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

August 23, 2021

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

Contents

Executive Summary

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

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

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

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

The TEP also identified 15 tasks that occur within the four stages. The TEP designed the Roadmap with the understanding that the measure developer can flexibly move these tasks within a stage or to a different stage to accommodate individual and organizational style and preference. This flexibility is referenced throughout the report.

Many of the tasks are highly technical and require specific knowledge or training. For example, one task is to develop and test a risk adjustment model, which requires an advanced understanding of statistics and public health. The Roadmap is not intended as a tutorial to train the reader on how to become a measure developer. Instead, it provides guidance to a person with the appropriate training, even if they are new to the field of measure development.

Introduction

Diverse healthcare stakeholders increasingly view PRO-PMs as an important opportunity to ensure that the patient's voice is captured in assessments of quality.[2] Policymakers, payers, and healthcare providers are beginning to use PRO-PMs to inform clinical decision making, improve quality of care, modify provider payment, and evaluate the value of medical technologies.[2] There are other added benefits when PROs are included in performance measurement: Data suggest that the process of patient

self-reporting itself can improve symptom management, quality of life, communication, and satisfaction with care.[3]

PROs represent the measurement of a patient's health and behavioral condition, or experience with the healthcare system, directly from the patient without interpretation of the patient's response by a clinician.[4] PROMs represent the tools and instruments that are used to collect the data (e.g., Patient-Reported Outcome Measurement Information System [PROMIS], Patient-Reported Outcomes version of the Common Terminology Criteria for Adverse Events [PRO-CTCAE], and the Patient Health Questionnaire 9 [PHQ-9]).[4] Depending on the measurement concept, PROMs can be used to collect data for PRO-PMs that aggregate patient-reported data to assess the quality of an accountable-healthcare entity (e.g., a hospital, health plan, or clinician).[4] In order to properly employ PROMs for these purposes, measure developers must test numerous criteria of the resultant PRO-PM, including its reliability, validity, and feasibility.

Although PRO-PMs make up a small percentage of the overall measures used in CMS' value-based purchasing (VBP) programs and alternative payment models (APMs), CMS views these performance measures as an important part of its evolving initiatives to incentivize high quality care. In 2010, the Affordable Care Act (ACA) was passed, and CMS adopted new VBP programs that shifted towards improving and rewarding value rather than volume.[5] CMS designed the VBP programs to ensure that healthcare was more patient centered by engaging patients as partners and creating care that focuses on patient preferences and desired outcomes of value to patients.[5] Several VBP programs apply to various provider settings, such as hospitals and outpatient centers.[5] In addition to VBPs, CMS utilizes APMs to incentivize eligible participants to provide high quality and cost-efficient care.[5] VBPs 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.[5]

Despite increasing support for PRO-PMs, their development and use are still emerging. Significant challenges hamper the broad adoption of these measures across healthcare. PROMs (which are the tools on which PRO-PMs are based) have not yet become standard practice in clinical use, and some healthcare professionals (including some clinicians and payers) possess limited understanding of their benefits. There is also a lack of thorough, accessible technical guidance that measure developers can use to identify PROMs for use in digital PRO-PMs and to develop high-impact outcome measures based on patient-reported data.[6]

This Roadmap provides support on what measure developers should consider when selecting high quality PROM(s) to develop a digital PRO-PM that is:

1. appropriate for regulatory purposes;
2. aligned with best practices related to developing digital PRO-PMs;
3. usable by public and private payers;
4. appropriately adjusted for risk; and
5. able to attribute fair and accurate linkages between a health outcome and the entity that has control over it.

For the purpose of the Building a Roadmap initiative, a high quality PROM specifically refers to a PROM that collects data—including patient responses, scores, or threshold data—that are appropriate for use

in a digital PRO-PM that is suitable for inclusion in a CMS VBP program or APM. NQF does not currently endorse, recommend, rank, or prioritize PROMs, and a PROM that is considered high quality for the Building a Roadmap initiative may not be considered high quality for other purposes (e.g., research).

Background

This work builds upon past NQF projects by presenting the Roadmap to guide measure developers in identifying high quality PROMs for use in digital PRO-PMs that are suitable for use in CMS' VBP programs and APMs and can be calculated and transmitted via electronic health record (EHR) systems and other health information technology (IT) systems.

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

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

- **Limited number of NQF-endorsed PRO-PMs:** At the time of this publication, only 29 PRO-PMs were endorsed by NQF, compared to approximately 200 NQF-endorsed process measures and 320 outcomes measures.[7] Different challenges contribute to the relatively small number of endorsed PRO-PMs, including resource limitations (e.g., finances, time, and staff) and a lack of clear guidance for development.[7]
- **CMS prioritization of reduced burden and elevated patient voices:** The CMS Meaningful Measures 2.0 initiative identifies priorities that include reduced measurement burden on clinicians and an increased presence of patient voices.[1] These priorities can be addressed, in part, through the utilization of digital PRO-PMs.
- **Relationship between PROMs and PRO-PMs:** While the majority of NQF-endorsed PRO-PMs collect data from a single specified PROM, the TEP agreed that there are benefits to developing performance measures that allow for the use of different PROMs to measure the quality concept of interest. These benefits include the following:
    - Increased autonomy for clinical settings to implement the PROMs that best meet the needs of their population as well as the organization's business needs (e.g., PROMs that are culturally sensitive and translated into languages that are most relevant to the patient population and PROMs that are free versus those with a licensing cost)

    - Flexibility to collect comparable outcomes data from different instruments (e.g., the use of the Short Form 12 [SF-12], the Veterans RAND 12 Item Health Survey [VR-12], or the PROMIS Global Health PROMs)
- **Integration with health IT systems:** The widespread use of PROMs and PRO-PMs requires improved integration with EHRs and other health IT systems. This is achieved through a combination of interoperability standards (e.g., Fast Healthcare Interoperability Resources [FHIR]) and coding schemes (e.g., Logical Observation Identifiers Names and Codes [LOINC]).
- **Digital quality measures:** Digital quality measures (dQMs) automatically pull data that are generated during the normal course of clinical care.[8] PROM owners do not always agree to allow terminologies such as LOINC to include codes for PROM subscales and total scores, which prevents the PROM from being used as part of a digital measure.

Upon the completion of the Environmental Scan Report, the TEP guided the development of the project's second report: the Interim Report. The Interim Report defines the attributes of high quality PROMs that are appropriate for use in a digital PRO-PM. These attributes are described in the Interim Report and presented in the form of an Attribute Grid that measure developers can use to analyze and compare different PROMs that measure similar outcomes (Appendix B). The attributes identified in the Interim Report are listed below:

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

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

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

- 2012: Methodological Issues in the Selection, Administration, and Use of Patient-Reported Outcomes in Performance Measurement in Health Care Settings (henceforth referred to as *Methodological Issues*; this report was updated in 2015 by its authors.)[9]
- 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*)

Goals and Objectives

The goal of the Roadmap is to provide guidance on developing PRO-PMs, from the identification of high quality PROMs appropriate for use in digital PRO-PMs to preparation for submission for NQF endorsement. The report is intended for novice and advanced measure developers alike. While the guidance in this report is generally applicable to all PRO-PMs, it specifically focuses on digital PRO-PMs that are intended for use in CMS' VBP programs and APMs.

The Roadmap provides an overview of the PRO-PM development process that a new developer can read within one hour. The Roadmap does not address every question or scenario related to the development of digital PRO-PMs; instead, it is a high-level guide that helps measure developers to understand a complex process, ask educated questions, and contribute to creating a more robust database of digital PRO-PMs.

Roadmap for the Development of a Digital PRO-PM

Overview and Visualization of the Roadmap

The Roadmap provides guidance on PRO-PM development. It is intended for measure developers at all levels of experience who aim to develop PROM-based performance measures that can be submitted for NQF endorsement. The Roadmap is organized into four stages:

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

The TEP also identified 15 tasks that occur within the four stages. The TEP recommends that the measure developer address each of these tasks during the development process. **In the same way that an actual roadmap offers many routes between two locations, the PRO-PM Roadmap recognizes that there are many different paths from identifying the need for a measure to submitting a fully tested PRO-PM for NQF endorsement.** While the four stages do generally occur in the sequence listed above, the 15 tasks are not entirely linear and are not bound to any specific stage. Different measure developers and organizations may follow different processes when developing 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, the task titled "Develop an Attribution Model" is included in stage 3 of the Roadmap; however, measure developers on the TEP expressed different preferences on when this task should occur, with some preferring to design the attribution model as part of the preliminary work that occurs during stage 1. The Roadmap is not intended to be prescriptive as to when each task

must be performed, and it acknowledges that significant flexibility exists within these tasks. However, every task is an important part of the digital PRO-PM development process and should be addressed by the measure developer. If a task is not relevant to the PRO-PM being developed, the measure developer should document the rationale for future reference.

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

The visualization of the Roadmap (Figure 1) illustrates the TEP's support for grouping the entire PRO-PM development process into four stages, with 15 tasks that can shift across stages. This reflects the flexibility that measure developers may take when developing a PRO-PM.

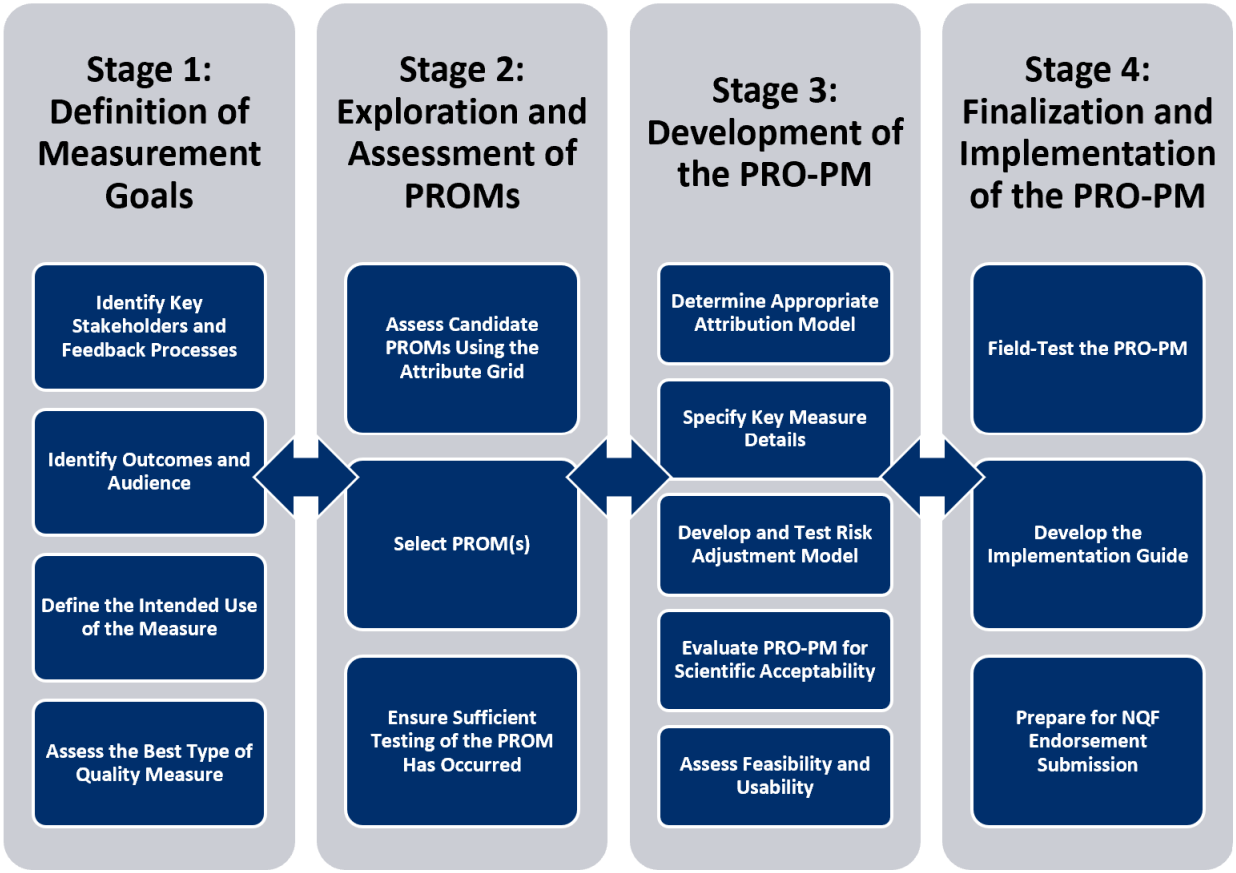

Figure 1. PRO-PM Roadmap: Each column contains one stage, and the bidirectional arrows indicate that tasks can move freely and be iterated across stages.

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

- **Stage 1: Definition of Measurement Goals.** The measure developer assembles the appropriate advisory stakeholders to help identify the measure's outcomes and audience, determine its intended use, and ascertain that a PRO-PM is the proper measurement approach. Intentional

and strategic thought during stage 1 can help establish a clear plan for the entire measure development life cycle. Successful execution of the tasks within this stage can influence the success of the entire Roadmap, up to preparation for NQF endorsement in the final stage.

- **Stage 2: Exploration and Assessment of PROMs.** This stage of the Roadmap includes the assessment and selection of the PROM(s) that will collect data for the PRO-PM. This stage also includes a detailed review of the testing that was completed for each PROM. The Interim Report for this project describes, in detail, the process and criteria for assessing the attributes of a PROM and determining whether it is appropriate for use in a PRO-PM. The Interim Report, along with the Attribute Grid it contains, can help a measure developer to select high quality PROM(s) for use with the performance measure.
- **Stage 3: Development of the PRO-PM.** Depending on the measure developer's experience, preference, or organizational policies, the tasks within this stage may vary. However, every task in stage 3 must be considered within the development process. If the measure developer and the key advisors determine that a task is not appropriate for the PRO-PM (e.g., if stratification by clinical or social risk is recommended instead of risk adjustment), the measure developer should thoroughly document these decisions for future reference.
- **Stage 4: Finalization and Implementation of the PRO-PM.** During this stage, the measure developer field-tests the PRO-PM, creates the implementation documentation, and prepares for the NQF Consensus Development Process (CDP), during which an NQF Standing Committee evaluates the PRO-PM and determines its suitability for endorsement. As mentioned previously, the preparation for tasks in stage 4 will likely begin early in the development process, even though these tasks are not completed until late in the process. Careful planning and preparation for these tasks will improve the likelihood of successfully testing the PRO-PM and submitting it for NQF endorsement.

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

Stage 1: Definition of Measurement Goals

As with any major project, planning and preparation at the early stages of measure development can positively influence the entire development life cycle. A measure developer should complete certain tasks prior to beginning the development of a PROM-based performance measure. This early work helps to ensure the measure will assist in improving the desired outcomes. The measure developer should address the following four tasks during stage 1, along with any tasks from other stages that may occur iteratively or that may be helpful in informing the preliminary work described below. As discussed earlier, an individual measure developer may decide to move some tasks into stage 1 based on individual preference or organizational policies.

Identify Key Stakeholders and Feedback Processes

The TEP highlighted the importance of engaging a variety of key advisory stakeholders to provide recommendations to the measure developer throughout the development process. The measure developer should identify and select representatives from key stakeholder groups early in the development process, ensure their feedback is incorporated into the development life cycle when appropriate, and secure their buy-in that the new measure will provide value to the stakeholder groups.

These stakeholders must include patients, patient advocacy groups, caregivers, and/or consumer groups, as well as an appropriate representation of the following stakeholders, which is dependent upon the intended use of the PRO-PM: clinicians, administrative staff (e.g., staff who will potentially administer or assist with PROM completion), payers, health IT staff, EHR vendors, data scientists, policy experts who represent the entities that will be affected by the measure, and any other stakeholder that will be directly affected by the PROM completion and collection process and/or the performance measure. Patients, clinicians, payers, and health IT perspectives are particularly important for digital PRO-PMs that are intended for use in VBP programs or APMs. In addition to identifying advisory stakeholders, 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).

Identify Outcomes and Audience

The measure developer and the advisory stakeholders should identify a consensus definition for the concept that is to be measured, including the desired outcomes that will be monitored and the primary audience of users. Examples of outcomes include functional status after a procedure, evaluation of symptom remission, or assessment of pain control throughout the care continuum. The primary audience of the measure includes whom the measure is intended to assess (e.g., the clinicians who performed a procedure). This task should occur before any PROMs are reviewed or any measure development occurs.

Define the Intended Use of the Measure

The measure developer should finalize the intended use of the measure prior to selecting PROMs or beginning development. Separate from the desired outcomes, the intended use of the measure identifies where or how this measure will be used once completed. Examples of intended use can include quality improvement purposes, comparison of provider performance, or inclusion within accountability programs, such as VBP programs or APMs. Determining the intended use ahead of time will help to ensure that the resulting measure is appropriate and likely to be successfully implemented.

Assess the Best Type of Quality Measure

Once the above tasks are completed, a measure developer must ensure that the goals of the measure are best achieved by a PRO-PM, as opposed to another type of measure (i.e., an outcome, process, or structural measure). A PRO-PM is a performance measure that is based on PROs assessed through data that are typically collected via a PROM and then aggregated for an accountable-healthcare entity.[10] Outcome measures reflect the impact of the healthcare service or intervention on the health status of patients, such as surgical complication rates or mortality rates.[11] Process measures indicate what a provider does to maintain or improve health, either for healthy people or those diagnosed with a healthcare condition.[11] Structural measures give consumers a sense of healthcare providers' capacity, systems, and processes to provide high quality care, such as provider-to-patient ratios and the number of board-certified physicians.[11] Due to the complexity of PRO-PM development, all measure types should be carefully considered prior to deciding on a PRO-PM as the correct type of quality measure.

When these tasks are complete, the measure developer should scan existing measure databases (e.g., NQF's Quality Positioning System (QPS), CMS' Measure Inventory Tool, and CMS' Pre-Rulemaking webpage that includes the Measures Under Consideration [MUC] list) to ensure that the definition,

outcome, audience, intended use, and measure type are all novel or significantly improve upon existing measures.

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

Stage 2: Exploration and Assessment of PROMs

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

Assess Candidate PROMs Using the Attribute Grid

The measure developer can use the Attribute Grid (Appendix B; described in detail in the Interim Report) to compare PROMs side-by-side and determine which are most suitable for use in the digital PRO-PM being developed.

To use the Attribute Grid, the measure developer must first identify at least one PROM that is a candidate for use in the digital PRO-PM. The developer can consult several resources to identify candidate PROMs that might be suitable data collection tools for the desired PRO-PM. The advisory stakeholders can be helpful in identifying which of the five domain(s) (i.e., health-related quality of life [HRQoL], functional status, symptoms and symptom burden, health behaviors, and experience with care) the candidate PROMs should address in order to collect data for the PRO-PM.

- The International Consortium for Health Outcomes (ICHOM) Standard Sets provide lists of recommended PROMs that have been vetted by diverse expert panels.
- Many professional societies have convened working groups of experts to review and recommend PROMs that are relevant within that specialty or discipline.
- Articles from peer-reviewed journals can provide insights into which PROMs are most widely used within a discipline, and which novel PROMs are offering noteworthy contributions.
- The descriptions or specifications of PRO-PMs typically list which PROM(s) are used to collect data for the performance measure; websites including NQF's QPS and CMS' Measures Inventory Tool include these details if they are available.
- The HealthMeasures program, funded by the National Institutes of Health (NIH), includes four PRO measurement sets that are each widely used and researched:
    - PROMIS
    - Quality of Life in Neurological Disorders (Neuro-QoL)
    - Adult Sickle Cell Quality of Life Measurement Information System (ASCQ-Me)
    - The NIH Toolbox
- NQF's PRO Best Practices Report can offer insights on the criteria that decision makers in clinical settings assess when selecting PROM(s) for the patients they serve. (See Appendix C for the PRO Best Practices Attribute Grid to assist with PROM selection in clinical settings.)

Once the measure developer identifies candidate PROMs, the Attribute Grid provides a structure for performing a side-by-side comparison of the benefits and limitations of each PROM. The Attribute Grid prompts 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 a CMS VBP program or APM). Patients and caregivers, clinicians, and other advisory stakeholders can help the measure developer to assess the candidate PROMs, particularly with regard to the first two attributes:

- 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 Attribute Grid includes several columns for comparing multiple PROMs (one column per PROM). The Attribute Grid also includes one row for each attribute being assessed. The measure developer may identify additional attributes that are relevant to the performance measure and add new rows to accommodate these attributes.

Select PROM(s)

Once the measure developer has identified candidate PROMs and completed the Attribute Grid, the selection process can begin. The measure developer reviews and compares the attributes to determine which PROMs are most suitable for the PRO-PM.

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 such, the measure developer has the flexibility to prioritize the attributes based on the needs of the PRO-PM that is being developed.

The measure developer will find quantifiable results for many of the cells in the completed Attribute Grid; for example, the attribute of "Psychometric soundness: reliability" should always include a numeric result. However, the Attribute Grid does not specify points at which results should be considered acceptable. While the 2012 Methodological Issues Report does recommend 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,[12] the TEP notes that the measure developer should use discretion in determining whether the existing reliability and validity testing are sufficient for selecting a PROM. **Because PRO-PM developers typically do not have the resources to independently test the reliability and validity of each PROM, developers should consider seeking PROMs in which the existing reliability and validity tests were performed in settings that align with the entities being measured by the PRO-PM.** The TEP encourages the measure developer to refer to existing tools and recommendations when selecting PROMs, such as literature from ICHOM or the Consensus-Based Standards for the Selection of Health Measurement Instruments (COSMIN) as well as previous NQF reports, including the 2012 Methodological Issues Report. Additionally, the TEP encourages the measure developer to utilize professional judgement and expertise in selecting PROMs, knowing that the pool of instruments and literature is constantly expanding.

Some attributes will have qualitative results. For example, while a measure developer can quantifiably measure the number of languages in which a PROM is translated, the cultural appropriateness of those translations is a more subjective assessment. Similarly, the selected PROM(s) must be feasible to implement and use, which includes the burden of the PROM to patients and clinicians, the presence of clear data collection workflows, and the ease with which the PROM fits within standards of care. These subjective attributes are included so that the measure developer will think holistically about a performance measure and consider the perspectives of multiple stakeholders, particularly patients and caregivers, when assessing PROMs.

During this task, the measure developer will determine whether the PRO-PM will use patient-reported data from a single PROM or from several different PROMs. While choosing a single PROM is likely to result in a simpler development process, it also requires the use of that PROM. **This can create issues in care settings if the PROM has licensing fees, does not have translations suitable for the patient population, requires the creation of new workflows, or is not culturally appropriate for the population.** VBP programs and APMs are likely to favor PRO-PMs that accept data from different PROMs. Although accepting data from different PROMs makes the development process more complex, **the Building a Roadmap TEP recommends that measure developers design PRO-PMs to accept data from multiple high quality PROMs**.

Ensure Sufficient Testing of the PROM Has Occurred

Many elements of testing—including reliability, validity, responsiveness, and mode of administration—are assessed using the Attribute Grid. However, a PRO-PM could require a PROM to undergo additional testing. During this task, the measure developer examines additional evidence supporting the PROM's suitability for use in the PRO-PM. As an example, this evidence may include the determination of whether the PROM was tested in the setting (e.g., inpatient surgery, ambulatory surgery center) that the PRO-PM will be implemented.

If the original testing of the PROM does not match the current intended use, the PRO-PM developer should contact the owner of the PROM to determine whether additional information is available.

Because the Attribute Grid allows for additional attributes that are pertinent to the PRO-PM being developed, the measure developer may consider adding these testing requirements to the Attribute Grid when assessing candidate PROMs for the relevant setting of care.

Stage 3: Development of the PRO-PM

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

Determine Appropriate Attribution Model

The measure developer must answer an important question during the PRO-PM development process: Who should be responsible for the patient outcomes? Attribution is used in quality measurement to assign accountability for a patient's outcomes to a clinician, group of clinicians, facility, or to accountable care organizations (ACOs) and health plans.[13] An attribution model is a set of rules that defines an accountable unit for a patient's healthcare outcomes; using an attribution model helps determine and specify who is responsible and how the measure can be implemented.[13]

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

1. Attribution models should be fair and accurate.
2. Attribution models are an essential part of development, implementation, and policy and program design.
3. Considered choices among available data are fundamental in the design of an attribution model.
4. Attribution models should be reviewed and updated regularly.
5. Attribution models should be transparent and consistently applied.
6. Attribution models should align with the goals and purpose of the program.[13]

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

Specify Key Measure Details

The technical specifications for a PRO-PM identify important details about the measure, such as the population (i.e., the patients whose PROM data will be aggregated for the measure), the inclusion and exclusion criteria, and the data sources (i.e., PROMs). In other words, the specification identifies the information and data that are needed to generate the measure result. Measure specifications provide the technical instructions on how the measure should be collected and used consistently, reliably, and effectively.[14] Developing measure specifications is an iterative process, and measure developers must include a variety of stakeholder perspectives.[14]

Creating full measure technical specifications includes defining the data source (including all PROMs that can collect data for the PRO-PM as well as any other required data sources), specifying the code

systems, constructing the data protocol, and documenting the measures. Elements that are typically part of technical specifications include the following:

- Measure name/title
- Description
- Population
- Numerator and denominator statements
- Inclusion and exclusion criteria and exceptions
- Data sources
- Key terms
- Data elements, codes, and code systems
- Unit of measurement of analysis
- Sampling methodology
- Risk adjustment
- Time intervals
- The calculation algorithm[14]

Digital quality measures (dQMs), including digital PRO-PMs, may include a narrative measure specification (i.e., a human-readable document) but provide specifications in machine-readable formats that have the potential to be updated remotely in real time, much in the way an app on a smartphone can be updated.[15] Clinical Quality Language (CQL) is used to create specifications for dQMs. It provides a more precise specification while reducing potential opportunities for interpretation or calculation errors.[15] The growing support for FHIR standards will enhance the use of dQMs and streamline the specification process.

Develop and Test Risk Adjustment Model

Risk adjustment refers to statistical methods used to account for patient-, community-, health plan-, or facility-level risk factors when computing outcome performance measures and resource use measures.[16] Patients bring these characteristics into healthcare, and they are outside of the clinician's control. Risk-adjusting to account for differences in health status and clinical factors (e.g., comorbidities or severity of illness) that are present at the start of care has been widely accepted and implemented within measure development.[17,18] A measure developer can determine how to approach these characteristics when developing a PRO-PM, as they may affect a patient's health outcome. Measures that are risk-adjusted at the facility level allow for the healthcare system or hospital to identify improvement opportunities, although risk adjustment can also mask certain patients with poor outcomes. By controlling for appropriate factors, however, risk adjustment can promote a fair comparison among different healthcare entities.

Risk adjustment has traditionally focused on clinical risk factors. The idea of incorporating social or functional status risk factors has been debated due to potential unintended consequences, such as masking disparities and institutionalizing differing standards of performance.[19] Risk stratification, defined as the division of a population or resource services into similar groups of data that enable analysis of specific subgroups, can be used as an alternative to risk adjustment to identify healthcare disparities.[19] NQF endorsement requires an evidence-based risk adjustment strategy that has demonstrated adequate discrimination and calibration and is based on patient factors that influence the measured outcome and are present at the start of care.[20] If risk adjustment is not appropriate for the

PRO-PM, the measure developer must document and provide rationale and evidence to support the lack of risk adjustment.[20,21]

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.[20] These criteria determine whether the PRO-PM, as specified, produces consistent (i.e., reliable) and credible (i.e., valid) results.[20] The measure developer **must not** conflate reliability and validity of the PRO-PM with reliability and validity of the chosen PROM(s) because these are completely separate processes.

Reliability comprises two subcriteria:

- The measure must be well defined and precisely specified so that it can be implemented consistently.
- The measure's data elements are able to be reproduced, repeating the same or similar results at a high proportion of the time when assessed in the same population in the same time period and/or that the measure score is accurate.[20]

Validity comprises six subcriteria that assess data elements, exclusions, risk adjustment model, computed measure scores, comparability of data sources (i.e., data from each PROM must be comparable), and nonresponse.[20] PRO-PMs must demonstrate reliability at both the performance measure level (i.e., the PRO-PM) and the data element level (i.e., the PROMs).[20,21]

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

Assess Feasibility and Usability

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."[20] Examples include clinical data that are routinely generated and used during care delivery (e.g., blood pressure readings or lab test results) and data available in EHRs or other electronic sources. Feasibility should be demonstrated by a data collection strategy that can be reasonably implemented.[21] The current state of PROM data collection can create feasibility challenges; nevertheless, the measure developer can mitigate these risks by using the Attribute Grid to identify high quality PROMs (e.g., those with an evidence base that supports successful data collection in the relevant setting from the specified patient population and those with assigned LOINC codes that are suitable for a digital PRO-PM).

Usability refers to the extent to which potential audiences are able to interpret and understand performance results. The measure developer should consider unintended consequences and a plan to assess whether they outweigh the evidence of improving healthcare quality.[21] The measure developer should also ensure that a plan is in place to provide performance results, assistance understanding the results, and a feedback mechanism to those being measured, and to consider this feedback when changes are made to the measure.[20]

Stage 4: Finalization and Implementation of the PRO-PM

The final stage of the Roadmap addresses field testing of the PRO-PM, development of implementation guidance, and preparation for NQF endorsement. As with the other stages, the measure developer may address some of 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. For example, field testing is complex, expensive, and time-consuming, and if appropriate planning and preparation do not occur during the previous stages, the timeline and budget can be significantly affected. As with many of the tasks in the Roadmap, certain tasks in this stage should be viewed as iterative, particularly related to implementation guidance and field testing.

Field-Test the PRO-PM

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

- Identification and confirmation of test sites to ensure testing occurs in the appropriate setting
- Implementation of necessary PROMs in the test sites, including defining workflows, training staff, and ensuring health IT supports (e.g., EHRs, patient portals) function as expected
- Development and delivery of test plan with necessary details (e.g., how, when, and where testing will occur; what patients are included in testing; and what data are captured)
- Documentation on how testing results will be assessed and how issues will be resolved
- Preparation of any legal documentation and communication, including internal review board (IRB) submissions or data use agreements (DUAs) required under the Health Insurance Portability and Accountability Act (HIPAA)

Ideally, test sites will receive an implementation guide so that this task (as well as the stage 3 task related to scientific acceptability) can benefit from iterative creation of the implementation guide.

An important aspect of field testing is the development of a comprehensive and accurate test plan that closely follows the NQF measure evaluation criteria. To ensure practical and logistical success of the PRO-PM, the measure developer should identify specific objectives and goals of the field-testing process and ensure these are reflected in the test plan. Data requirements should be addressed in the test plan, including preliminary strategies for addressing issues that are likely to arise due to issues with data availability, data accuracy, insufficient sample sizes, or other common challenges. The test plan should also outline strategies to analyze and harmonize data from different test sites.

Develop the Implementation Guide

A systematic approach to implementation can minimize unique challenges from varied clinical settings and contribute to the success of implementation across diverse clinical environments. The measure developer should treat the creation of the implementation guidance as an iterative process that incorporates information from every stage of the PRO-PM development life cycle, including context on data collection via the PROM(s) and lessons learned during the field-testing process. 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 developer's intent for each cohort definition and how to ensure the reliability of the information sought in local data systems.

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

Prepare for Endorsement Submission

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

Not all of the NQF measure evaluation criteria are explicitly addressed in the Roadmap. The TEP deemed certain criteria (e.g., scientific acceptability, feasibility, and usability) to be particularly challenging, resource-intensive, or otherwise noteworthy during the digital PRO-PM development life cycle, and explicitly addressed these criteria in the Roadmap. **However, all of the measure evaluation criteria are important and should be closely considered during the development of a PRO-PM, regardless of whether or not the measure developer intends to seek NQF endorsement.** Extensive information 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 (effective September 2019)
- The Measure Developer Guidebook

The CMS Measures Management System Blueprint is also a valuable source of guidance, and the measure developer should closely review this document.

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

Conclusion

Digital PRO-PMs are an important component of the future of healthcare that will elevate the patient's voice in the assessment of healthcare quality. However, because only a few dozen PRO-PMs currently exist, it is critical that measure developers (particularly those developers who are at the beginning of their careers) receive resources to help them better understand and navigate the development process. When combined with the 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 PRO-PM development (i.e., Definition of Measurement Goals, Exploration and Assessment of PROMs, Development of the PRO-PM, and Finalization and Implementation of the PRO-PM) and the 15 iterative tasks that exist within those stages, measure developers can clearly understand the PRO-PM development process in less than one hour. The Roadmap is not a textbook that will answer every question related to the development of digital PRO-PMs, but it is a primer that will help measure developers understand a complicated process, ask informed and insightful questions, and ultimately aid in building a more robust database of digital PRO-PMs.

References

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

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

3 Basch E, Spertus J, Adams Dudley R, et al. Methods for Developing Patient-Reported Outcome-Based Performance Measures (PRO-PMs). *Value in Health*. 2015;18(4):493-504.

4 Centers for Medicare & Medicaid Services (CMS). Supplemental Materials to MMS Blueprint: Patient Reported Outcome Measures. September 2020:6.

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

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

7 NQF. NQF: Quality Positioning System. 2021. https://www.qualityforum.org/QPS.

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

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

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

11 Agency for Healthcare Research and Quality (AHRQ). Types of Health Care Quality Measures. http://www.ahrq.gov/talkingquality/measures/types.html. Published 2015. Last accessed August 2021.

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

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

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

15 Hamlin B, O'Kane ME, Barr M, et al. COVID-19 Underscores The Need For Digital Quality Measurement | Health Affairs Blog. https://www.healthaffairs.org/do/10.1377/hblog20200824.987845/full/. Last accessed August 2021.

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

17 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.

18 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.

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

20 NQF. Measure evaluation criteria and guidance for evaluating measures for endorsement. National Quality Forum. https://www.qualityforum.org/WorkArea/linkit.aspx?LinkIdentifier=id&ItemID=92804. Published 2019.

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

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

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

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

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

26 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.

27 NQF. NQF health information technology glossary: A guide to HIT jargon. June 2015. https://www.qualityforum.org/Measuring_Performance/HIT_Glossary.aspx.. Last accessed April 2021.

28 NQF. NQF: Measuring performance. https://www.qualityforum.org/Measuring_Performance/Measuring_Performance.aspx. Last accessed February 2021.

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

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

Appendices

Appendix A: Glossary of Terms

Alternative Payment Models (APMs)

A payment approach that gives added incentive payments to provide high quality and cost-efficient care. APMs can apply to a specific clinical condition, care episode, or population. [22]

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.[23]

Attribute

A characteristic or trait of a PROM. Past National Quality Forum (NQF) reports have used *attribute* and *characteristic* synonymously.[10,24] Throughout the Roadmap, *attributes* primarily refer to the characteristics that make a PROM suitable for use in a PRO-PM.

Attribute Grid

A table designed to provide a systematic method to perform a side-by-side comparison of PROMs on the basis of meaningful PROM attributes.[24]

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.[13]

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.[25]

Crosswalk

A concordance table to convert scores from one scale to the other and vice versa.[26] Crosswalks can allow harmonization of PROMs that measure similar outcomes (e.g., health-related quality of life [HRQoL] after a knee replacement surgery), which may facilitate multicenter collaboration or allow sites to switch PROMs without loss of historic comparison data.[26]

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.[23]

Digital Quality Measures (dQMs)

Digital quality measures originate from sources of health information that are captured and can be transmitted electronically and via interoperable systems.[1] These measures utilize data that are generated during the normal course of clinical care. Other types of dQMs include information generated

from medical devices, such as ventilators and digitized information from patient portals or other modules.[8]

Electronic Clinical Quality Measures (eCQMs)

These are the most recognizable of the digital measures and are specified for use in the Medicare and Medicaid EHR Incentive Programs. Eligible professionals, eligible hospitals, and critical access hospitals are required to submit eCQM data from certified EHR technology to help measure and track the quality of healthcare services provided within the healthcare system. These measures use data associated with providers' ability to deliver high quality care or related to long-term goals for quality healthcare.[27]

Interpretability

The degree to which the meaning of the scores can be easily understood by any group requiring use of the scores. A PRO measure should have documentation to support interpretation of scores, including the meaning of low and high scores and guidance on the minimally important difference in scores between groups and/or over time.[10]

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.[27]

Minimal Clinically Important Difference (MCID)

This is the smallest improvement needed after treatment that would be considered worthwhile from the patient's perspective.[23] 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.[23]

Patient-Reported Outcome (PRO)

1. The measurement of a patient's health and behavioral condition, or experience with the healthcare system, directly from the patient without interpretation of the patient's response by a clinician[4]
2. Any information on the outcomes of healthcare obtained directly from patients without modification by clinicians or other healthcare professionals[10]

Patient-Reported Outcome Measure (PROM)

1. The tools and instruments that are used to collect PRO data.[4] Depending on the measurement concept, PROMs can be used to collect data for PRO-PMs that aggregate patient-reported data to assess the quality of an accountable-healthcare entity (e.g., a hospital, health plan, or clinician).[4]
2. Any standardized or structured questionnaire regarding the status of a patient's health condition, health behavior, or experience with healthcare that comes directly from the patient (i.e., a PRO). The use of a structured, standardized tool, such as a PROM, will yield quantitative data that enables comparison of patient groups or providers.[10]

Patient-Reported Outcome Performance Measure (PRO-PM)

A performance measure that is based on PROs assessed through data often collected through a PROM and then aggregated for an accountable-healthcare entity.[10]

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.[28]

Psychometric Soundness

How consistently and accurately an assessment measures what it purports to measure.[25] 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).[29]

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.[30]

Appendix B: Attributes of High Quality PROMs for Use in PRO-PMs

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

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

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

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

Appendix C: PROM Attribute Grid From 2020 Report

Originally, this was Appendix F from the final NQF PRO Best Practices Report, published in September 2020. The PRO Best Practices TEP designed this Attribute Grid to assist decision makers in clinical settings with the selection and implementation of PROMs. Although the TEP did not design it with measure developers in mind, developers may find attributes in this grid that are beneficial to incorporate into the grid shown in Appendix B.

PROM	PROM 1	PROM 2	PROM 3	PROM 4
Covers desired PROs:	*	*	*	*
Contains goal attainment and goal attainment follow-up questions	*	*	*	*
Symptoms	*	*	*	*
Impacts	*	*	*	*
Costs/fees	*	*	*	*
Languages/translations available	*	*	*	*

PROM	PROM 1	PROM 2	PROM 3	PROM 4
Length (number of items)	*	*	*	*
Psychometric soundness: burden, including time and effort	*	*	*	*
Psychometric soundness: clear conceptual and measurement models Clinical applicability to desired population	**Concepts included:** **Intended population**:	**Concepts included:** **Intended population**:	**Concepts included:** **Intended population**:	**Concepts included:** **Intended population**:
Psychometric soundness: reliability (include sample size, various estimates if provided, and applicable population(s)) Good, better, or best reliability	**Test-retest reliability**: **Internal Consistency (Cronbach's a):**	**Test-retest reliability**: **Internal Consistency (Cronbach's a):**	**Test-retest reliability**: **Internal Consistency (Cronbach's a):**	**Test-retest reliability**: **Internal Consistency (Cronbach's a):**
Psychometric soundness: validity (include various estimates if provided and notes applicable population(s)) Good, better, or best validity	**Construct Validity (Population):**	**Construct Validity (Population):**	**Construct Validity (Population):**	**Construct Validity (Population):**
Psychometric soundness: responsiveness—ability to detect change Good, better, or best actionability	*	*	*	*
Psychometric soundness: clear documentation on how to interpret scores Good, better, or best interpretability	**Minimal clinically important difference**: summary or total score change	**Minimal clinically important difference**: summary or total score change	**Minimal clinically important difference**: summary or total score change	**Minimal clinically important difference**: summary or total score change

Appendix D: Committee Members, Federal Liaisons, and NQF Staff

Committee Members

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

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

David Andrews, PhD
Patient Advisor
Aiken, South Carolina

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Federal Liaisons

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

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

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

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

Rhona Limcango, PhD
Agency for Healthcare Research and Quality

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

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

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

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

NQF Staff

Kathleen Giblin, RN
Senior Vice President, Quality Innovation

Alejandra Herr, MPH
Senior Managing Director

Maha Taylor, MHA
Managing Director

Chuck Amos, MBA
Director

Teresa Brown, MHA, MA
Senior Manager

Deidra Smith, MBA, PMP
Senior Project Manager

Jhamiel Prince, MA
Analyst

Juanita Rogers, MS, CHES
Analyst

CMS Staff

LaWanda Burwell
Task Order COR

Sophia Chan
IDIQ COR

Gequincia Polk
IDIQ COR

Patrick Wynne
IDIQ COR

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

Appendix E: Public Comments and Responses

Public commenting will be welcomed and updated after the finalization of Draft 2 of this report.