webpage • NQF Measure Evaluation Criteria and Guidance for Evaluating Measures for Endorsement (PDF, September 2021) • The CMS Blueprint Measure Lifecycle Overview webpage • The CMS Blueprint QuickStart Guide (PDF, May 2022)

Assess the Best Type of Quality Measure

The measure developer will need to determine which type of measure—outcome, process, or structure—is best suited to achieve the identified measurement goals. If an outcome measure is appropriate, the measure developer will then ascertain whether the goals are best achieved by a digital PRO-PM. A PRO-PM is a type of outcome measure that is based on PROs, assessed through data collected via a PROM, and aggregated for an accountable healthcare entity.[22] 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.

Developing PRO-PMs, and specifically digital PRO-PMs, can be complex, time consuming, and expensive. Once the measure developer has determined that the measure concept should proceed as a digital PRO-PM, early planning on several tasks in stages 2, 3, and 4 can help mitigate challenges and delays later in the development process. Some important questions to begin considering now include the following:

- What health equity issues affect the target population (e.g., language requirements and access to broadband)? (stages 2 and 3)
- Will data collection be limited to a single PROM, or can different PROMs collect data for the PRO-PM? (stage 2, task 3)
- What is the measure's attribution model? (stage 3, task 1)

- How will the measure be risk-adjusted? (stage 3, task 1)
- What sampling methodology might ensure a representative sample, and what strategies might lessen the chances of a low response rate? (stage 3, task 1)
- Which test sites might be appropriate for beta testing? (stage 3, task 3)
- What is the potential testing plan for alpha testing? (stage 3, task 4)
- When will the measure developer have adequate information to complete the Intent to Submit process? (stage 4, task 2)

Stage 2: Exploration and Assessment of PROMs

Once the initial strategic work is completed, the measure developer can begin an objective assessment and selection process to determine which PROM(s) are the most appropriate data collection tools for the PRO-PM. 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. The Roadmap focuses its guidance on measure developers who are selecting from existing PROMs rather than those who are working with PROM developers to create a de novo PROM for use with a PRO-PM, although the attributes of a high quality PROM are applicable to either scenario.

The Interim Report and the Roadmap are designed to be used together, and the Interim Report provides more 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. When identifying candidate PROMs, measure developers should keep stage 1 findings top of mind. For example, if stage 1 assessments determined that the PROM must have a validated computerized adaptive testing (CAT) version available in both English and Spanish in order to reduce burden and inequities for the target population, the measure developer can immediately disregard PROMs that do not meet those initial requirements.

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) sets of Patient-Centered Outcome Measures provide lists of recommended PROMs that have been vetted by diverse expert panels.
- The Consensus-Based Standards for the Selection of Health Measurement Instruments (COSMIN) website contains extensive resources on identifying candidate PROMs.
- 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 D; 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. 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/or consumers from the stakeholder advisory group will be important participants in 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 should be engaged during this process.

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 existing literature can provide guidance in interpreting the results. For example, the 2012 NQF Methodological Issues Report 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.[23] 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, as well as previous NQF reports.

Many elements of testing—including psychometric soundness and mode of administration—are reviewed using the Attribute Grid. However, the NQF endorsement process for a digital PRO-PM could require a PROM to undergo additional testing. For example, endorsement of a PRO-PM that is also an eCQM will require validity testing for each question of a PROM, and the endorsement process will scrutinize any substantial differences in validity across questions. As a result, the measure developer should 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 tests of psychometric soundness 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 PROM developer to determine whether additional information is available. The interdependent relationship between PROMs and PRO-PMs is why the measure developer should attempt to engage the PROM developer as a member of the stakeholder advisory group.

Select PROM(s) for Inclusion in PRO-PM

Based on the analysis of the Attribute Grid, the measure developer should determine which PROM(s) best meet the needs of the PRO-PM. During this task, the measure developer and stakeholder advisory group will determine whether the PRO-PM will utilize data from a single PROM or from multiple different PROMs. There are advantages and disadvantages to either approach:

- **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 (e.g., a PROM with multiple validated translations).
    - 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., HRQoL, symptoms and symptom burden, or health functioning); however, a crosswalk must exist or be developed to ensure the different PROMs consistently assess quality (e.g., harmonized cut points and scoring).

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. Measure development involves designing a quality measure that is suitable for assessing the performance of the accountable entity; it includes defining the care targeted for measurement, the level of provider being evaluated, the approach to attributing patients to providers, and any model development required to adjust for difference in case mix across providers. Measure testing involves formally evaluating the measure's reliability and validity for use as a quality measure.

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 at which point the measure

developer initiates these tasks, each must be addressed and iteratively updated as new information is gained throughout the development process. If the measure developer and the stakeholder advisory group determine that certain tasks are unnecessary for a specific PRO-PM, the measure developer should thoroughly document the rationale for these decisions in the PRO-PM's supporting documentation.

Specify Key Measure Details

The technical specification for a PRO-PM identifies the information and data that are needed to generate the measure result. This includes 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). Measure specifications provide technical instructions on how the measure should be collected and used consistently, reliably, and effectively.[24] Developing measure specifications is an iterative process that requires a variety of stakeholder perspectives.[24] 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.

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
- Attribution model
- Risk adjustment
- Time intervals
- The calculation algorithm[24]

Specify Information for Related PROMs

Technical specifications for PRO-PMs require information about the PROM(s) that are used as data collection instruments. The measure developer will access some of this information during the stage 2 tasks, but other details may be needed, such as acceptable modes of PROM administration, instructions for recording PROM results that are blank or contain multiple responses to a single question, and the definition of a completed PROM. Although the PROM developer has already validated the PROM(s), the measure developer must ensure that the measure specifications are detailed enough to ensure that the collection of PROM data is accurately informing the PRO-PM.

Specify Information for Digital PRO-PMs

The following examples present a side-by-side comparison of a digital (i.e., machine-readable) specification and a narrative (i.e., human-readable) specification for a quality measure. The examples compare the "Initial Patient Population" sections of CMS #146 *Appropriate Testing for Children with Pharyngitis*. Links are provided to specifications on the HL7 Clinical Quality Language (CQL) website and the CMS Quality Payment Program website.

Machine-readable format:

```
* ======================================================================
* Initial Patient Population =
*     AND: "Patient Characteristic Birthdate: birth date" >= 2 year(s) starts before start of "Measurement Period"
*     AND: "Patient Characteristic Birthdate: birth date" < 18 year(s) starts before start of "Measurement Period"
*     AND:
*         AND: "Occurrence A of Encounter, Performed: Ambulatory/ED Visit" during "Measurement Period"
*         AND: "Medication, Order: Antibiotic Medications" <= 3 day(s) starts after start of "Occurrence A of Encounter, Performed: Ambulatory/ED Visit"
*         AND:
*             OR: "Occurrence A of Encounter, Performed: Ambulatory/ED Visit" during
*                 OR: "Occurrence A of Diagnosis, Active: Acute Pharyngitis"
*                 OR: "Occurrence A of Diagnosis, Active: Acute Tonsillitis"
*             OR:
*                 OR: "Occurrence A of Diagnosis, Active: Acute Pharyngitis"
*                 OR: "Occurrence A of Diagnosis, Active: Acute Tonsillitis"
*                 starts during "Occurrence A of Encounter, Performed: Ambulatory/ED Visit"
*
```

Human-readable format:

Initial Population: children 3-18 years of age who had an outpatient or emergency department (ED) visit with a diagnosis of pharyngitis during the measurement period and an antibiotic ordered on or three days after the visit

eCQMs (including PRO-PMs) have unique endorsement criteria that are described in NQF's Measure Endorsement Criteria and Guidance. These criteria include providing machine-readable specifications using the latest accepted versions of eCQM technical specifications (i.e., Health Quality Measure Format [HQML], Quality Data Model [QDM], Clinical Quality Language [CQL], and FHIR) and value sets vetted through the National Library of Medicine's (NLM) Value Set Authority Center (VSAC). Developers should test and verify the behavior of their digital measure logic using a constructed patient test deck. The publicly available BONNIE tool supports testing of the accuracy of the measure logic. Developers must also include a narrative measure specification (i.e., a human-readable document) along with the machine-readable specifications. Figure 2 illustrates machine-readable and human-readable specifications.

As described in the section on interoperability, FHIR will facilitate the use of dQMs by advancing data standards. To date, however, eCQMs are primarily developed and reported using the QDM rather than FHIR. CMS and others are working to streamline the specification process and offer enhanced tools for measure specification development and testing.[2,14,17] NQF is considering its role in developing standards for evaluating FHIR-based measures. Measure developers should align efforts with current NQF, CMS, and/or ONC guidelines (e.g., the dQM Strategic Roadmap).

Determine Appropriate Attribution Model

The measure developer must answer an important question during the digital PRO-PM development process: For which providers is the patient outcome a signal of the quality of care? Attribution is used in quality measurement to assign accountability for a patient's outcomes to the accountable entity being assessed by the measure; in other words, the model assigns specific patients to the specific parties being measured (e.g. a clinician, group of clinicians, facility, or accountable care organizations [ACOs] and

health plans).[15] An *attribution model* is a set of rules that defines the accountable unit for a patient's healthcare outcomes. It may utilize, for example, patient attestation of who their provider is or analyses of claims data. Attribution models are needed for PRO-PMs because patients may acquire healthcare related to the measured outcome from multiple providers serving various payers; an attribution model specifies how patients will be assigned to providers when the measure is implemented to ensure that providers' measure scores will reliably reflect patients the measured providers cared for and outcomes the providers had the ability to influence.[15] NQF's Attribution: Principles and Approaches Report provides an Attribution Model Selection Guide as a useful resource for determining the appropriate model selection. The NQF report on Attribution for Critical Illness and Injury offers additional guidance on geographic/population-based attribution models.

During this task, measure developers should engage pertinent advisory stakeholders to discuss effects on accountable entities. This will help clarify the connection between the PRO-PM data and the accountable entity's performance.

Develop and Test Risk Adjustment Model

For outcome measures, including PRO-PMs, NQF endorsement requires an evidence-based risk adjustment strategy to account for differences in case mix across measured entities that affect the risk of the outcome independent of quality or an explanation of why risk adjustment is not required. Patients may have demographic, clinical, and social risk factors that vary widely across the measured providers, such as hospitals or clinician groups, and need to be considered for risk adjustment in model development. Developers must provide for the risk model a demonstrated conceptual and statistical rationale, which considers patient factors that are outside of the provider's control, which 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 the risk adjustment model.[25] 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.[25,26]

Guidance on risk adjustment modeling is evolving. Risk-adjusting performance measures, namely outcome and cost/resources use measures, has traditionally accounted for differences in patient health status and clinical factors (e.g., comorbidities or severity of illness) that are present at the start of care. This approach to risk adjustment has been widely accepted and implemented within measure development.[27,28] However, patients can also bring certain social characteristics (e.g., income, education, housing instability, food insecurity, and urbanicity/rurality) into their engagement with the healthcare system, which can influence healthcare outcomes, and that are outside the provider's control. The idea of incorporating these social factors into risk adjustment models for quality measurement has been debated, in brief, due to competing concerns: On one hand, not adjusting may be unfair to providers caring for more patients with these risk factors, while on the other hand, adjusting could lead to potential unintended consequences, such as masking disparities and institutionalizing differing health outcome expectations for different patient groups.

NQF is publishing updated recommendations and analyses of best practices for risk adjustment models in late 2022 on the Risk Adjustment Guidance project page.

Sampling Methodology

The sampling methodology describes the approach to identifying a representative sample of accountable entities that will be measured by the PRO-PM and quantifying the implications of nonresponse rates observed in testing. Although nonresponse is a consideration for PRO-PMs that does not apply to most other types of performance measures (e.g., measures based on laboratory values are not dependent on a patient voluntarily completing a questionnaire), measure developers must assess nonresponse rates for PRO-PMs because they can bias the measure score. These concerns can be mitigated with thoughtful planning of a sampling strategy for the PRO-PM with the goal of establishing a minimal sample size that supports a reliable measure score. The sampling strategy should identify adequate numbers, approaches for randomization, and be representative of the intended patient population. NQF does not provide prescriptive rules for identifying a sampling methodology, so the measure developer needs to carefully consider and prepare to illustrate how risks of bias are addressed.

Create Testing Plan(s)

Comprehensive and accurate testing plans that closely follow NQF's measure evaluation criteria are an important aspect of developing measures. To ensure practical and logistical success of the digital PRO-PM, the test plan should identify and address the goals of the testing process. The testing plan should identify data requirements, including preliminary strategies for addressing issues of 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. Separate test plans are created for alpha testing and beta testing. Alpha testing, or formative testing, determines the feasibility of testing and implementing the draft specifications; subsequently, this testing plan is usually prepared early in the measure development life cycle. Beta testing, or field testing, expands feasibility testing and also assesses scientific acceptability and usability. 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. Testing plans for beta testing 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)[29]

Select Test Sites and Prepare for Testing

Beta testing is a resource-intensive and expensive process that is critical to the success of the digital PRO-PM. Although beta testing is listed as a stage 3 task, the measure developer should plan for and coordinate beta testing of the digital PRO-PM early in the development process.

Because PRO-PMs rely on data that are voluntarily provided by patients, selecting PRO-PM test sites carries unique challenges. Test sites are hesitant to accept the burden of testing PRO-PMs that might not be endorsed or adopted. Measure developers involved with the Building a Roadmap initiative identified this as one of the most significant challenges that faces PRO-PM developers. Unlike testing a new performance measure that depends on information that is routinely collected during clinical care, collecting PROs may require new clinical workflows, staff training, patient education, and/or IT infrastructure. Strategic approaches to this challenge do exist nonetheless:

Identify candidate test sites that are already using the PROM(s): Once the measure developer completes the stage 2 task of identifying PROM(s) for data collection, begin surveying potential test sites to identify who is already using the PROM as part of their established clinical workflows. By focusing on test sites that are already using the PROM, the burdens of implementing a new PROM are significantly reduced, if not eliminated. Measure developers and PROM developers can collaborate in this effort, which reinforces the benefits of including PROM developers on the stakeholder advisory group. Journal articles about the PROM can also help measure developers identify individuals and clinics who have already adopted a PROM. Notably, digital PRO-PMs must be tested in at least two sites with at least two different EHRs. If the measure developer is only able to identify one suitable test site, other strategies in this section might apply.

Collaborate with partners to incentivize test sites: Some stakeholders in the measure development process, such as federal agencies, EHR vendors, private companies, or foundations, can be well positioned to develop unique methods of incentivization for test sites. As an example, an EHR vendor who is seeking a competitive advantage in the market might be willing to incur costs related to building a digital PROM in the EHR. Federal agencies such as CMS can be uniquely positioned to incentivize test sites that participate in VBP programs, APMs, or other innovative payment programs. Private companies or foundations may be able to provide funding to offset the costs of training staff and implement clinical workflows that support the PROM. While stakeholders need to be aware of and avoid potential conflicts of interest, incentives can help alleviate the burden that test sites face.

Explore creative solutions to reduce the burden of implementation: Measure developers are well positioned to find solutions that will lower barriers resulting from training staff, educating patients, and developing IT infrastructure for a new PROM. As an example, a measure developer could offset the burden of building a PROM into the test site's EHR by designing a spreadsheet-based method to analyze and aggregate PROM data and providing staff to capture and manage the data. A different approach might be to utilize PROM data from existing registries under a data use agreement (DUA) or business associates agreement (BAA). A third approach could be to engage a vendor who administers PROMs and stores patient responses without affecting the test site's workflows or data systems. While creative solutions such as these might not meet the testing requirements of endorsing a digital PRO-PM, they might enable the measure developer to apply for trial use under NQF's trial use criteria for eCQMs, as described below.

Consider NQF Trial Use endorsement of a digital PRO-PM: The NQF Trial Use program recognizes the testing burden of eCQMs that are ready for implementation but cannot be adequately tested to meet NQF endorsement criteria. A digital PRO-PM with Trial Use Approval receives a three-year period for additional testing. While not every eCQM will proceed to endorsement (some fail due to the results of testing during the Trial Use period), it is an important program that can offset the burden of testing digital PRO-PMs.

Patients are valuable partners in this task. Their insights and experiences are critical on topics related to workflows regarding PROM administration that affect testing. For example, patients who are part of the target population are more likely than clinicians to understand issues related to social determinants of health and health equity (e.g., barriers that prevent patients from completing PROMs digitally via patient portals or smart phones) or convenience (e.g., challenges with arriving early for appointments to fill out PROMs in the waiting room).

Conduct Alpha Testing

Alpha testing is an initial assessment that allows the measure developer to determine the feasibility of assembling the valid and reliable data required for the measure and calculating the measure score using the measure calculation software. 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."[25] Feasibility should be demonstrated by a data collection strategy that can be reasonably implemented.[30] 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.[25] 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.[29] 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 of the measure calculation software to demonstrate that the measure logic will work (e.g., BONNIE testing for eCQMs). Although the current state of PROM data collection can create feasibility challenges, the measure developer can mitigate these risks by using the Attribute Grid in the Interim Report (Appendix D) 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 more extensive than alpha testing and 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 can 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.[30]

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.[25] These criteria determine whether the PRO-PM, as specified, produces consistent (i.e., reliable) and credible (i.e., valid) results.[25] **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 concepts.**

Reliability comprises two subcriteria, and validity comprises six subcriteria, which are described in detail in NQF's Measure Evaluation Criteria on the Submitting Standards webpage.

Evaluate PRO-PM for Usability and Use

Criteria on usability and use assess the extent to which potential audiences (e.g., patients, clinicians, health plans, and policymakers) are using or could use the results of the performance measure to achieve the goal of delivering or choosing high quality, efficient healthcare. Details on testing for usability and use are also addressed in NQF's Measure Evaluation Criteria on the Submitting Standards webpage.

As with many of the tasks in stage 3, evaluation for scientific acceptability and usability and use is a complex process. The measure developer should begin planning for beta 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 the 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 low reliability and/or validity, or the stakeholder advisory group expresses 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.[25]

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 review. 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 is responsible for developing implementation guidance and should consider using 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

Measures endorsed by NQF have undergone careful evaluation through a multistakeholder consensus-building endorsement process, a process designed intentionally to garner highly diverse stakeholder perspectives. These stakeholders consist of doctors, hospitals and other healthcare providers, employers, health plans, public agencies, community coalitions, patients, and caregivers. This process ensures all NQF measures meet thorough standards for performance measures.

Preparing for NQF endorsement review is specific to the measure steward or developer. As referenced in stage 1 of this report, the measure developer should determine whether they intend to submit the PRO-PM for NQF endorsement review 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 review.

If the measure developer and the stakeholder advisory group determine in stage 1 that NQF endorsement is appropriate for the digital PRO-PM, it is essential that the measure developer understand the NQF Intent to Submit process. Developers who have guided PRO-PMs through the endorsement process note that Intent to Submit is an intensive process that should be started as early as possible. The Submitting Standards webpage provides an overview of the Intent to Submit process and links to an Intent to Submit checklist.

The Roadmap is not a comprehensive resource for seeking NQF endorsement. 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 website.

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

The Roadmap is both a guide for measure developers and a catalyst to elevate patients' voices. It is a guide because it offers measure developers straightforward information on the unique aspects of developing PRO-PMs that are suitable for NQF endorsement and use in CMS' VBP programs and APMs. It is a catalyst because it will facilitate the creation of additional PRO-PMs that measure what matters to patients, using data that patients provide.

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 17 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 questions, and ultimately aid in building a more robust database of digital PRO-PMs.

References

1. Discern Health. Patient-Reported Outcome Performance Measures: Current Environment And Next Steps. *Pharm Res Manuf Am*. Published online December 2017:23.

2. eCQI. Digital Quality Measures. dQMs - Digital Quality Measures. Published August 23, 2022. https://ecqi.healthit.gov/dqm

3. Centers for Medicare & Medicaid Services (CMS). Meaningful Measures 2.0: Moving from Measure Reduction to Modernization. Published June 10, 2022. Accessed August 24, 2022. https://www.cms.gov/meaningful-measures-20-moving-measure-reduction-modernization

4. Centers for Medicare & Medicaid Services (CMS). CMS National Quality Strategy. What is the CMS National Quality Strategy? Published June 10, 2022. Accessed August 12, 2022. https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/Value-Based-Programs/CMS-Quality-Strategy

5. Centers for Medicare & Medicaid Services (CMS). Meaningful Measures Framework. Meaningful Measures Framework. Published 2022. Accessed August 12, 2022. https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/QualityInitiativesGenInfo/CMS-Quality-Strategy

6. Centers for Medicare & Medicaid Services (CMS). MIPS Explore Measures - QPP. Quality Payment Program. Accessed March 28, 2022. https://qpp.cms.gov/mips/explore-measures?tab=qualityMeasures&py=2022

7. NQF. *NQF: Quality Positioning System*.; 2022. https://www.qualityforum.org/QPS

8. Weldring T, Smith SMS. Patient-Reported Outcomes (PROs) and Patient-Reported Outcome Measures (PROMs). *Health Serv Insights*. 2013;6:61-68. doi:10.4137/HSI.S11093

9. CMS. *Supplemental Materials to MMS Blueprint: Patient Reported Outcome Measures*.; 2022:9. https://www.cms.gov/files/document/blueprint-patient-reported-outcome-measures.pdf

10. CMS. *Digital Quality Measurement Strategic Roadmap*. CMS; 2022:65. https://ecqi.healthit.gov/sites/default/files/CMSdQMStrategicRoadmap_032822.pdf

11. National Quality Forum (NQF). Building a Roadmap From Patient-Reported Outcome Measures to Patient-Reported Outcome Performance Measures Environmental Scan. Published online June 2022. https://www.qualityforum.org/Publications/2022/06/Building_a_Roadmap_From_Patient-Reported_Outcome_Measures_to_Patient-Reported_Outcome_Performance_Measures_Environmental_Scan.aspx

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

13. eCQI. What are dQMs? Published August 23, 2022. Accessed August 25, 2022. https://ecqi.healthit.gov/dqm?qt-tabs_dqm=1

14. eCQI. Electronic Clinical Quality Measures. Get Started with eCQMs. Published July 21, 2022. https://ecqi.healthit.gov/ecqms

15. NQF. NQF: Attribution - Principles and Approaches. Published 2016. Accessed June 28, 2021. https://www.qualityforum.org/Publications/2016/12/Attribution_-_Principles_and_Approaches.aspx

16. NQF. *Patient-Reported Outcomes: Best Practices on Selection and Data Collection - Final Technical Report*.; 2020. Accessed August 12, 2022. https://www.qualityforum.org/Publications/2020/09/Patient-Reported_Outcomes__Best_Practices_on_Selection_and_Data_Collection_-_Final_Technical_Report.aspx

17. Office of the National Coordinator for Health Information Technology (ONC). *What Is HL7 FHIR?* Office of the National Coordinator for Health Information Technology (ONC); 2021. Accessed March 14, 2022. https://www.healthit.gov/sites/default/files/page/2021-04/What%20Is%20FHIR%20Fact%20Sheet.pdf

18. Argentieri R, Myers E, Posnack S, Tripathi M. Thinking Outside the Box: The USCDI+ Initiative. Health IT Buzz. Published October 8, 2021. Accessed August 22, 2022. https://www.healthit.gov/buzz-blog/health-it/thinking-outside-the-box-the-uscdi-initiative

19. Meaningful Measures 2.0: Moving from measure reduction to modernization | CMS. Accessed April 12, 2021. https://www.cms.gov/meaningful-measures-20-moving-measure-reduction-modernization

20. Agency for Healthcare Research and Quality (AHRQ). Frequently Asked Questions About CAHPS. Accessed August 12, 2022. https://www.ahrq.gov/cahps/faq/index.html

21. Centers for Disease Control and Prevention. Paving the Road to Health Equity. Paving the Road to Health Equity. Published November 2020. Accessed August 22, 2020. https://www.cdc.gov/minorityhealth/publications/health_equity/index.html

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

23. Cella D, Hahn EA, Jensen SE, Butt Z, Nowinski CJ, Rothrock N. Methodological issues in the selection, administration and use of patient-reported outcomes In performance measurement in health care settings. Published online September 28, 2012. http://www.qualityforum.org/WorkArea/linkit.aspx?LinkIdentifier=id&ItemID=72156

24. CMS. Specification | CMS. Published 2021. Accessed August 12, 2022. https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/MMS/MSP-Specification

25. NQF. *Measure Evaluation Criteria and Guidance for Evaluating Measures for Endorsement*.; 2021. https://www.qualityforum.org/WorkArea/linkit.aspx?LinkIdentifier=id&ItemID=88439

26. NQF. Submitting Standards. Submitting Standards. Published 2022. https://www.qualityforum.org/Measuring_Performance/Submitting_Standards.aspx

27. 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. doi:10.1097/00005650-200208000-00011

28. 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. doi:10.1161/CIRCOUTCOMES.113.000520

29. CMS. CMS Measures Management System Blueprint. Blueprint Measure Lifecycle. Published May 2022. Accessed August 12, 2022. https://mmshub.cms.gov/blueprint-measure-lifecycle-overview

30. NQF. NQF: Measure evaluation criteria. Accessed April 11, 2021. https://www.qualityforum.org/Measuring_Performance/Submitting_Standards/Measure_Evaluation_Criteria.aspx

31. CMS. Alternative payment models (APMs) overview - QPP. CMS.gov. Accessed August 12, 2022. https://qpp.cms.gov/apms/overview

32. McGlothlin AE, Lewis RJ. Minimal clinically important difference: Defining what really matters to patients. *JAMA*. 2014;312(13):1342-1343. doi:10.1001/jama.2014.13128

33. NQF. *Building a Roadmap From Patient-Reported Outcome Measures to Patient-Reported Outcome Performance Measures Interim Report*.; 2021. Accessed August 25, 2022. https://www.qualityforum.org/Publications/2021/08/Building_a_Roadmap_From_Patient-Reported_Outcome_Measures_to_Patient-Reported_Outcome_Performance_Measures_-_Interim_Report.aspx

34. Deutsch A, Smith L, Gage B, Kelleher C, Garfinkel D. *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

35. 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. doi:10.2106/JBJS.19.00916

36. NQF. *NQF Health Information Technology Glossary: A Guide to HIT Jargon*.; 2015. Accessed August 12, 2022. https://www.qualityforum.org/Measuring_Performance/HIT_Glossary.aspx..

37. Regenstrief Institute. LOINC and Health Data Standards. Published 2022. https://www.regenstrief.org/centers/loinc

38. CMS. Measure Management System Glossary. Glossary. Published 2022. Accessed August 12, 2022. https://mmshub.cms.gov/glossary/m

39. NQF. *NQF: Measuring Performance*. Accessed July 25, 2022. https://www.qualityforum.org/Measuring_Performance/Measuring_Performance.aspx

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

41. CMS. Value-based programs. What are value-based programs? Published March 31, 2022. Accessed August 24, 2022. https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/Value-Based-Programs/Value-Based-Programs

Appendices

Appendix A: Committee Members, Federal Liaisons, Key Informant Interviewees, NQF Staff, and CMS 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 (AHRQ)

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

Clifford A. Smith, PhD, ABPP-Cn
U.S. Department of Veterans Affairs (VA), Office of Mental Health and Suicide Prevention

David Au, MD, MS
VA

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

Key Informant Interviewees

Lisa Anderson, MSN, RN-BC
Clinical Informatics Lead, Mathematica

David Clayman, DPM, MBA
Clinical Program Lead, Mathematica

Kyle Cobb, MS
U.S. Department of Health and Human Services, Office of the National Coordinator for Health Information Technology (ONC)

Collette Cole, BSN, RN, CPHQ
Clinical Measurement Developer, Minnesota (MN) Community Measurement

Michelle Dardis, PhD, MBA, MSN
Clinical Quality and Informatics Leader, The Joint Commission

Rebecca Etz, PhD
Cultural Anthropologist, Co-Director of The Larry A. Green Center and Associate Professor of Family Medicine and Population Health at Virginia Commonwealth University

Grace Glennon, MS, RD, CHES
Health Outcomes Researcher, Yale/Yale New Haven Hospital (YNHH) Center for Outcomes Research and Evaluation (CORE)

Ben Hamlin, DrPH, MPH
Senior Research Informaticist, National Committee of Quality Assurance (NCQA)

Laura Jantos, LFHIMSS
Patient Advocate, Healthcare Technology/Digital Health

Samantha Tierney, MPH
Senior Scientist, Clinical Policy, American College of Physicians (ACP)

Sandra Mitchell, PhD, CRNP, FAAN
Senior Scientist and Program Director of the Outcomes Research Branch, NIH, National Cancer Institute

Ashley Wilder Smith, PhD, MPH
Branch Chief of the Outcomes Research Branch, NIH, National Cancer Institute

NQF Staff

Kathleen Giblin, RN
Senior Vice President, Quality Innovation

Elizabeth Drye, MD, SM
Chief Scientific Officer, Measurement Science and Application

Alejandra Herr, MPH
Senior Managing Director

Chuck Amos, MBA
Senior Director

Poonam Bal, MHSA
Senior Director

Kimberly Streett, RN, MSN

Director

Teresa Brown, MHA, MA
Director

Deidra Smith, MBA, PMP
Senior Project Manager

Carina Smith, MPP, MBA
Manager

Evelyn Thomas, MPH, CHES
Specialist

CMS Staff

LaWanda Burwell, ScD
Task Order COR

Gequincia Polk
IDIQ COR

Helen Dollar-Maples
Director, DPMS/QMVIG/CCSQ

Marsha Smith, MD, MPH, FAAP
Medical Officer, DPMS/QMVIG/CCSQ

Appendix B: Relevant Resources, Stage 1 Task 1

NQF Resources

Resources and publications about PROs and PRO-PMs

- The 2012 white paper, PRO-Based Performance Measures for Healthcare Accountable Entities
- The 2012 white paper, Methodological Issues in the Selection, Administration, and Use of Patient-Reported Outcomes in Performance Measurement in Health Care Settings
- The 2015 update to the 2012 Methodological Issues white paper by David Cella et al, titled Patient-Reported Outcomes in Performance Measurement
- The 2013 Committee Final Report, Patient-Reported Outcomes in Performance Measurement
- The 2020 Committee Final Report, Patient-Reported Outcomes: Best Practices on Selection and Data Collection
- The 2021 Building a Roadmap Interim Report on high quality PROMs for use in PRO-PMs

Resources and publications about the Consensus Development Process

- CDP homepage
- The Submitting Standards webpage
- The Measure Evaluation Criteria webpage
- The Measure Evaluation Criteria and Guidance for Evaluating Measures for Endorsement (PDF, September 2021)
- Measure Developer Guidebook for Submitting Measures to NQF (PDF, Version 6.5, July 2022)
- Intent to Submit Checklist and Guidance (PDF, May 2021)

NQF reports on pertinent areas of measure development

- The Best Practices for Developing and Testing Risk Adjustment Models webpage (2022)
- The Risk Adjustment for Sociodemographic Factors webpage (2014)
- The Attribution - Principles and Approaches webpage (2016)
- The Attribution for Critical Illness and Injury webpage (2021)

U.S. Department of Health & Human Services Resources

CMS reports on pertinent areas of measure development

- The CMS Measures Management System website
- The CMS Blueprint Measure Lifecycle Overview webpage
- The CMS Blueprint QuickStart Guide (PDF, May 2022)
- The Supplement on Patient-Reported Outcome Measures (PDF, May 2022)
- The Supplement on Risk Adjustment in Quality Measures (PDF, May 2022)

CMS and eCQI resources on digital quality measures

- The eCQI Resource Center website
- Overview of electronic clinical quality measures (eCQMs)
- dQM Strategic Roadmap webpage
- Digital Quality Measurement Strategic Roadmap (PDF, March 2022)
- Digital Quality Measurement Strategic Roadmap – Executive Summary (PDF, March 2022)
- Explanations of Qualified Clinical Data Registries (QCDR)
- Overview of Clinical Quality Language (CQL)
- Overview of United States Core Data for Interoperability (USCDI)

- The ONC USCDI+ website, which describes the initiative and provides links to additional resources

Additional Resources on Interoperability and Technical Issues

- The Fast Healthcare Interoperability Resources (FHIR) website
- The FHIR Implementation Guide Registry
- The FHIR Confluence site
- The Regenstrief Institute homepage on Logical Observation Identifiers, Names and Codes (LOINC)

Other Resources

- The Council of Medical Specialty Societies (CMSS) resources on clinical registries
- The Assistant Secretary for Planning and Evaluation (ASPE) Report to Congress: Social Risk Factors and Performance Under Medicare's Value-Based Purchasing Programs (2016)
- The ASPE Second Report to Congress: Social Risk Factors and Performance Under Medicare's Value-Based Purchasing Programs (2020)

Appendix C: Glossary of Terms

Accountable Unit

The entity whose performance is being measured, which could be a hospital, health plan, clinician, etc. Performance measurement can be applied to any setting and level of analysis.[15]

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, a care episode, or a population.[31]

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.[32]

Attribute

A characteristic or trait of a PROM. Past National Quality Forum (NQF) reports have used *attribute* and *characteristic* synonymously.[16,22] Throughout the Roadmap, *attributes* primarily refer to the characteristics that make a PROM suitable for use in a PRO-PM.

Attribute Grid

A table in the Interim Report designed to facilitate a side-by-side comparison of different PROMs against 12 attributes of high quality PROMs.[33]

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.[15]

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.[34]

Crosswalk

A concordance table to convert scores from one scale to the other and vice versa.[35] 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.[35]

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.[32]

Digital Quality Measures (dQMs)

Quality measures, organized as self-contained measure specifications and code packages, that use one or more sources of health information that is captured and can be transmitted electronically via interoperable systems. Data sources for dQMs may include administrative systems, electronically submitted clinical assessment data, case management systems, electronic health records, laboratory systems, prescription drug monitoring programs, instruments (for example, medical devices and wearable devices), patient portals or applications (for example, for collection of patient-generated data such as a home blood pressure monitor, or patient-reported health data), health information exchanges, or registries, and other sources.[13]

Electronic Clinical Quality Measures (eCQMs)

An electronic clinical quality measure (eCQM) is a clinical quality measure expressed and formatted to use data from electronic health record (EHRs) and/or health information technology systems to measure healthcare quality, ideally data captured in a structured form during the process of patient care. For the measured entity to report an eCQM from an EHR, eCQM developers format the Health Quality Measure Format using the Quality Data Model to define the data elements and Clinical Quality Language to express the logic needed to evaluate a provider or organization's performance.

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, representative mean(s) and standard deviation(s) in the reference population, and guidance on the minimally important difference in scores between groups and/or over time.[22]

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.[36,37]

Measured Entities

Measured entities are the front-line clinicians and their organizations, including health information technology, collecting quality measurement data. Measured entities are the implementers of quality measures. The effect of quality measure data collection on clinician workflow can be negative. There may be effects on their payments, positive and negative, with respect to reporting and actual performance on quality measures. Because of these potential effects, measured entities should be involved in all aspects of the Measure Lifecycle.[38]

Minimal Clinically Important Difference (MCID)

This is the smallest improvement needed after treatment that would be considered worthwhile from the patient's perspective.[32] 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.[32]

Patient-Reported Outcome (PRO)

The status of a patient's (or person's) health or behavioral condition that comes directly from the patient without interpretation of the patient's response by a clinician.[9]

Patient-Reported Outcome Measure (PROM)

The tools and instruments that are used to collect PRO data.[9] Depending on the measurement concept, PROMs can be used to collect data for PRO-PMs.[9]

Patient-Reported Outcome Performance Measure (PRO-PM)

How PROs are calculated. A way to aggregate the information from patients into a reliable, valid (tested) measure of performance at the healthcare entity level (e.g., a hospital, health plan, or clinician).[9]

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.[39]

Psychometric Soundness

How consistently and accurately an assessment measures what it purports to measure.[34] 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).[40]

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.[41]

Appendix D: 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	*	*	*	*
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)	*	*	*	*

* Indicates the table cell left intentionally blank

Appendix E: 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 **(opening date)**, through **(closing date)**. **TBD** prompts were offered to guide public commenters on key areas of interest. **TBD** comments from **TBD** 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.

Comments and responses will be added in Draft 3 of the report.