NATIONAL QUALITY FORUM
Driving measurable health improvements together

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

INTERIM REPORT – DRAFT 2

June 1, 2021

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

Contents

Building a Roadmap From Patient-Reported Outcome Measures to Patient-Reported Outcome Performance Measures 1
Contents 2
Executive Summary 1
Introduction and Background 2
Brief NQF History With PROs 3
Terminology 4
Discussion of PRO Best Practices Report's Attribute Grid 5
Key Environmental Scan Findings Pertinent to the Interim Report 5
Prioritized Domains for the Interim Report 7
Goals and Objectives 8
Attributes of High Quality PROMs for Use in Performance Measures 9
Rationale for New Attribute Grid 9
Attributes and Definitions 9
Additional Considerations About the Attributes 12
Objectivity and Subjectivity 12
Flexibility of the Attribute Grid 13
Use Case: Assessing the Attribute Grid 13
Use Case Findings 14
Limitations of the Interim Report and the Attribute Grid 14
Conclusion 15
References 16
Appendices 19
Appendix A: Glossary of Terms 19
Appendix B: PROM Attribute Grid From 2020 Report 20
Appendix C: Attributes of High Quality PROMs for Use in PRO-PMs 22
Appendix D: PHQ-9 Use Case 23
Appendix E: Committee Members, Federal Liaisons, and NQF Staff 25
Appendix F: Public Comments and TEP Response 28

Executive Summary

Building a Roadmap From Patient-Reported Outcome Measures to Patient-Reported Outcome Performance Measures (henceforth referred to as "Building a Roadmap") is an initiative funded by the Centers for Medicare & Medicaid Services (CMS) and led by the National Quality Forum (NQF). The goal of the project is to provide step-by-step guidance to measure developers at all career stages on developing a fully tested patient-reported outcome performance measure (PRO-PM) for use in CMS accountability programs. As part of this guidance, NQF will identify key attributes of a high quality patient-reported outcome measure (PROM) that is suitable for use in a digital PRO-PM.

This initiative builds on a long-standing relationship between NQF and CMS to advance work in patient-reported outcomes (PROs). Two recent collaborative projects stemmed from measure developers' requests to CMS for guidance in developing PRO-PMs. Recognizing that successful development and implementation of PRO-PMs is dependent on frontline clinical use of PROs and PROMs, CMS and NQF launched the Patient-Reported Outcomes: Best Practices on Selection and Data Collection project (henceforth referred to as "PRO Best Practices") in 2019 to provide guidance to practices and health systems on how to select meaningful PROs and select/implement PROMs. Following the publication of the PRO Best Practices Final Technical Report in September 2020, NQF and CMS launched the Building a Roadmap initiative to guide measure developers in identifying high quality PROMs for use in digital PRO-PMs.

The initial year of the Building a Roadmap project will include the publication of three reports:

1. An Environmental Scan Report that outlines the current state of guidance on and practice in developing PROM-based PRO-PMs;
2. An Interim Report that identifies and describes the attributes of high quality PROMs for use in digital PRO-PMs; and
3. A Technical Guidance Report that guides measure developers at all career stages in the development of PROM-based digital PRO-PMs for regulatory purposes (e.g., CMS Value-Based Purchasing [VBP] programs and alternative payment models [APMs]).

This Interim Report will reflect the work of the Building a Roadmap Technical Expert Panel (TEP) across six web meetings. During its first four meetings, the TEP identified and described the attributes of high quality PROMs that are well suited to serve as data collection instruments for digital PRO-PMs that can be used in VBP programs, APMs, and other regulatory settings and/or innovative payment models. During its fifth and sixth meetings, the TEP will refine these attributes based on TEP members' expertise and insights gleaned from public comments on this report. The attributes identified by the TEP thus far are described in detail in this report and listed below:

- Desired PROs from patient and/or caregiver perspective (including cultural appropriateness)
- Defined and actionable cut points or targets, anchors, and/or meaningful change
- Outcome measured in PROM is the result of care for which relevant clinical quality is being measured
- Clear conceptual and measurement models
- Psychometric Soundness: Reliability
- Psychometric Soundness: Validity

- Psychometric Soundness: Responsiveness and/or actionability
- Usability/Feasibility of Use: Low burden, including length of tool, time/effort to complete
- Usability/Feasibility of Use: Fits with standard of care and related workflows
- Usability/Feasibility of Use: Language/translations
- Usability/Feasibility of Use: Standardized codes available

The TEP assessed these attributes against the attributes of a sample of PROMs identified by CMS as potential examples of high quality PROMs (i.e., they are currently used in federal programs, for data collection in existing PRO-PMs, and/or by a range of healthcare stakeholders). The goal of the assessment was to ensure the set of attributes adequately represents those PROMs that are particularly well suited as data collection instruments for performance measures. Where possible, the TEP identified PROMs that are currently used in NQF-endorsed PRO-PMs, CMS APMs, and/or for Medicare coverage determination.

Introduction and Background

As priorities in healthcare quality and delivery have changed, many healthcare stakeholders have called for an increased emphasis on patient centeredness without compromising on quality. Patient-centered care aims to ensure that the patient voice is included in treatment and care delivery and that health outcomes are based largely on the patient's health goals.[1] PROMs provide important information about the patient that is not available through clinical documentation or insurance claims by using data that are directly reported by the patient.[2] This information promotes patient-centered care and captures various dimensions of the quality of the patient's care. Given that this information is self-reported by the patient, it is easier to identify and improve outcomes that are most important to the patient.

In addition to the growing emphasis on patients being more involved in their care, changes in healthcare quality and delivery reinforce the value of PRO-PMs for health systems, researchers, clinicians, and policymakers. At the policy level, PRO-PMs add the patient voice to value-based payment reform and ensure the consumers' perspective on experience and outcomes is considered alongside quality and cost efficiency when adjusting provider reimbursement.[3] For clinicians and health systems, PRO-PMs can be used to assess treatment decisions and to monitor patient progress towards identified treatment goals.

At the time of this report's publication, only 35 PRO-PMs were endorsed by NQF, making up approximately 6 percent of the total number of NQF-endorsed quality measures.[4] The CMS-funded Building a Roadmap initiative aims to increase the development of PRO-PMs and prepare developers for the NQF endorsement process; to accomplish this, the initiative will provide measure developers with guidance on selecting high quality PROMs as data collection instruments for digital performance measures that are suitable for use in CMS' VBP programs and APMs. This project focuses on digital measures because of their ability to collect data with minimal burden; maximize response rates and thus, patient representation; and leverage electronic health records (EHRs) with data collection that creates efficiencies and improved treatment within healthcare. NQF is leading the work and has convened a 25-member multistakeholder TEP. The TEP includes patients and patient advocates, measure developers, health information technology (IT) professionals, clinicians, researchers, quality measurement experts, and other healthcare professionals with relevant perspectives on PROs. Within the Interim Report, unless a fact or recommendation is explicitly attributed to a specific source, information was gathered from the TEP and synthesized by NQF.

Building a Roadmap is the second of two recent CMS-funded initiatives that emerged as a result of measure developers seeking guidance from the federal agency on how to develop digital PRO-PMs more effectively. CMS recognized that PRO-PMs depend on PRO data captured as part of clinical care; therefore, the agency opted to first fund the PRO Best Practices initiative. PRO Best Practices provided guidance to healthcare staff in clinical settings on selecting and implementing meaningful PROs and PROMs. The PRO Best Practices Final Technical Report was published in September 2020, and Building a Roadmap was launched shortly afterward.

The initial year of the Building a Roadmap initiative will center on the development and publication of three reports:

1. Environmental Scan Report: This report assesses the current state of guidance on and practice in developing PROM-based PRO-PMs. It provides an overview of CMS' goals for digital quality measurement, a description of resources that measure developers can rely upon to identify candidate PROMs for use in PRO-PMs, an overview of the NQF endorsement process for performance measures, and a discussion of major challenges in the development of PRO-PMs.
2. Interim Report: This report identifies and describes the attributes of high quality PROMs for use in digital PRO-PMs for CMS' VBP programs and APMs and provides guidance on determining whether a PROM is well suited for performance measurement.
3. Technical Guidance Report: This report will offer guidance to measure developers at all career stages on the development of PROM-based digital PRO-PMs. While the guidance in this report will be generally applicable to all PRO-PMs, it will specifically focus on digital performance measures that are intended for use in CMS VBP programs and APMs.

Brief NQF History With PROs

Over the past decade, NQF has actively participated in the development of numerous reports intended to further the use of PROs and PROMs in clinical settings as well as the use of PRO-PMs to assess the performance of healthcare organizations.

Measure developers, clinicians, and researchers have developed hundreds of PROMs that collect patient-level outcomes data. While NQF does recognize the critical role of PROMs in collecting these data, NQF does not endorse PROMs. The organization's endorsement process focuses on quality measures that assess healthcare entities, such as health systems and health plans, at the aggregate level rather than the individual patient level. As such, NQF endorses PRO-PMs but does not endorse instruments or scales (including PROMs) on their own.[5] If a PROM is explicitly identified in the specification of a PRO-PM, NQF's Scientific Methods Panel will review the PROM for reliability and validity as part of the endorsement process. However, NQF remains agnostic to the specific PROM and reviews it only to the extent that it meets an acceptable scientific standard as an element of the PRO-PM.

NQF's endorsed PRO-PMs span different domains (e.g., health-related quality of life [HRQoL], functional status, symptoms and symptom burden, health behaviors, and experience with care), conditions and diseases (e.g., joint replacement, depression), and settings (e.g., ambulatory, inpatient, long-term care, and hospice). The Environmental Scan Report provides more details on the following reports:

- 2012: Methodological Issues in the Selection, Administration, and Use of Patient-Reported Outcomes in Performance Measurement in Health Care Settings—The first of two CMS-funded reports commissioned by NQF focused on selecting PROMs for use in performance measurement.[6] (The report was updated in 2015 by its authors.[7])
- 2012: PRO-Based Performance Measures for Healthcare Accountable Entities—The second commissioned report focused on reliability and validity of PRO-PMs.[8]
- 2013: Patient-Reported Outcomes in Performance Measurement—NQF convened a CMS-funded TEP whose reports laid the groundwork for future PRO-PM development, testing, endorsement, and implementation.[5]
- 2017: Measuring What Matters to Patients: Innovations in Integrating the Patient Experience into Development of Meaningful Performance Measures—NQF and PatientsLikeMe partnered on this Robert Wood Johnson Foundation-funded report that reiterated the importance of patient-centered quality measurement and online patient communities to measure developers.[9]
- 2020: Patient-Reported Outcomes: Best Practices on Selection and Data Collection—With CMS funding, NQF convened a multistakeholder TEP that identified best and promising practices for practices and health systems to follow when selecting and implementing PROs and PROMs.[2]
- 2021: Building a Roadmap From Patient-Reported Outcome Measures to Patient-Reported Outcome Performance Measures—This is the project page for the current initiative, which contains a link to the Environmental Scan Report.

Terminology

In this report, NQF will continue to use established terminology from the 2013 report to distinguish between PROs, PROMs, and PRO-PMs (Table 1). Because the terminology regarding PRO-PMs is highly technical, the report includes a glossary of key terms (Appendix A: Glossary of Terms).

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

Concept	Definition	Example
Patient-Reported Outcome (PRO)	Any information on the outcomes of healthcare obtained directly from patients without modification by clinicians or other healthcare professionals.[5]	Symptom: depression
Patient-Reported Outcome Measure (PROM)	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.[5]	Patient Health Questionnaire 9 (PHQ-9)©, a standardized tool to assess depression

Concept	Definition	Example
PRO-Based Performance Measure (PRO-PM)	A performance measure that is based on patient-reported outcomes assessed through data often collected through a PROM and then aggregated for an accountable healthcare entity.[5]	Percentage of patients with diagnosis of major depression or dysthymia and initial PHQ-9 score >9 with a follow-up PHQ-9 score <5 at 6 months (NQF #0711)

Discussion of PRO Best Practices Report's Attribute Grid

The 2020 PRO Best Practices Final Technical Report included an Attribute Grid developed by the TEP to aid decision makers in clinical settings with the selection of PROMs (Appendix B: PROM Attribute Grid from the 2020 Report). The TEP designed the grid as "a systematic method to perform a side-by-side comparison of PROMs on the basis of meaningful PROM attributes."[2] As illustrated in Appendix B, clinicians can use the grid to easily compare the attributes of different PROMs that measure similar PROs (e.g., the Oxford Knee Score [OKS] and the Knee Injury and Osteoarthritis Outcome Score, Joint Replacement [KOOS JR] for functional outcomes after a total knee replacement surgery) to determine which is most applicable to its clinicians and patients.

The Attribute Grid allows decision makers to assess both subjective and objective attributes of PROMs. Subjective attributes address questions such as how well each PROM assesses the PROs that are most important to clinicians and patients. Objective attributes address measurable aspects of a PROM, such as its reliability and validity. The TEP designed the grid to be flexible: The instructions guide the user to "Use the sample attribute grid … and add/remove rows based on organizational goals and priorities."[2]

When designing the Building a Roadmap initiative, NQF initially planned to use the PRO Best Practices Attribute Grid to identify high quality PROMs for use in performance measures. However, as the Building a Roadmap TEP members began reviewing the PRO Best Practices Attribute Grid, they determined that a slightly different set of attributes is necessary when reviewing PROMs for performance measurement. The TEP developed a new Attribute Grid, which is presented later in this report.

Key Environmental Scan Findings Pertinent to the Interim Report

While the Environmental Scan Report for this initiative is a stand-alone document that describes the current state of identifying high quality PROMs for use in digital performance measures, some of its findings are directly pertinent to this Interim Report and are summarized here. The findings focus on the following items:

- Limited number of PRO-PMs
- CMS priorities and reduced measure burden
- Relationship between PROMs and PRO-PMs
- Interoperability

Limited Number of PRO-PMs

PRO-PMs provide an opportunity to inform clinical decision making, adjust provider payment, assess and improve quality of care, and ensure the patient voice is captured in assessments of their health.[10] Although both peer-reviewed and grey literature reflect that the healthcare industry understands the

importance of PRO-PMs, significant challenges exist in their development and implementation. Challenges include, but are not limited to, identifying thresholds of meaningful change that matter most to patients, developing PRO-PMs within financial and other resource constraints, lack of extensive guidance on the development process, and interoperability and usability.[10] At the time of this publication, only 35 PRO-PMs were endorsed by NQF.[4] Because of this low number—for comparison, NQF's Quality Positioning System (QPS) currently lists more than 200 endorsed process measures—NQF and CMS are exploring opportunities to increase development of PRO-PMs.[4] One such opportunity is providing stakeholders with a practical methodology that describes how to develop, test, implement, and interpret PRO-PMs.

In addition to the 35 PRO-PMs that are currently endorsed by NQF, 11 PRO-PMs related to HRQoL, experience with care, and symptom and symptom burden are no longer endorsed by NQF.[44] There are many possible reasons why a measure may lose its endorsement status, such as a measure developer's efforts to create a more effective PRO-PM or multiple competing measures that serve the same purpose. The fact that nearly one-third of all historic PRO-PMs are no longer endorsed, however, underlines the importance of creating guidance that supports the development and testing of new performance measures.

CMS Priorities and Reduced Measure Burden

A key priority of CMS' Meaningful Measures 2.0 initiative is reducing measure burden through the use of digital measures.[11] This initiative aims to transform measures into fully digital measures by 2025 and elevate patient voices through the use of patient-reported measures.[11] CMS is committed to ensuring that measures are aligned across value-based programs and incorporate the patient voice. The Office of Burden Reduction and Health Informatics within CMS will be critical in leading, supporting, and coordinating methods that will support burden reduction, interoperability initiatives, and actions to maintain the patient voice across CMS.[12]

Relationship Between PROMs and PRO-PMs

When developing PROM-based performance measures, developers must consider the advantages and disadvantages of specifying a PRO-PM to utilize data from a single PROM or from multiple PROMs. Arguments in favor of a one-to-one relationship (i.e., one PROM captures and contributes data to one performance measure) include the following:

- the ease of tailoring a PRO-PM specification to align with the data structure of a single PROM;
- the relative simplicity of modeling a performance measure on one PROM's scoring methodology;
- the ability for measure developers to focus on one PROM that exhibits high quality attributes; and
- the comparative ease of maintaining a single-PROM performance measure.

In reviewing NQF's QPS, a repository of quality measures that are currently or were at one time endorsed by NQF, it was discovered that the vast majority of PRO-PMs describe a one-to-one relationship. Conversely, NQF #0700, a previously endorsed measure that assesses HRQoL in patients with chronic obstructive pulmonary disease (COPD), does accept data from three PROMs: (1) the Chronic Respiratory Disease Questionnaire (CRQ), (2) the St. George's Respiratory Questionnaire (SGRQ), or (3) the COPD Assessment Test (CAT).[4] Additionally, crosswalks exist that allow scores from one PROM

to be converted for use in another PROM, such as the 2020 publication of a validated crosswalk between the KOOS JR and the OKS.[13] As a result of crosswalks, measures such as NQF #2653, a currently endorsed measure on functional status following total knee replacement surgery that utilizes data from the OKS, were redesigned incorporating the KOOS JR tool with a validated crosswalk score that provides comparability between the two target scores of either an OKS score greater than or equal to 37 or a KOOS JR score greater than or equal to 71 at one year postoperatively. Theoretically, PRO-PMs can accept scores from multiple PROMs, as long as they are converted via a validated crosswalk; in practice, this should not be done unless the developer of the PRO-PM has confirmed that the crosswalk score is valid in the performance measure.

Despite these arguments, the TEP's discussions trend toward support for a many-to-one relationship, in which clinicians can choose from a list of multiple PROMs to inform the performance measure. Importantly, this approach offers clinicians the flexibility to use the PROM that is most appropriate for their setting based on criteria such as licensing costs and availability of validated translations relevant to the patient population. The Environmental Scan Report discusses potential avenues raised by the TEP that developers can use to map the data fields, cut points (i.e., markers in PROMs that indicate the need to screen for a diagnosis or provide treatment), and scores of disparate PROMs to a single performance measure.

Interoperability

In 2011, CMS established the Medicare and Medicaid EHR Incentive Programs to encourage providers, hospitals, and critical access hospitals to adopt, implement, upgrade, and demonstrate meaningful use of certified electronic health record technology (CEHRT). These programs, currently known as the Promoting Interoperability Programs, introduce a new phase of EHR measurement with an increased focus on interoperability and patient access to health information.[14] One pathway to achieving this goal is Fast Health Interoperability Resources (FHIR), a standard from Health Level Seven International (HL7) that exchanges healthcare information electronically through EHRs and other health IT systems.[15]

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, including FHIR and coding schemes, such as Logical Observation Identifiers Names and Codes (LOINC). 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.

Prioritized Domains for the Interim Report

PRO-PMs can be grouped within five domains of PROs: (1) HRQoL, (2) functional status, (3) symptoms and symptom burden, (4) health behaviors, and (5) patient experience.[7] After considering the current state of PRO-PMs and the most pressing needs for PROM-based performance measurement, CMS, NQF, and the TEP agreed to focus this work on HRQoL, functional status, and symptoms and symptom burden. Several factors led the TEP to not focus on the health behavior or patient experience domains. These drivers are described below.

One driver for this decision is how well each domain is addressed in existing NQF-endorsed measures. For example, the TEP decided not to include the patient experience domain because it is relatively well represented in the body of NQF-endorsed PRO-PMs. As of April 1, 2021, 10 of the 35 NQF-endorsed PRO-PMs relate to the patient experience domain.[4] Conversely, HRQoL and symptom-based PRO-PMs

are particularly under-represented in NQF-endorsed performance measures, and while a few endorsed measures of functional status exist, this domain is not as robust as it would ideally be. This led the TEP to determine that this work should focus on the HRQoL, symptom, and functional status domains.

A second driver is the prioritization of domains that more directly assess the clinical performance of healthcare entities (e.g., providers, health systems, and health plans). This is because the Building a Roadmap initiative focuses on areas in which the actions and decisions of healthcare entities more directly influence outcomes.[1] While researchers have demonstrated that clinicians do have the ability to influence the health behaviors of patients, the patient behaviors domain centers on the actions and behaviors of patients; therefore, it is not the appropriate focus for this project. Similarly, patient representatives on the PRO Best Practices TEP noted that the experience domain typically evaluates how a provider or practice serves patients (e.g., how well does the staff communicate with patients, how easy is it to schedule an appointment).[2] Multiple public comments on the PRO Best Practices Final Report also noted an emerging trend to differentiate between PROMs and PREMs, or patient-reported experience measures.

The third driver is the need to focus on performance measures whose underlying data can be independently captured by healthcare entities. PROM data for HRQoL, functional status, and symptoms can typically be collected and analyzed by the healthcare entity that is involved in care. This provides healthcare entities with more autonomy in selecting PROMs and collecting data, and it also supports the aim of administering PROMs to the entire population of eligible patients as opposed to a sample. However, PROMs for patient experience measures, such as the Consumer Assessment of Healthcare Providers and Systems (CAHPS), typically utilize methodologies that require external partners to collect and analyze patient-reported data. Many PROMs that measure experience, such as the Hospital Consumer Assessment of Healthcare Providers and Systems (HCAHPS), are standardized and have stringent rules about administration.[16] Because of the different levels of control over data as well as the methodological differences in how data are captured and analyzed, this report does not focus on patient experience in PRO-PMs.

CMS, NQF, and the TEP recognize that the health behaviors and patient experience domains are both valid and important aspects of PROs. The authors' prioritization of HRQoL, functional status, and symptoms and symptom burden for the Building a Roadmap initiative is not, in any way, intended to diminish the importance of other PRO domains.

Goals and Objectives

The goal of the Interim Report is to help measure developers understand what defines a high quality PROM for use in performance measures and what attributes of a high quality PROM are most conducive to the development of a digital PRO-PM that is appropriate for regulatory purposes. This report will build on the findings of the Environmental Scan Report by identifying attributes of high quality PROMs that are used for performance measures, including digital PRO-PMs and measures used for APMs, VBP programs, and/or Medicare coverage determinations. The report will also describe the process of assessing the attributes to ensure they are complete and accurate.

The Interim Report will detail the recommendations of the TEP regarding the attributes of a high quality PROM. These recommendations are informed by the extensive combined experiences of the

multistakeholder TEP, as well as by pertinent information elicited through a literature review, public comments, and feedback from federal liaisons to the TEP. The Interim Report expands on the information about PROMs in the 2020 NQF report titled Patient-Reported Outcomes: Best Practices on Selection and Data Collection as well as the Building a Roadmap Environmental Scan Report. Additionally, it will inform the ultimate deliverable from the first year of the Building a Roadmap initiative: the Technical Guidance Report.

Attributes of High Quality PROMs for Use in Performance Measures

The previously published PRO Best Practices Final Technical Report includes an Attribute Grid for PROM Selection as an effective tool during the PROM selection process for reviewing and detailing select criteria relevant to a clinical setting. This Attribute Grid provides a systematic method to perform a side-by-side comparison on predetermined and meaningful attributes of PROMs and is intended to guide the selection process by those in the field seeking to select and use PROMs.[2]

Rationale for New Attribute Grid

As noted earlier, NQF initially planned to use the previously published Attribute Grid for PROM Selection (Appendix B: PROM Attribute Grid from 2020 Report) for this project. The TEP reviewed this Attribute Grid and determined that a slightly different set of attributes is needed to review PROMs for performance measurement. As a result, Panel members guided the compilation of a list of attributes of high quality PROMs for use in performance measures. These attributes were developed using the following definition of high quality for this scope of work: **A high quality PROM for performance measurement is a PROM that is suitable to be the foundation for a digital PRO-PM that can be used to evaluate the performance of healthcare entities.**

This definition aimed to include attributes that align with CMS' regulatory purposes and are conducive to developing high-impact digital PRO-PMs that are endorsed by NQF. Additional considerations for the development of this definition follow below. A high-quality PROM for performance measurement must perform the following actions and meet the following criteria:

- capture and acknowledge outcomes that are important to patients;
- include psychometric soundness;
- go above and beyond reliable, valid, feasible, low burden, and low or no cost;
- be tested and reliable in real-world settings with different collection modes;
- include readily interpretable and actionable scores within the definition;
- have assigned LOINC codes; and
- incorporate the quality of PRO-PMs, emphasizing the intersection of PROMs and PRO-PMs.

Attributes and Definitions

The final list of attributes is listed below and within Appendix C: Attributes of High Quality PROMs for Use in PRO-PMs. The TEP developed this list of attributes through extensive discussion, considering a broad audience while ensuring specific and applicable attributes of PROMs for the development of performance measures.

Covers Desired PROs From Patient and/or Caregiver Perspective

PROM-based performance measures are seen as a byproduct of delivering better care and improving patient engagement related to patient treatment, alleviating symptoms, and improving quality of life. A patient member of the TEP emphasized that clarity of the PROM (i.e., understanding how the intent and weight of the questions relate to the outcome) helps patients to understand its relevance to the desired PROs.

Outcome Measured in PROM Is Result of Care for Which Relevant Clinical Quality Is Being Measured

There was strong agreement among the TEP that what the PROM is capturing must be related to clinical care and useful from a clinical perspective for clinicians to use and buy into the PROM. Additionally, the availability of the outcomes at the point of care is important for buy-in and continued and actionable use by clinicians. When selecting PROMs for performance measures, developers should consider whether the PROM measures the area of interest (e.g., measuring depression or total knee replacement would require different PROMs) and uses the appropriate unit of attribution (e.g., care team, clinician, and/or hospital).

Defined and Actionable Cut Points or Targets, Anchors, and/or Defined Meaningful Change

TEP members widely agreed that when building performance measures, the PROM(s) that serve as the basis for the PRO-PM must have defined and comparable cut points or targets. The PROM(s) must measure PROs in a way to effect meaningful change, such as using anchor-based methods to measure minimal clinically important difference (MCID).[17] Cut points must be evidence based.

Clear Conceptual and Measurement Models

Documentation of the conceptual and measurement models should exist. Conceptual and measurement models provide a clear rationale for what a PROM measures and why it is important, which ultimately contributes to interpretability and actionability. The conceptual model identifies the concept(s) included in the PROM as well as the target population, while the measurement model explains how the concept(s) relate to the items in the questionnaire.[18] Clear conceptual and measurement models provide a clear explanation of what is being measured, what relationships exist between the different components of the PROM, and the populations who use it to submit information.

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/or neuroscientific information).[19] Psychometric testing is used in NQF evaluation as evidence of scientific acceptability of measure properties by showing testing reliability, validity, and adequacy of risk adjustment.[20]

Reliability

Reliability testing demonstrates that the measure data elements are repeatable, producing the same 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 precise.[20] For PRO-PMs and composite performance measures, reliability should be demonstrated for the computed performance score.[20]

Reliability testing applies to both the data elements and computed measure score. Examples of reliability testing for data elements include, but are not limited to, inter-rater/abstractor or intra-rater/abstractor studies, internal consistency for multi-item scales, and test-retest for survey items.[20] Reliability testing of the measure score addresses precision of measurement (e.g., signal-to-noise).[20]

Validity

Validity testing demonstrates that the measure data elements are correct and/or the measure score correctly reflects the quality of care provided, adequately identifying differences in quality.[20] For PRO-PMs, validity should be demonstrated for the computed performance score.[20]

Validity testing applies to both the data elements and computed measure score.[20] Validity testing of data elements typically analyzes agreement with another authoritative source of the same information.[20] Examples of validity testing of the measure score include, but are not limited to, testing hypotheses that the measures scores indicate quality of care (e.g., measure scores are different for groups known to have differences in quality assessed by another valid quality measure or method), correlation of measure scores with another valid indicator of quality for the specific topic, or relationship to conceptually related measures (e.g., scores on process measures to scores on outcome measures).[20] Face validity of the measure score as a quality indicator may be adequate if it is accomplished through a systematic and transparent process by identified experts and explicitly addresses whether performance scores resulting from the measure as specified can be used to distinguish good from poor quality.[20]

Responsiveness and/or Actionability

Responsiveness refers to the ability of the clinician or provider to understand and act upon the PROM score or change score. Responsiveness is dependent on the second attribute in the grid: defined and actionable cut points or targets, anchors, and/or defined meaningful change. Clinicians and patients emphasized the importance of this attribute, as buy-in and successful implementation both hinge on providers using these PROMs and acting on their outcomes.

Usability/Feasibility of Use

Low Burden (e.g., Length, Time/Effort to Complete)

Both the patient and clinician realize inherent burden in PROM use and implementation. Methods such as reducing the length of a PROM, using computer-assisted technology (CAT) to streamline unnecessary questions, and using electronic collection as a natural part of the workflow can help to reduce burden.[2] From a measure developer perspective, considering the burden of these PROMs during the build of related performance measures can help to ensure success for these measures.

Fits With Standard of Care and Related Workflows (e.g., Incorporated and Discussed at Point of Care)

In conjunction with actionability and interpretability, the availability of outcomes at the point of care is critical to implementing PROMs and incorporating them within the clinical workflow. When the result of the PROM is available to the clinician during the patient visit, the provider can compare present outcomes to historical results, discuss them with the patient, and act on the scores, all in real time. This allows the clinician to take meaningful action on the results.

Cultural Appropriateness, Language, and Translations With Culturally Appropriate Items

Cultural appropriateness considers whether the questions on a PROM accurately reflect the way life is lived in a culture, society, or population; it is particularly relevant when a PROM is being used in a different culture than the one for which it was developed. For example, a PROM that asks whether a patient can walk on a forest trail after surgery may not be culturally appropriate for patients in areas that do not have forests (e.g., urban areas).

PROMs should not just be translated into different languages but should also be validated and culturally adapted within each additional language offered. Literal translations may not capture important differences or changes that would affect scoring. Preferred language options should represent an organization's patient population.

Standardized Codes Available

To use PROM results as the basis for performance measures, standardized codes such as LOINC are required to translate each patient response or result into a comparable electronic data point. LOINC codes assign a separate code for each unit (e.g., a question on a PROM) to define and standardize results for accurate comparison.[21] Not having codes to translate and store the data creates a challenge when creating a digital performance measure. Different PROMs may measure the same domain dissimilarly; therefore, a structured approach to categorize different PROM scores is necessary for successful PRO-PM development.

Guidance on Standardized Data Collection (Including Modes and Methods)

PROMs being considered for use in performance measures should have clear documentation or research that describes comparability of data gathered via different modes (e.g., was the PROM self-administered or completed during an interview?) and methods (e.g., was the PROM completed via paper, text message, or patient portal?).[2] Different settings often have different limitations with regard to data collection (e.g., technology or internet challenges collecting the information prior to a visit or patients with chronic conditions or special needs may be unable to complete a written questionnaire), and measure developers will benefit from understanding if PROM data are generalizable across settings and populations. Additionally, documentation or literature should explain how a PROM is validated across different modes and methods, along with specific data collection requirements for each.

Additional Considerations About the Attributes

Objectivity and Subjectivity

Most of the attributes in the Attribute Grid can be assessed objectively through a literature review: defined cut points and anchors, reliability, validity, responsiveness, actionability, burden (including tool length and time to complete), number of translations, LOINC codes, and documentation on modes and methods. However, even with these seemingly straightforward attributes, PRO-PM developers may find that assessing an attribute is not obvious. For example, it is easy to find evidence that the PHQ-9 has been translated into more than 90 languages, but determining whether each translation is culturally adapted is subjective as well as difficult (if not impossible).

Some of the attributes are intentionally subjective. The first attribute, "Covers desired PROs from patient and/or caregiver perspective," requires the measure developer to first identify a "desired" PRO and then review PROMs to see which ones measure outcomes that are meaningful to patients. The

presence of subjective attributes is not intended to be a barrier to measure development but to challenge the developers to think creatively and broadly about a performance measure instead of simply relying on quantitative assessments, such as reliability and validity. Additionally, the presence of subjective attributes helps to prompt measure developers to consider measurement from the perspective of what is most meaningful to patients.

Flexibility of the Attribute Grid

As with the PRO Best Practices Attribute Grid, the Building a Roadmap Attribute Grid is a tool that measure developers can use to systematically and consistently review PROMs that might be suitable for use in a PRO-PM. It can and should be used flexibly, and if approved, an optional report in the Building a Roadmap initiative will study how measure developers are using and modifying the Attribute Grid.

The Attribute Grid is not intended to be prescriptive. It does not dictate whether measure developers should select one PROM or multiple PROMs to collect data for a PRO-PM. It does not require that every attribute be met for a PROM to be used in a performance measure, nor does it specify that a certain number of attributes must be met. The attributes are also not definitive: Additional attributes can and should be added if they are suitable for a specific performance measure. The Attribute Grid does not generate a score, nor are pass/fail criteria defined.

Use Case: Assessing the Attribute Grid

During its fourth meeting, the TEP scrutinized and refined the attributes within the Attribute Grid to ensure that:

- the TEP had a clear, shared understanding of the meaning of each attribute;
- no attributes were included erroneously; and
- no attributes were inadvertently omitted from the grid.

To complete this activity, the TEP assessed a use case of the PHQ-9 that was prepared by NQF staff. TEP members compared the PHQ-9 against the attributes contained within the Attribute Grid. Additionally, the TEP members considered whether the PHQ-9 triggered any additional attributes absent from the Attribute Grid. This activity led to further TEP discussions about the intended meanings of several attributes, which resulted in refinements to the Attribute Grid. After the web meeting, NQF staff invited specific TEP members with measure development experience to individually and voluntarily complete a similar use case exercise with a PROM that they either developed or knew well.

NQF selected the PHQ-9 as the PROM for the group use case for several reasons: its use in federal programs (e.g., CMS MIPS #370 and Health Resources and Services Administration [HRSA] Uniform Data System Quality of Care Measures), its 20-plus year history of clinical use, its widespread adoption, the breadth of published research about it, and its use in multiple NQF-endorsed PRO-PMs.[22,23] In funding the Building a Roadmap initiative, CMS identified several PROMs, including the PHQ-9, as potential high quality candidates for analysis throughout the course of this work due to their use in APMs or for Medicare coverage determinations. As noted in the Environmental Scan Report, NQF and CMS accepted developers and stewards of several PROMs and PRO-PMs as members of the TEP due to their unparalleled familiarity with and expertise of certain PROMs. Potential conflicts of interest were

identified during the first meeting of the TEP. The lack of potential conflicts regarding the PHQ-9 made it an ideal instrument for the TEP use case.

The Building a Roadmap initiative does not recommend any PROM, nor does it identify any specific PROM as being "high quality." NQF does not currently endorse, recommend, rank, or prioritize PROMs. This report includes PROMs only as components of the use cases that verify the accuracy and completeness of the Attribute Grid.

The Environmental Scan Report identifies resources that measure developers can use when seeking recommendations for PROMs to consider in the PRO-PM development process; these resources include, but are not limited to, the International Consortium for Health Outcomes (ICHOM) Standard Sets, PROMs recommended by specialty societies or health systems, and the National Institutes of Health (NIH) HealthMeasures initiative. The 2020 PRO Best Practices Report also provides guidance to clinicians and administrators on selecting PROs and PROMs in care delivery settings.

(Please note that Draft 2 of the Interim Report was due before the voluntary use cases could be completed. If any TEP members complete additional use cases, these will be added to Appendix D in Draft 3.)

Use Case Findings

The use case review process found that the TEP accurately identified attributes of high quality PROMs for use in performance measures. NQF made minor adjustments to the Attribute Grid to reflect observations and refinements that emerged during the review process. The Attribute Grid provides an effective resource to help measure developers objectively assess candidate PROMs to determine whether they are suitable as data collection tools for PRO-PMs.

(Please note that Draft 2 of the Interim Report was due to CMS before the findings of any voluntary use cases could be assessed. NQF will update the "Findings" section in Draft 3 if additional relevant information emerges.)

Limitations of the Interim Report and the Attribute Grid

Two primary limitations hindered the development and testing of the Attribute Grid: the contextual factors that influence whether a PROM is "high quality" and the limited resources available to create use cases to assess the Attribute Grid.

Ideally, the TEP could develop static criteria to identify a high quality PROM. However, the TEP quickly determined that high quality is a dynamic term when applied to PROMs. One factor influencing this dynamism is the diversity of situations in which stakeholders can use a PROM. For example, a PROM might be a highly effective screening tool at the patient level but might not be as effective when applied to a healthcare entity. In this case, the PROM is high quality in one setting but not in another. Similarly, a PROM that is proven to be reliable and valid for measuring post-surgical outcomes at a 12-month interval might not be effective at 12 weeks; this PROM is high quality for a PRO-PM that measures postoperative outcomes at one year but is low quality at three months. A second contextual factor that influences the dynamic nature of high quality PROMs is the length of time that a PROM has been publicly available. This report developed a use case on the PHQ-9 because, in part, the PROM had been

in existence for more than 20 years, and the developers' website lists more than 400 articles studying different facets of the instrument.[24] For newer PROMs, however, the amount of available information on an instrument's attributes may be limited due to a lack of funding for the developers to test the PROM across multiple populations, inadequate resources to develop guidance on data collection standards, lack of time for researchers to develop and test culturally appropriate translations, or lack of widespread clinical adoption to generate novel research about using the PROM in different settings and populations. As such, there is no universal definition of a high quality PROM, but rather a conditional definition that depends on the context in which the PROM is being used.

Given the vast number of PROMs that are used in various federal programs and widely adopted by various clinical specialties, NQF and the TEP would have liked to create additional use cases and invest more time in testing these use cases. However, the Building a Roadmap initiative was designed to be completed with a finite amount of time and resources, and the assessment of the Attribute Grid was necessarily limited. Despite this constraint, NQF and the TEP find the Attribute Grid to be a valuable and practical resource for measure developers to use when assessing candidate PROMs to determine whether they are suitable to use with digital PRO-PMs that are intended for CMS VBP programs, APMs, or other regulatory and/or payment programs.

Conclusion

Stakeholders in the measure development process must clearly understand the aspects of performance they strive to measure. Even when this clarity exists, however, measure developers can struggle with determining which PROM or PROMs will best capture PROs data for performance measurement. Although it would be impossible to generate a simple list of high quality PROMs that are suitable for all potential performance measures, the TEP identified a set of attributes that measure developers can use to assess candidate PROMs and select those that are well suited for a specific digital PRO-PM. The similarities and differences between this Attribute Grid and the grid presented in NQF's September 2020 PRO Best Practices Report highlight the importance of psychometric soundness and the need for measure developers to consider the context in which a PROM will be used before selecting instruments. The Attribute Grid adds a much-needed tool for measure developers and will be an important part of the guidance presented in the Final Technical Report for Building a Roadmap From Patient-Reported Outcome Measures to Patient-Reported Outcome Performance Measures.

References

1 Santana MJ, Manalili K, Jolley RJ, et al. How to practice person-centered care: A conceptual framework. *Health Expect*. 2018;21(2):429-440.

2 National Quality Forum (NQF). *Patient-Reported Outcomes: Best Practices on Selection and Data Collection - Final Technical Report*. 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.

3 Squitieri L, Bozic KJ, Pusic AL. The role of patient-reported outcome measures in value-based payment reform. *Value in Health*. 2017;20(6):834-836.

4 NQF. NQF: Quality Positioning System™. https://www.qualityforum.org/QPS. Last accessed January 2021.

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

6 Cella D, Hahan E, Jensen S, 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.

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

8 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.

9 NQF, PatientsLikeMe. Measuring what matters to patients: Innovations in integrating the patient experience into development of meaningful performance measures. August 2017. http://www.qualityforum.org/WorkArea/linkit.aspx?LinkIdentifier=id&ItemID=85770.

10 Patient-reported outcome performance measures: Current environment and next steps. https://www.phrma.org/-/media/Project/PhRMA/PhRMA-Org/PhRMA-Org/PDF/P-R/PhRMA-PRO-PM-Issue-Brief-120417_Final.pdf. Published 2017. Last accessed April 2021.

11 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.

12 Office of Burden Reduction and Health Informatics | CMS. https://www.cms.gov/About-CMS/Agency-Information/CMSLeadership/Office_OBRHI. Last accessed April 2021.

13 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.

14 Centers for Medicare & Medicaid Services (CMS). Promoting Interoperability Programs | CMS. https://www.cms.gov/Regulations-and-Guidance/Legislation/EHRIncentivePrograms. Last accessed February 2021.

15 Health Level 7 International. Overview - FHIR v4.0.1. https://www.hl7.org/fhir/overview.html. Last accessed February 2021.

16 Health Services Advisory Group. CAHPS Hospital Survey (HCAHPS) quality assurance guidelines, v. 16.0. March 2021. https://hcahpsonline.org/globalassets/hcahps/quality-assurance/hcahps-qag-v16.0-march-2021.pdf. Last accessed April 2021.

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

18 NQF. *NQF: Patient-Reported Outcomes Environmental Scan*.; 2019. http://www.qualityforum.org/Publications/2019/12/Patient-Reported_Outcomes_Environmental_Scan.aspx. Last accessed December 2020.

19 Wright JD, ed. *International Encyclopedia of the Social & Behavioral Sciences*. 2. ed. Amsterdam: Elsevier; 2015.

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

21 Forrey AW, McDonald CJ, DeMoor G, et al. Logical observation identifier names and codes (LOINC) database: a public use set of codes and names for electronic reporting of clinical laboratory test results. *Clin Chem*. 1996;42(1):81-90.

22 CMS. Quality ID #370 (NQF 0710): Depression Remission at Twelve Months. November 2020. https://qpp.cms.gov/docs/QPP_quality_measure_specifications/CQM-Measures/2021_Measure_370_MIPSCQM.pdf. Last accessed May 2021.

23 Bureau of Primary Health Care. Uniform Data System: Reporting Instructions for Calendar Year 2020 Health Center Data. August 2020. https://bphc.hrsa.gov/sites/default/files/bphc/datareporting/pdf/2020-uds-manual.pdf. Last accessed May 2021.

24 PHQScreeners. Screener Overview. Patient Health Questionnaire (PHQ) Screeners. https://www.phqscreeners.com/select-screener. Last accessed April 2021.

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

26 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.

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 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.

30 PHQScreeners. Instruction Manual: Instructions for Patient Health Questionnaire (PHQ) and GAD-7 Measures. https://www.phqscreeners.com/images/sites/g/files/g10016261/f/201412/instructions.pdf. Last accessed April 2021.

31 Kroenke K, Spitzer RL, Williams JBW. The PHQ-9: Validity of a brief depression severity measure. *J Gen Intern Med*. 2001;16(9):606-613.

32 Oncology Nursing Society. Patient Health Questionnaire (PHQ9): English. https://www.ons.org/assessment-tools/patient-health-questionnaire-phq9-english. Last accessed April 2021.

33 Baas KD, Cramer AOJ, Koeter MWJ, et al. Measurement invariance with respect to ethnicity of the Patient Health Questionnaire-9 (PHQ-9). *J Affect Disord*. 2011;129(1-3):229-235.

34 Reich H, Rief W, Brähler E, et al. Cross-cultural validation of the German and Turkish versions of the PHQ-9: an IRT approach. *BMC Psychology*. 2018;6(1):26.

35 Lupascu N, Timar B, Albai A, et al. Validation and cross-cultural adaptation of the depression Patient's Health Questionnaire 9 in the Romanian population of patients with Type 2 Diabetes Mellitus. *DMSO*. 2019;12:841-849.

36 LOINC. Panel Browser: Patient Health Questionnaire (PHQ). LOINC from Regenstrief. https://loinc.org/panels/category/survey-instruments/behavioral-health-psychiatry-substance-abuse-survey-instruments/patient-health-questionnaire-phq/. Last accessed April 2021.

37 Fann JR, Berry DL, Wolpin S, et al. Depression screening using the Patient Health Questionnaire-9 administered on a touch screen computer. *Psychooncology*. 2009;18(1):14-22.

38 Pinto-Meza A, Serrano-Blanco A, María MT, et al. Assessing Depression in Primary Care with the PHQ-9: Can It Be Carried Out over the Telephone? *Journal of general internal medicine*. 2005;20:738-742.

Appendices

Appendix A: Glossary of Terms

Alternative Payment Models

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.[25]

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.[17]

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.[8]

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.[8]

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.[11] 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.[26]

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.[5]

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. non-profit 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.[17]

Patient-Reported Outcome (PRO)

Any information on the outcomes of healthcare obtained directly from patients without modification by clinicians or other healthcare professionals.[5]

Patient-Reported Outcome Measure (PROM)

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.[5]

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.[5]

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. Validity and reliability are key aspects to attaining psychometric soundness.[8] 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).[1919]

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.[29]

Appendix B: PROM Attribute Grid From 2020 Report

The following Attribute Grid was published in the Final Technical Report for the CMS-funded initiative titled Patient-Reported Outcomes: Best Practices on Selection and Data Collection, which was published in September 2020. The Attribute Grid was presented as a tool to aid in the comparison and selection of PROMs for use in clinical settings, and guidance on the recommended use of the Attribute Grid was presented in the report on pages 12-13 and 41-47. In its original form, certain rows repeated multiple

times to ensure key selection criteria were captured. For example, "Covers desired PROs" was repeated across five rows to ensure all PROs were represented during PROM selection.

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				
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):**

PROM	PROM 1	PROM 2	PROM 3	PROM 4
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 C: Attributes of High Quality PROMs for Use in PRO-PMs

The following example 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 result of care for which relevant clinical quality is being measured				
Defined and actionable cut points or targets, anchors, and/or defined meaningful change				
Clear conceptual and measurement models				
Psychometric Soundness: Reliability				
Psychometric Soundness: Validity				
Psychometric Soundness: Responsiveness and/or actionability				
Usability/Feasibility of Use: Low burden, including length of tool and time and effort to complete				

Usability/Feasibility of Use: Fits with standard of care and related workflows (e.g., incorporated and discussed at point of care)				
Usability/Feasibility of Use: • Cultural appropriateness • Language • Culturally adapted translations (i.e., available and validated in multiple languages)				
Usability/Feasibility of Use: Standardized codes (e.g., LOINC) available				
Usability/Feasibility of Use: Standardized data collection (to include mode and methods, applicable guidance)				

Appendix D: PHQ-9 Use Case

The following example shows a completed Attribute Grid that assesses one PROM: the PHQ-9.

ATTRIBUTE	PROM 1: PHQ-9
Covers desired PROs from patient and/or caregiver perspective	The PHQ-9 tool is validated for use as a measure to assess the level of depression severity (for initial treatment decisions) as well as an outcome tool (to determine treatment response).[4]
Outcome measured in PROM is result of care for which relevant clinical quality is being measured	The outcome of PHQ-9 reflects change over time in response to treatment interventions (e.g., counseling, pharmacotherapy, and/or psychotherapy) and other factors that may or may not be related to treatment.
Defined and actionable cut points or targets, anchors, and/or defined meaningful change	Guide for Interpreting PHQ-9 Scores[30] • Score of 0-4; no/minimal depression; no proposed treatment actions. • Score of 5-9; mild depression; watch and wait and repeat PHQ-9 at follow-up. • Score of 10-14; moderate depression; develop a treatment plan that considers counseling, follow-up, and/or pharmacotherapy. • Score of 15-19; moderate/severe depression; treat using pharmacotherapy and/or psychotherapy. • Score of 20-27; severe depression; immediately initiate pharmacotherapy and consider expedited referral to a mental health specialist. Sensitivity: 88% (PHQ-9 score >= 10)[31] Specificity: 88% (PHQ-9 score >= 10)[31] Positive Predictive Value: 31% (cut point = 9) to 51% (cut point = 15)[31]

ATTRIBUTE	PROM 1: PHQ-9
Clear conceptual and measurement models	The conceptual and measurement models are clearly documented on www.phqscreeners.com.
Psychometric Soundness: Reliability	Internal Reliability: Excellent – Cronbach's α of 0.89 in PHQ Primary Care Study and 0.86 in PHQ Ob-Gyn Study[31] Test-Retest Reliability: Excellent – Correlation was 0.84 between PHQ-9 completed by patient in the clinic and telephonic administration by a mental health professional within 48 hours; mean scores were nearly identical (5.08 vs 5.03).[31]
Psychometric Soundness: Validity	Criterion validity: demonstrated in 580 primary care patients who underwent an independent reinterview by a mental health professional[31] Construct validity: established by strong association between PHQ-9 scores and functional status, disability days, and symptom-related difficulty[31] External validity: replicated findings from 3,000 primary care patients in a sample of 3,000 OB/GYN patients[31] ROC analysis: The area under the curve for the PHQ-9 in diagnosing major depression was 0.95, suggesting a test that discriminates well between persons with and without major depression.[31]
Psychometric Soundness: Responsiveness and/or actionability	Major depressive disorder (MDD) is suggested if the following two cases occur:[31] • Of the 9 items, 5 or more are checked as at least "more than half the days" • Either item #1 or 2 is checked as at least "more than half the days" Other depressive syndrome is suggested if the following two cases occur:[31] • Of the 9 items, 2 to 4 are checked as at least "more than half the days" • Either item #1 or 2 is checked as at least "more than half the days"
Usability/Feasibility of Use: Low burden, including length of tool and time and effort to complete	Examples of relevant attributes include the following: • Average time to complete: 3.6 minutes[32] • Number of questions: 9 • Administration: self-administered[31] • Scoring: less than one minute (time to add nine single-digit numbers)[31]
Usability/Feasibility of Use: Fits with standard of care and related workflows (e.g., incorporated and discussed at point of care)	PHQ-9 is well documented as being effective in primary care settings (e.g., PHQ-2 given at check-in, positive result leads to PHQ-9 administered by a medical assistant).

ATTRIBUTE	PROM 1: PHQ-9
Usability/Feasibility of Use: • Cultural appropriateness • Language • Culturally adapted translations (i.e., available and validated in multiple languages)	Official website offers translation into more than 90 languages but notes that few translations have been validated with an independent structured psychiatric interview.[24,30] Although a thorough assessment of cultural appropriateness and adaptation is outside the scope of this use case, a cursory search of peer-reviewed articles shows cultural validations of the PHQ-9 in Surinam Dutch and Dutch men and women[33], Turkish immigrants and Germans[34], and Romanians with type 2 diabetes mellitus.[35]
Usability/Feasibility of Use: Standardized codes (e.g., LOINC) available	The LOINC Panel Browser offers standard codes for PHQ-9 (LOINC 44249-1) as well as other versions.[36]
Usability/Feasibility of Use: Guidance on standardized data collection (to include mode and methods)	Numerous studies exist on modes and methods within specific populations, such as use of touchscreens for patients with cancer[37] and telephonic administration in primary care settings.[38]

Appendix E: 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
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, National Cancer Institute

Ashley Wilder Smith, PhD, MPH
National Institutes of Health, 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
US Department of Veterans Affairs

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

NQF Staff

Sheri Winsper, RN, MSN, MSHA
Senior Vice President, Quality Measurement

Maha Taylor, MHA
Managing Director

Chuck Amos, MBA
Director

Teresa Brown, MHA, MA
Senior Manager

Bejier Edwards, MBA, PMP
Project Manager

Jhamiel Prince, MA
Analyst

Juanita Rogers, MS, CHES
Analyst

CMS Staff

LaWanda Burwell
Task Order COR

Sophia Chan
Senior Research Scientist

Patrick Wynne
IDIQ COR

Maria Durham
Director, DPMS/QMVIG/CCSQ

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

Appendix F: Public Comments and TEP Response

(Please note that NQF staff will add Public Comments and TEP responses to Draft 3 of the Interim Report.)