NATIONAL QUALITY FORUM
Driving measurable health improvements together

Building a Roadmap From Patient-Reported Outcome Measures to Patient-Reported Outcome Performance Measures: Interim Report

FINAL DRAFT

August 18, 2021

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

Contents

Note to Readers

Given the technical nature of this Interim Report, the authors included a Terminology section within the report as well as a Glossary in Appendix A. The authors encourage readers, particularly those who do not have an advanced understanding of the measure development process for patient-reported outcome performance measures (PRO-PMs), to review these sections before reading the report.

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 of all experience levels on developing a fully tested PRO-PM for use in CMS accountability programs. The Interim Report contributes to this goal by identifying key attributes of a high quality patient-reported outcome measure (PROM). For the purposes of this Report, a high quality PROM for performance measurement is one that is suitable to be the foundation of a digital PRO-PM that can be used to evaluate the performance of healthcare entities. This Interim Report presents an Attribute Grid that measure developers can use to assess, compare, and select PROMs for use in PRO-PMs and provides guidance on determining whether a PROM is high quality.

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 the successful development and implementation of PRO-PMs are 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 provide guidance to measure developers on the development, testing, and implementation of PROM-based digital performance measures.

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

1. An Environmental Scan Report that outlines the current state of guidance on and practice in developing PROM-based performance measures
2. An Interim Report that identifies and describes the attributes of high quality PROMs for use in digital PRO-PMs
3. A Technical Guidance Report that guides measure developers of all experience levels 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 reflects the work of the Building a Roadmap Technical Expert Panel (TEP) across six web meetings. The TEP identified, described, and refined 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. The attributes identified by the TEP are described in detail in the body of the report and 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 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 confirm that 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

Healthcare stakeholders, including CMS, have called for an increased emphasis on more efficient and effective quality measurement that is focused on outcomes, patient-reported measures, and digital measures.[1] Patient-centered care aims to ensure that the patient voice is included in both treatment and care delivery and that priority health outcomes are based largely on the patient's health goals.[2] 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.[3] 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 target 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 determining provider reimbursement.[4] 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 29 PRO-PMs were endorsed by NQF, making up approximately 6 percent of the total number of NQF-endorsed quality measures.[5] The CMS-funded Building a Roadmap initiative aims to increase the development of PRO-PMs and prepare developers for the NQF endorsement process by providing step-by-step guidance on developing digital PRO-PMs. To accomplish this, the initiative provides 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. The focus on digital measurement offers several advantages: collecting data and calculating scores with minimal burden; maximizing response rates, and thus, patient representativeness; and leveraging electronic health records (EHRs), registries, and other tools/systems.

NQF is leading the Building a Roadmap 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.

The initial year of the Building a Roadmap initiative centers 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.

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 builds 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 also presents an Attribute Grid that measure developers can use to compare and select PROMs for use in PRO-PMs and provides guidance

on determining whether a PROM is well suited for performance measurement. It also describes the process of assessing the attributes to ensure they are complete and accurate.

The Interim Report details 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 informs the ultimate deliverable from the first year of the Building a Roadmap initiative: the Technical Guidance Report.

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 specific PROMs only as examples, e.g., components of the use cases that verify the accuracy and completeness of the Attribute Grid.

Background

NQF has a long history with PROs, PROMs, and PRO-PMs, and key publications are listed below. This Interim Report is not intended to replace or redo any of this past work, but rather to build on it in ways that offer more structure and guidance to measure developers who are selecting PROMs for use in PRO-PMs. The section titled Commonalities and Differences Between the Interim Report and Past Reports expands upon the relationship between the Interim Report and past reports.

Building a Roadmap is the second of two recent CMS-funded initiatives that emerged in response to 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 in 2019. 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 to describe how to move from the implementation of PROs and PROMs to the development of digital PRO-PMs.

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.[6]	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 enable comparison of patient groups or providers.[6]	Patient Health Questionnaire 9 (PHQ-9)©, a standardized tool to assess depression
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.[6]	Percentage of patients with diagnosis of major depression or dysthymia and initial PHQ-9 score >9 with a follow-up PHQ-9 score <5 at 6 months (NQF #0711)

The Interim Report specifically discusses *attributes* of PROMs, a term that is similar to others used in the field of performance measurement and in this report. Therefore, it is important to explicitly differentiate between attributes, attribution, and the Attribute Grid.

- **Attribute:** Throughout the Interim Report and the entire Building a Roadmap initiative, the word *attribute* is used to describe certain characteristics or traits of a PROM that make it suitable for use in a PRO-PM. Although the 2013 report titled *Patient Reported Outcomes in Performance Measurement* used the word *characteristics*, the authors of this report use *attributes* to remain consistent with the terminology in the 2020 PRO Best Practices work.
- **Attribute Grid:** An *attribute grid* in the context of NQF's work on PROs is a table designed to facilitate a side-by-side comparison of different PROMs against a given set of attributes. The 2020 PRO Best Practices Final Technical Report presents an attribute grid specifically intended to help decision makers in clinical settings select PROMs using a set of attributes that help to identify those that are most suitable for their patient populations and clinical priorities. This Interim Report presents an attribute grid for a different purpose than the original: to help measure developers select PROMs using a set of attributes that help to identify those that are well suited for use in digital PRO-PMs.

- **Attribution:** This is 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.[7]

Brief NQF History With PROs

Measure developers, clinicians, and researchers have developed hundreds of PROMs that collect patient-level outcomes data. While NQF recognizes 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 clinical practices, health systems, and health plans. As such, NQF endorses PRO-PMs but does not endorse instruments or scales (including PROMs) on their own.[6] NQF's Scientific Methods Panel (SMP) will review any PROM(s) identified in the specification of a PRO-PM for reliability and validity as part of the endorsement process. However, NQF remains agnostic to the specific PROM and reviews only the extent to which it meets an acceptable scientific standard as an element of the PRO-PM.

NQF-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).

Over the past decade, NQF has also 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. The Environmental Scan Report provides more details on these reports listed below:

- 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 is the first of two CMS-funded reports commissioned by NQF focused on selecting PROMs for use in performance measurement.[8] (The report was updated in 2015 by its authors.[9])
- 2012: PRO-Based Performance Measures for Healthcare Accountable Entities—The second commissioned report focused on the reliability and validity of PRO-PMs.[10]
- 2013: Patient-Reported Outcomes in Performance Measurement—Henceforth referred to as the *2013 PRO Report*, this is the result of a CMS-funded project in which NQF convened a TEP whose reports laid the groundwork for future PRO-PM development, testing, endorsement, and implementation.[6]
- 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 (RWJF)-funded report that reiterated the importance of patient-centered quality measurement and online patient communities to measure developers.[11]
- 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.[3]
- 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.

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 resources and guidance for 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 topics:

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

Limited Number of PRO-PMs

Significant challenges, including the resources (e.g., finances, time, and staff) to develop PRO-PMs and the lack of detailed guidance on the development process, hinder the development and implementation of PRO-PMs.[12] At the time of this publication, only 29 PRO-PMs were endorsed by NQF.[5] 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 the development of PRO-PMs.[5] The Building a Roadmap initiative is one such opportunity to provide guidance to measure developers.

CMS Priorities and Reduced Measure Burden

A key priority of CMS' Meaningful Measures 2.0 initiative is reducing measure burden to clinicians through the use of digital measures.[1] This initiative aims to transform measures into fully digital measures by 2025 and elevate patient voices through the use of patient-reported measures.[1] 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 ensure the inclusion of the patient voice across CMS.[13] The attributes within this Interim Report will assist measure developers with identifying PROMs that are well suited for the digital, patient-centered measurement that these priorities require.

Relationship Between PROMs and PRO-PMs

When developing PROM-based performance measures, developers must consider the advantages and disadvantages of specifying a PRO-PM that relies on data from a single PROM (i.e., a one-to-one relationship) as opposed to a PRO-PM that can employ data from multiple PROMs (i.e., a many-to-one relationship).

A review of NQF's QPS, a repository of quality measures that are currently or were at one time endorsed by NQF, revealed that the vast majority of PRO-PMs have a one-to-one relationship with a specific PROM. Conversely, NQF #0700, a previously endorsed measure that assesses HRQoL in patients with chronic obstructive pulmonary disease (COPD), can utilize data from three different PROMs: (1) the Chronic Respiratory Disease Questionnaire (CRQ), (2) the St. George's Respiratory Questionnaire (SGRQ), and (3) the COPD Assessment Test (CAT).[5]

Arguments in favor of a one-to-one relationship 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.

Developers may prefer a many-to-one relationship, in which the measure developer includes more than one PROM in the PRO-PM technical specification. Arguments in favor of a many-to-one relationship include the following:

- the ability to provide the accountable entity (e.g., clinician, facility, system, or plan) flexibility to use the PROM that is most appropriate for their setting;
- the flexibility to collect data from multiple PROMs in the same domain (e.g., a single PRO-PM could accept data from the Short Form 12 [SF-12], the Veterans RAND 12 Item Health Survey [VR-12], and the Patient-Reported Outcomes Measurement Information System Global Health [PROMIS Global Health] PROMs);
- the capability to accept data from PROMs that are freely available as well as those with a licensing fee; and
- the ability to support PROMs that have translations or are more culturally appropriate for specific patient populations.

During their discussion, the TEP's support trended toward the use of PRO-PMs with a many-to-one relationship with PROMs.

When deciding between one-to-one and many-to-one relationships, developers must consider crosswalks that estimate scores between PROMs when developing a PRO-PM. Crosswalks can benefit one-to-one relationships by allowing scores from one PROM to be converted for use in another PROM (e.g., the 2020 publication of a validated crosswalk between the Knee Injury and Osteoarthritis Outcome Score for Joint Replacement [KOOS JR] and the Oxford Knee Score [OKS][14]). However, crosswalks do not exist for every PROM, and the estimated scores introduce some risk of measurement error that does not exist in many-to-one relationships. For example, NQF #2653, a currently endorsed measure on functional status following total knee replacement surgery based on the OKS, was redesigned incorporating the KOOS JR tool with a validated crosswalk score that provides comparability between the two target scores (either an OKS score ≥ 37 or a KOOS JR score ≥ 71 one year postoperatively). PRO-PMs can theoretically accept scores from PROMs via validated crosswalks, but the measure developer should confirm that the translated score equates to the original score's interpretation.

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, introduced a new phase of EHR measurement with an increased focus on interoperability and patient access to health information.[15] Since 2011, there has been much work done related to moving interoperability forward. One pathway to achieving this goal is Fast Health

Interoperability Resources (FHIR), a standard from Health Level Seven International (HL7) that can be used to exchange healthcare information electronically through EHRs and other health IT systems.[16]

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 with allowing terminologies such as LOINC to include codes for PROM subscales and total scores, which may prevent the PROM from being used as part of a digital measure. This integration requires concerted and committed effort between stakeholders in the testing and implementation process, particularly measure developers, health IT vendors, and clinical sites where testing will occur. Interoperability is critical to the advancement of digital PRO-PMs; this report encourages measure developers to consider each PROM's ability to interact with EHRs and other health IT systems when selecting high quality PROMs for use in performance measures.

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.[9] 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 when identifying attributes of high quality PROMs that are suitable for performance measurement. Several factors led the TEP to not focus on the health behavior or patient experience domains in this report. 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 July 1, 2021, 10 of the 29 NQF-endorsed PRO-PMs relate to the patient experience domain.[5] Conversely, HRQoL and symptom-based PRO-PMs are particularly underrepresented in NQF-endorsed performance measures. While there are a few endorsed measures of functional status, this domain is not as robust as it would ideally be. This led the TEP to determine that the Interim Report 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 are anticipated to 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).[3] Multiple public comments on the PRO Best Practices Final Report also noted an emerging trend to differentiate between PROMs and patient-reported experience measures (PREMs).

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, in part to ensure patient responses are confidential and will not affect the delivery of care. 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.[17] 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. Additionally, mixed mode data collection (e.g., mail, web, telephone administration, and personal interviews) is currently recommended for CAHPS surveys, making this domain less appropriate for digital measurement.[18]

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 Interim Report is not, in any way, intended to diminish the importance of other PRO domains.

Attributes of High Quality PROMs for Use in Performance Measures

To adequately assess and ensure the use of high quality PROMs that are suitable for performance measurement, a similar approach to the previous 2020 PRO Best Practices Final Technical Report was explored. By using a consistent Attribute Grid approach, these reports build on one another and offer ease of use for the reader. The respective grids capture important attributes in determining which PROM(s) are appropriate for the desired outcomes.

Discussion of PRO Best Practices Attribute Grid

The 2020 PRO Best Practices Final Technical Report included an Attribute Grid developed by that initiative's 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."[3] As illustrated in Appendix B, decision makers in clinical settings can use the grid to easily compare the attributes of different PROMs that measure similar PROs (e.g., the OKS and the KOOS JR for functional outcomes after a total knee replacement surgery) to determine which is most applicable to their 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."[3]

Rationale for New Attribute Grid

When designing the Building a Roadmap initiative, NQF initially planned to use the PRO Best Practices Attribute Grid as-is to identify high quality PROMs for use in performance measures. The TEP reviewed

this Attribute Grid and determined that a slightly different set of attributes is needed since a PROM that is well suited for clinical settings might not be well suited for digital performance measurement.

As a result, TEP 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 foster the inclusion of PROM attributes that align with CMS' regulatory purposes and are conducive to developing high-impact digital PRO-PMs that can be endorsed by NQF. Additional considerations for the development of this definition follow below. A high quality PROM for performance measurement must:

- capture outcomes that are important to patients;
- be psychometrically sound;
- be usable and feasible in a diverse array of clinical settings;
- be successfully tested in settings in which data are collected using different collection modes;
- include readily interpretable and actionable scores within its documentation (e.g., specifications); and
- have assigned LOINC codes or other mechanisms that make it suitable for digital measurement.

Attributes and Definitions

The final list of attributes is detailed below, as well as in 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 an opportunity to reflect the outcomes of care, as well as increased patient engagement in treatment, alleviating symptoms, and improving quality of life. There are recent transformations within healthcare with the goal to increase a person's control of the planning and delivery of their healthcare services and supports.[19] To be truly person-centered, outcome measurement must include informed choice and reflect people's meaningful priorities, goals, and desires.[19] A patient member of the TEP emphasized that clarity of the PROM (i.e., a clear connection between the intent and weight of the questions and the domain) helps patients to understand its relevance to the desired PROs. While assessment of this attribute is subjective, examples of factors that measure developers might consider include documentation of whether and how patients were included in the development of the PROM, average patient response rates compared to other PROMs used for the same PRO, peer-reviewed literature reflecting effective use of the PROM with marginalized populations, or a demonstrably clear connection between the questions and the desired PROs.

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

There was strong agreement within the TEP that what the PROM captures must be influenced by clinical care and provide actionable information to clinicians. Additionally, the availability of the outcomes at

the point of care is important for buy-in and sustained use by clinicians, as well as increased engagement from patients. When selecting PROMs for performance measures, developers should consider whether the PROM measures the outcome of interest and can be attributed to the appropriate members of the healthcare team.

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

A PROM that is usable for PRO-PMs should have clear documentation on how to interpret scores.[9] TEP members widely agreed that when building performance measures, the PROM that serves as the basis for the PRO-PM must have defined cut points or targets. If multiple PROMs can form the basis of the PRO-PM, comparable cut points must be established and evidence based throughout the different PROMs. The PROM(s) must measure PROs in a way to indicate what difference in score comprises a meaningful change, such as using anchor-based methods to measure minimal clinically important difference (MCID).[20] There also may be challenges interpreting changes in PROM results at the individual level, given that PROMs are generally developed and validated at the group level. High quality PROMs should include the appropriate statistical information for interpretation at both levels.

Clear Conceptual and Measurement Models

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 defines and describes the concept(s) included in the PROM as well as the target population, while the measurement model explains how the concept(s) relate to each other as well as to the items in the questionnaire.[9] 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 that use it to submit information.

Psychometric Soundness

Psychometrics is a field of study concerned with the theory and methods of psychological measurement. It deals with how psychological constructs (e.g., intelligence, neuroticism, or depression) can be optimally related to other sources of data (e.g., outcomes of psychological tests, genetic profiles, and/or neuroscientific information).[21] Psychometric soundness refers to the PROM's psychometric characteristics of reliability, validity, and responsiveness.

Reliability

Reliability of a PROM means the instrument is free from random error. Reliability is well defined in the literature and should demonstrate internal consistency of multi-item scales and reproducibility (or stability over time) through test-retest reliability.[8,9]

Validity

Validity means the PROM reflects what it is intended to measure.[8,9] As with reliability, validity testing for a PROM is well defined in the literature and includes content validity (i.e., how well a PROM samples a representative range of the content it is intended to measure), construct and criterion-related validity (i.e., evidence that the PROM supports the expected results that are similar or dissimilar to the measured PRO, as well as the expected differences between known populations and that the PROM's scores are related to a criterion measure), and responsiveness, as described below.

Responsiveness

Responsiveness refers to the ability of the PROM to detect change in the patient's clinical condition over time and in response to clinical interventions. Responsiveness is dependent on the second attribute in the grid: defined and actionable cut points or targets, anchors, and/or defined meaningful change.[9]

Usability/Feasibility of Use

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

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 select the most pertinent questions for the individual, and using electronic collection, such as integration into an EHR vendor platform, can further incorporate PROM collection as a natural part of the workflow to reduce burden.[3] 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. Burden contributes to a PROM's overall feasibility and is one factor that affects whether clinical and support staff will actively engage patients in completing PROMs, as well as whether patients will participate in completing questionnaires that collect the requested data.

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

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 current outcomes to historical results, plot their course over time and relative to clinical events or interventions, discuss them with the patient, and act on the scores, all in real time. This allows the clinician and patient 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 only be translated into different languages but should also be culturally adapted and validated within each language offered. Literal translations will generally fail to capture important differences or changes that would affect scoring. Preferred language options should represent an organization's patient population.

Availability of Standardized Clinical Terminology and Codes

To use PROM results as the basis for digital 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.[22] Not having defined clinical terminologies and associated 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 and consistent approach to organize PROM results is necessary for successful PRO-PM development. For further detail, please see the

crosswalk definition within Appendix A: Glossary of Terms or the crosswalk discussion above in the Relationship Between PROMs and PRO-PMs section.

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 text message or patient portal?).[3] Different settings present different limitations with regard to data collection (e.g., technology or internet challenges that limit pre-visit data collection, patients with chronic conditions or special needs who may be unable to complete a written questionnaire). Measure developers will benefit from understanding whether PROM data are generalizable across settings and populations. Additionally, documentation or literature should explain how a PROM is developed and 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; however, determining whether each translation is culturally adapted is subjective as well as difficult (if not impossible). This is primarily because these data are not uniformly available.

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 determine those that measure outcomes which 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 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 and should use to systematically and consistently review PROMs that might be suitable for use in a PRO-PM. However, the Attribute Grid is also designed to be flexible.

The Attribute Grid is not intended to be prescriptive; nevertheless, the TEP strongly encourages measure developers to use it when selecting PROM(s) for use in a PRO-PM. The Attribute Grid 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 exhaustive: Additional attributes can and should be added if they are suitable for a specific performance measure or

population; 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 does not generate a score, and pass/fail criteria are not defined within the attributes.

Commonalities and Differences Between the Interim Report and Past Reports

The Interim Report and the attributes it contains are not intended to replace the important recommendations from previous NQF reports that offer guidance on identifying PROMs for use in PRO-PMs, including the Methodological Issues Report and the 2013 PRO Report, as well as the 2015 publication titled Patient-Reported Outcomes in Performance Measurement. Instead, this report offers an opportunity to closely review the past recommendations in light of the evolution in digital healthcare and the changes that have occurred, identify those that remain accurate and essential today, and expand upon novel recommendations for attributes of high quality PROMs for use in digital PRO-PMs.

After their discussion concluded, the TEP reached consensus that the "characteristics" from the Methodological Issues Report and the 2013 PRO Report (i.e., Conceptual and Measurement Model, Reliability, Validity, Interpretability of Scores, Burden, Alternative Modes and Methods, Cultural and Language Adaptations, and EHRs) remain significant today, reiterating the relevance of the prior work, and are either directly or indirectly addressed in this Interim Report. While the Interim Report does cite the earlier work, particularly with regard to attributes of psychometric soundness, the majority of the attribute definitions in this report contain additional information that was recommended by the TEP.

One notable expansion relevant to digital PRO-PMs is the intersection between attributes related to EHRs in the past reports and the attribute of "Availability of Standardized Clinical Terminology and Codes" in the current work. While the earlier reports discussed EHRs at a fairly high level, significant work has occurred in the past eight years to advance knowledge and functionality on how PROMs and PRO-PMs interact. While significant work remains in the area of digital PRO-PMs, the Interim Report (and the upcoming Technical Guidance Report) offers more guidance to measure developers about selecting PROMs that are well suited for digital measurement.

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 used a PROM that may be suitable to use with digital PRO-PMs that are intended for CMS' VBP programs, APMs, or other regulatory and/or payment programs—the PHQ-9—as a use case to assess the Attribute Grid. If the Attribute Grid is designed appropriately, its use should confirm that the PHQ-9 is a high quality PROM for use in digital PRO-PMs.

TEP members conducted the assessment by comparing the PHQ-9 against the attributes contained within the Attribute Grid. Additionally, the TEP members considered whether the PHQ-9 triggered any additional attributes critical for defining a high quality PROM 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' Merit-Based Incentive Payment System [MIPS] #370 and the Health Resources and Services Administration [HRSA] Uniform Data System Quality of Care Measures), its 20+ year history of clinical use, its widespread adoption, the breadth of published research about it, and its use in multiple NQF-endorsed PRO-PMs.[23,24] 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.

Appendix D includes the use case of the PHQ-9 and Appendix E includes a use case of the Kansas City Cardiomyopathy Questionnaire (KCCQ). NQF does not currently endorse, recommend, rank, or prioritize PROMs. The inclusion of the PHQ-9 and the KCCQ within the use cases is not an endorsement or recommendation. These PROMs were included to verify the accuracy of the Attribute Grid and to demonstrate its use to readers.

The Environmental Scan Report identified 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.

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.

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 attributes 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.[25] 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 and the amount of evidence that exists on its use.

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.

Although the above limitations directly affected the Interim Report, there are numerous opportunities to advance future work related to digital PRO-PMs.

Next Steps

The recommendations contained within this Interim Report create the foundation for the Technical Guidance Report (to be published in fall 2021). 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 Interim Report describes these attributes, and the Technical Guidance Report will present a roadmap that guides measure developers through the stages of developing digital PRO-PMs for potential use in VBP programs or APMs. The attributes discussed in this Interim Report will be an important contribution to that roadmap.

Beyond the Building a Roadmap initiative, there are other important steps that stakeholders across the spectrum of healthcare should take to advance the use of PRO-PMs. Digital performance measurement is in its early stages and the use of PROs is not ubiquitous, so stakeholders can advance the field of digital PRO-PMs in many ways, including:

- Actively working to drive patient-centered efforts, such as CMS' Meaningful Measures 2.0;

- Advocating for, and implementing PROMs in clinical settings; and
- Exploring and acting upon opportunities to incentivize the use of PROMs.

As priorities in healthcare continue to evolve towards elevating patient voices, PRO-PMs will become increasingly important. The attributes discussed in this Interim Report are one small but valuable step forward in the ongoing journey toward high quality, patient-centered care.

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

3 National Quality Forum (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.

4 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.

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

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

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

8 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.

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 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.

11 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.

12 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.

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

14 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.

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

16 Health Level 7 International (HL7). Overview - FHIR v4.0.1. https://www.hl7.org/fhir/overview.html. Last accessed February 2021.

17 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.

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

19 NQF. NQF: Person Centered Planning and Practice Final Report. https://www.qualityforum.org/Publications/2020/07/Person_Centered_Planning_and_Practice_Final_Report.aspx. Published 2020. Last accessed July 2021.

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

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

22 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.

23 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.

24 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.

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

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

27 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.

28 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.

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

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.

31 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.

32 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.

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

34 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.

35 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.

36 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.

37 Logical Observation Identifiers Names and Codes (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.

38 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.

39 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.

40 Spertus JA, Jones PG, Sandhu AT, et al. Interpreting the Kansas City Cardiomyopathy Questionnaire in Clinical Trials and Clinical Care: JACC State-of-the-Art Review. *J Am Coll Cardiol*. 2020;76(20):2379-2390.

41 Ekman I, Chassany O, Komajda M, et al. Heart rate reduction with ivabradine and health related quality of life in patients with chronic heart failure: results from the SHIFT study. *Eur Heart J*. 2011;32(19):2395-2404.

42 Flynn KE, Piña IL, Whellan DJ, et al. Effects of exercise training on health status in patients with chronic heart failure: HF-ACTION randomized controlled trial. *JAMA*. 2009;301(14):1451-1459.

43 Jayaram NM, Khariton Y, Krumholz HM, et al. Impact of Telemonitoring on Health Status. *Circ Cardiovasc Qual Outcomes*. 2017;10(12):e004148.

44 Lewis EF, Claggett BL, McMurray JJV, et al. Health-Related Quality of Life Outcomes in PARADIGM-HF. *Circ Heart Fail*. 2017;10(8):e003430.

45 Lewis EF, Kim H-Y, Claggett B, et al. Impact of Spironolactone on Longitudinal Changes in Health-Related Quality of Life in the Treatment of Preserved Cardiac Function Heart Failure With an Aldosterone Antagonist Trial. *Circ Heart Fail*. 2016;9(3):e001937.

46 Nassif ME, Windsor SL, Tang F, et al. Dapagliflozin Effects on Biomarkers, Symptoms, and Functional Status in Patients With Heart Failure With Reduced Ejection Fraction. *Circulation*. 2019;140(18):1463-1476.

47 Kosiborod MN, Jhund PS, Docherty KF, et al. Effects of Dapagliflozin on Symptoms, Function, and Quality of Life in Patients With Heart Failure and Reduced Ejection Fraction: Results From the DAPA-HF Trial. *Circulation*. 2020;141(2):90-99.

48 Maurer MS, Schwartz JH, Gundapaneni B, et al. Tafamidis Treatment for Patients with Transthyretin Amyloid Cardiomyopathy. *New England Journal of Medicine*. 2018;379(11):1007-1016.

49 Arnold SV, Stone GW, Mack MJ, et al. Health Status Changes and Outcomes in Patients With Heart Failure and Mitral Regurgitation: COAPT Trial. *J Am Coll Cardiol*. 2020;75(17):2099-2106.

50 Arnold SV, Spertus JA, Vemulapalli S, et al. Quality-of-Life Outcomes After Transcatheter Aortic Valve Replacement in an Unselected Population: A Report From the STS/ACC Transcatheter Valve Therapy Registry. *JAMA Cardiol*. 2017;2(4):409-416.

51 Arnold SV, Reynolds MR, Wang K, et al. Health Status After Transcatheter or Surgical Aortic Valve Replacement in Patients With Severe Aortic Stenosis at Increased Surgical Risk: Results From the CoreValve US Pivotal Trial. *JACC Cardiovasc Interv*. 2015;8(9):1207-1217.

52 Reynolds MR, Magnuson EA, Wang K, et al. Health-related quality of life after transcatheter or surgical aortic valve replacement in high-risk patients with severe aortic stenosis: results from the PARTNER (Placement of AoRTic TraNscathetER Valve) Trial (Cohort A). *J Am Coll Cardiol*. 2012;60(6):548-558.

53 Joseph SM, Novak E, Arnold SV, et al. Comparable performance of the Kansas City Cardiomyopathy Questionnaire in patients with heart failure with preserved and reduced ejection fraction. *Circ Heart Fail*. 2013;6(6):1139-1146.

54 Dunlay SM, Gheorghiade M, Reid KJ, et al. Critical elements of clinical follow-up after hospital discharge for heart failure: insights from the EVEREST trial. *Eur J Heart Fail*. 2010;12(4):367-374.

55 Arnold SV, Spertus JA, Lei Y, et al. How to define a poor outcome after transcatheter aortic valve replacement: conceptual framework and empirical observations from the placement of aortic transcatheter valve (PARTNER) trial. *Circ Cardiovasc Qual Outcomes*. 2013;6(5):591-597.

56 Hejjaji V, Cohen DJ, Carroll JD, et al. Practical Application of Patient-Reported Health Status Measures for Transcatheter Valve Therapies. *Circulation: Cardiovascular Quality and Outcomes*. 2021;14(3):e007187.

57 Heidenreich PA, Spertus JA, Jones PG, et al. Health status identifies heart failure outpatients at risk for hospitalization or death. *J Am Coll Cardiol*. 2006;47(4):752-756.

58 Pokharel Y, Khariton Y, Tang Y, et al. Association of Serial Kansas City Cardiomyopathy Questionnaire Assessments With Death and Hospitalization in Patients With Heart Failure With Preserved and Reduced Ejection Fraction: A Secondary Analysis of 2 Randomized Clinical Trials. *JAMA Cardiol*. 2017;2(12):1315-1321.

59 Butler J, Khan MS, Mori C, et al. Minimal clinically important difference in quality of life scores for patients with heart failure and reduced ejection fraction. *Eur J Heart Fail*. 2020;22(6):999-1005.

60 Dreyer RP, Jones PG, Kutty S, et al. Quantifying clinical change: discrepancies between patients' and providers' perspectives. *Qual Life Res*. 2016;25(9):2213-2220.

61 Spertus J, Peterson E, Conard MW, et al. Monitoring clinical changes in patients with heart failure: a comparison of methods. *Am Heart J*. 2005;150(4):707-715.

62 Spertus JA, Jones PG. Development and Validation of a Short Version of the Kansas City Cardiomyopathy Questionnaire. *Circ Cardiovasc Qual Outcomes*. 2015;8(5):469-476.

63 Green CP, Porter CB, Bresnahan DR, et al. Development and evaluation of the Kansas City Cardiomyopathy Questionnaire: a new health status measure for heart failure. *J Am Coll Cardiol*. 2000;35(5):1245-1255.

64 Hejjaji V, Tang Y, Coles T, et al. Psychometric evaluation of the Kansas City Cardiomyopathy Questionnaire in men and women with heart failure. *Circ Heart Fail*. 2021.

65 Kosiborod M, Soto GE, Jones PG, et al. Identifying Heart Failure Patients at High Risk for Near-Term Cardiovascular Events With Serial Health Status Assessments. *Circulation*. 2007;115(15):1975-1981.

66 Soto GE, Jones P, Weintraub WS, et al. Prognostic Value of Health Status in Patients With Heart Failure After Acute Myocardial Infarction. *Circulation*. 2004;110(5):546-551.

67 Johansson I, Joseph P, Balasubramanian K, et al. Health-Related Quality of Life and Mortality in Heart Failure: The Global Congestive Heart Failure Study of 23 000 Patients From 40 Countries. *Circulation*. 2021;143(22):2129-2142.

68 Greene SJ, Butler J, Spertus JA, et al. Comparison of New York Heart Association Class and Patient-Reported Outcomes for Heart Failure With Reduced Ejection Fraction. *JAMA Cardiology*. 2021;6(5):522-531.

69 Mondesir FL, Zickmund SL, Yang S, et al. Patient Perspectives on the Completion and Use of Patient-Reported Outcome Surveys in Routine Clinical Care for Heart Failure. *Circulation: Cardiovascular Quality and Outcomes*. 2020;13(9):e007027.

70 Wohlfahrt P, Zickmund SL, Slager S, et al. Provider Perspectives on the Feasibility and Utility of Routine Patient-Reported Outcomes Assessment in Heart Failure: A Qualitative Analysis. *Journal of the American Heart Association*. 2020;9(2):e013047.

71 Carroll JD, Mack MJ, Vemulapalli S, et al. STS-ACC TVT Registry of Transcatheter Aortic Valve Replacement. *J Am Coll Cardiol*. 2020;76(21):2492-2516.

72 Spertus JA, Birmingham MC, Butler J, et al. Novel Trial Design: CHIEF-HF. *Circulation: Heart Failure*. 2021;14(3):e007767.

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.[26]

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.[20]

Attribute

A characteristic or trait of a PROM. Past National Quality Forum (NQF) reports have used *attribute* and *characteristic* synonymously.[3,6] In the Interim Report, *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.[3]

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.[7]

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.[10]

Crosswalk

A concordance table to convert scores from one scale to the other and vice versa.[14] 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.[14]

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.[10]

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.[27]

Electronic Clinical Quality Measures (eCQMs)

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

Interpretability

The degree to which the meaning of the scores can be easily understood by any group requiring use of the scores. A PRO measure should have documentation to support interpretation of scores, including the meaning of low and high scores and guidance on the minimally important difference in scores between groups and/or over time.[6]

Logical Observation Identifiers, Names, and Codes (LOINC)

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

Minimal Clinically Important Difference (MCID)

This is the smallest improvement needed after treatment that would be considered worthwhile from the patient's perspective.[20] MCID can be calculated using three different methods: consensus or delphi method, which depends on consensus of an expert panel; anchors (described above); and distribution-based method, which relies on the statistical analysis of the distribution of outcome scores.[20]

Patient-Reported Outcome (PRO)

Any information on the outcomes of healthcare obtained directly from patients without modification by clinicians or other healthcare professionals.[6]

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.[6]

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.[6]

Performance Measures (PMs)

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

Psychometric Soundness

How consistently and accurately an assessment measures what it purports to measure. Validity and reliability are key aspects to attaining psychometric soundness.[10] 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).[21]

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: PROM Attribute Grid From 2020 Report

The following Attribute Grid was published in the Final Technical Report for the Centers for Medicare & Medicaid Services (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 were 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)	*	*	*	*

PROM	PROM 1	PROM 2	PROM 3	PROM 4
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 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	*	*	*	*
Interpretable scores, defined and actionable cut points or targets, and anchors and/or defined meaningful change	*	*	*	*
Clear conceptual and measurement models	*	*	*	*
Psychometric Soundness: Reliability	*	*	*	*
Psychometric Soundness: Validity	*	*	*	*
Psychometric Soundness: Responsiveness	*	*	*	*
Usability/Feasibility of Use: Low burden (e.g., length, time/effort to complete) and feasibility	*	*	*	*
Usability/Feasibility of Use: Fits with standard of care and related workflows (e.g., actionable, incorporated, and discussed at point of care)	*	*	*	*
Usability/Feasibility of Use: • Cultural appropriateness • Language • Translated with culturally appropriate items	*	*	*	*
Usability/Feasibility of Use: Availability of standardized clinical terminology and codes	*	*	*	*
Usability/Feasibility of Use: Guidance on standardized data collection (including modes and methods)	*	*	*	*

Appendix D: Patient Health Questionnaire-9 (PHQ-9) Use Case

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

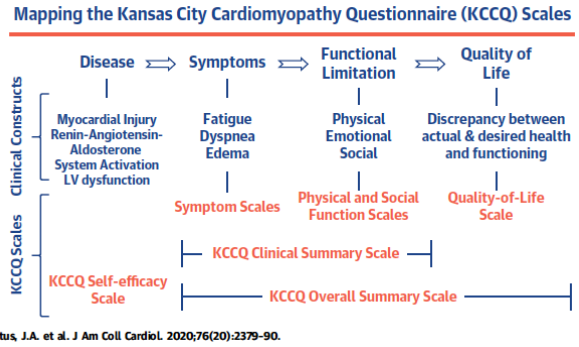
ATTRIBUTE	PROM 1: PHQ 9
Covers desired PROs from patient and/or caregiver perspective	The PHQ-9 tool is validated for use as a measure to assess the level of depression severity (for initial treatment decisions) as well as an outcome tool (to determine treatment response).[5]
Outcome measured in PROM 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.
Interpretable scores, defined and actionable cut points or targets, and anchors and/or defined meaningful change	Guide for Interpreting PHQ-9 Scores[31] • 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)[32] Specificity: 88% (PHQ-9 score >= 10)[32] Positive Predictive Value: 31% (cut point = 9) to 51% (cut point = 15)[32]
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[32] 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).[32]
Psychometric Soundness: Validity	Criterion validity: demonstrated in 580 primary care patients who underwent an independent reinterview by a mental health professional[32] Construct validity: established by strong association between PHQ-9 scores and functional status, disability days, and symptom-related difficulty[32] External validity: replicated findings from 3,000 primary care patients in a sample of 3,000 OB/GYN patients[32] 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.[32]

ATTRIBUTE	PROM 1: PHQ 9
	half the days"
Usability/Feasibility of Use: Low burden (e.g., length, time/effort to complete) and feasibility	Examples of relevant attributes include the following: • Average time to complete: 3.6 minutes[33] • Number of questions: 9 • Administration: self-administered[32] • Scoring: less than one minute (time to add nine single-digit numbers)[32]
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	Official website offers translation into more than 90 languages but notes that few translations have been validated with an independent structured psychiatric interview.[25,31] 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[34], Turkish immigrants and Germans[35], and Romanians with type 2 diabetes mellitus.[36]
Usability/Feasibility of Use: Availability of standardized clinical terminology and codes	The LOINC Panel Browser offers standard codes for PHQ-9 (LOINC 44249-1) as well as other versions.[37]
Usability/Feasibility of Use: Guidance on standardized data collection (including modes and methods)	Numerous studies exist on modes and methods within specific populations, such as use of touch screens for patients with cancer[38] and telephonic administration in primary care settings.[39]

Appendix E. Kansas City Cardiomyopathy Questionnaire (KCCQ) Use Case

The following use case provides a second example of how the Attribute Grid can be used to assess a single PROM, the KCCQ.

ATTRIBUTE	PROM 1: KCCQ
Covers desired PROs from patient and/or caregiver perspective	The KCCQ has been demonstrated through numerous concept elicitation and cognitive debriefing studies to capture the key manifestations of heart failure that are important to patients; a foundation for FDA qualification. From a caregiver's perspective, extensive efforts have formed clinically-interpretable frameworks for interpreting cross-sectional and changes in scores over time.[40]
Outcome measured in PROM is result of care for which relevant clinical quality is being measured	Numerous clinical trials have used the KCCQ and demonstrated that treatments for heart failure result in improvements in patients' health, as measured by the KCCQ. These have included studies of drugs, devices and population-health management interventions.[41–52]
Interpretable scores, defined and actionable cut points or targets, and anchors and/or defined meaningful change	Extensive work has been done to define clinical interpretation for the KCCQ.[40] This has been done for cross-sectional scores in terms of risks for subsequent clinical events in multiple clinical settings (acute decompensated heart failure,[53] 1-week after discharge,[54] prior to and after valvular interventions[55,56] and in the outpatient setting[57]), the prognostic significance of serial KCCQ assessments and which is most significant,[58] and how to interpret changes in KCCQ scores for both population means and individual patients.[59–61] Cross-sectional scores: 0-24: Very poor to poor health status 25-49: Poor to fair health status 50-74: Fair to good health status 75-100: Good to excellent health status Changes in scores: ≤-20: A very large clinical deterioration ≤-10 to >-20: A moderate to large clinical deterioration ≤-5 to >-10: A small, but clinically important deterioration <5 to >-5: Clinically stable ≥5 to <10: A small, but clinically important improvement ≥10 to <20: A moderate to large clinical improvement ≥20: A very large clinical improvement

<table>
<tr><th>ATTRIBUTE</th><th>PROM 1: KCCQ</th></tr>
<tr><td>Clear conceptual and measurement models</td><td>The different domains of the KCCQ conform to the key domains by which HF manifests itself to patients:
 

Spertus, J.A. et al. J Am Coll Cardiol. 2020;76(20):2379–90.</td></tr>
<tr><td>Psychometric Soundness: Reliability</td><td>The internal reliability is high from both patients with HFrEF and HFpEF,[53] as shown below:
<table>
<tr><th></th><th colspan="2">Cronbach's α</th></tr>
<tr><th></th><th>HFrEF (EF≤40)*</th><th>HFpEF (EF≥50)*</th></tr>
<tr><td>Overall summary</td><td>0.94</td><td>0.96</td></tr>
<tr><td>Clinical summary</td><td>0.92</td><td>0.93</td></tr>
<tr><td>Physical limitation</td><td>0.87</td><td>0.88</td></tr>
<tr><td>Symptoms</td><td>0.89</td><td>0.91</td></tr>
</table>
The test-retest reliability is also very high when assessed over 6 weeks in stable outpatients:[62]
<table>
<tr><th></th><th colspan="2">KCCQ-12</th><th colspan="2">Full KCCQ</th></tr>
<tr><th></th><th>Mean±SD</th><th>Intra-Class Correlation</th><th>Mean±SD</th><th>Intra-Class Correlation</th></tr>
<tr><td>Physical limitation</td><td>2.8±14.4</td><td>0.85</td><td>2.4±13.5</td><td>0.86</td></tr>
<tr><td>Symptom frequency</td><td>1.0±14.5</td><td>0.83</td><td>1.0±14.5</td><td>0.83</td></tr>
<tr><td>Quality of life</td><td>3.4±19.8</td><td>0.76</td><td>2.4±16.2</td><td>0.82</td></tr>
<tr><td>Social limitation</td><td>3.8±16.0</td><td>0.86</td><td>4.0±16.4</td><td>0.85</td></tr>
<tr><td>Summary score</td><td>2.6±10.5</td><td>0.91</td><td>2.3±9.7</td><td>0.92</td></tr>
</table></td></tr>
<tr><td>Psychometric Soundness: Validity</td><td>Content validity has been established through >4 independent cognitive debriefing studies, as required by FDA criteria for qualifying a Clinical Outcome Assessment.
Construct validity has been established by comparing each domain of the KCCQ with other measures of those concepts.[63,64]
Predictive validity has been demonstrated in multiple studies in multiple clinical settings,[53–58,65,66] including an international study or 23,000 patients from 40 countries.[67]</td></tr>
</table>

ATTRIBUTE	PROM 1: KCCQ
Psychometric Soundness: Responsiveness	The KCCQ has been very responsive to interventions for HF, including medical therapy, telephonic disease management, surgical, percutaneous and electrophysiological interventions, exercise therapy, palliative care, etc. As described in the section above on meaningfulness of scores/changes in scores, there is very strong data supporting the responsiveness and interpretability of the KCCQ and it has recently been shown to be much more sensitive and prognostically important than clinician-assigned New York Heart Associations.[68]
Usability/Feasibility of Use: Low burden (e.g., length, time/effort to complete) and feasibility	The 12-item KCCQ takes 2-3 minutes to complete and has been readily adapted to support collection through applications and patient portals in the EMR. A recent qualitative study of patients completing the KCCQ as part of routine clinical care demonstrated (96%) felt that the use of the KCCQ made their evaluation more accurate, but only 15 (62%) felt that their providers effectively discussed their KCCQ results with them, suggesting a need for providers to become more familiar with the use and interpretation of the KCCQ and a desire by patients to share this information with their providers.[69]
Usability/Feasibility of Use: Fits with standard of care and related workflows (e.g., actionable, incorporated and discussed at point of care)	The KCCQ has, and is, being implemented in multiple clinics[69,70] and procedures, including pre-procedure and follow-up of the American College of Cardiology/Society of Thoracic Surgery Trans-Valvular Therapeutics registry.[71] The actionability of scores directly emanates from the interpretability of scores/changes in scores, the conceptual domains measured and the evaluation of an individual patient to define what interventions might best influence the scores that are most impaired in that individual. By improving patients' KCCQ scores, not only is there health status improved, but their risk for mortality and hospitalizations also decreases.[58]
Usability/Feasibility of Use: • Cultural appropriateness • Language • Translated with culturally appropriate items	The KCCQ has been translated into >100 languages and cultures. Each translation has undergone extensive iterative processes to ensure a valid translation, including 2 independent forward translations, backward translation and reconciliation, feedback from the developer and pilot testing with patients and clinicians in that language/culture.
Usability/Feasibility of Use: Availability of standardized clinical terminology and codes	LOINC codes are available, as are CPT3 codes.

ATTRIBUTE	PROM 1: KCCQ
Usability/Feasibility of Use: Guidance on standardized data collection (including modes and methods)	Standard guidelines for the acquisition of the KCCQ are available and multiple modes have been used to collect it in clinical care and in trials. In fact, a recent clinical trial was conducted during the COVID-19 pandemic without any in-person visits.[72]

Appendix F: 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, Center for Clinical Standards and Quality

Janis Grady, RHIT, FAC-COR III
Centers for Medicare & Medicaid Services, 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
U.S. 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

Alejandra Herr, MPH
Senior Managing Director

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 G: Public Comments and TEP Response

The draft Interim Report was posted on the project webpage for public and National Quality Forum (NQF) member comment from June 1, 2021, through June 21, 2021. Six prompts were offered to guide public commenters on key areas of interest. The comments below are grouped by prompt, and the TEP's response is listed immediately beneath each comment. During the commenting period, NQF received 15 total comments from eight organizations. Comments were elicited through various avenues, including the public commenting tool and additional organizational outreach. Unless otherwise noted, public comments are presented as they were received by NQF and have not been edited, with the exception of correcting minor spacing, spelling, and punctuation issues.

What general comments do you have on the report?

The Council on Quality and Leadership

COMMENT

We would like to commend NQF on their efforts to recognize patient-reported outcome performance measures. We represent the Council on Quality and Leadership (CQL), an international not-for-profit organization dedicated to the definition, measurement, and improvement of personal quality of life. For over 50 years, CQL has taken the leadership initiative in developing progressive measures and indicators of quality in services and supports, personal quality of life outcome measures, and social capital. CQL emerged from national efforts to improve conditions in public institutions in the late 1960s. CQL established the first and subsequent sets of standards and performance indicators for children and adults that were later adopted as Federal Standards by HCFA (now CMS). CQL's standards were incorporated in the historic Wyatt v. Stickney decision, and the federal courts later incorporated them into legal settlements in TX, ND, CA, AK, and WV. CQL has collaborated with state systems for over five decades to design and deliver statewide quality management models that are person-centered, integrated, and evidence based. CQL models have been shown to empower states to build capacity and utilize data to hold systems accountable while targeting limited resources. Based on our experience, background, and research, we would like to draw attention to the following important aspects of quality for your consideration as you build a roadmap for PRO-PMs.

A quality service system necessitates a truly person-centered outcome performance measurement process. The person-centered measures should be designed to assist the person to achieve personally defined outcomes, while also contributing to the assurance of their health and welfare. To truly be person-centered, outcome measurement must reflect people's meaningful priorities/goals, reflect their desires, and include informed choice.

It is important that quality metrics are not utilized simply because they are easy to measure. In addition, whenever possible, it is important to avoid utilizing process-based assessments and metrics. We believe that quality assurance metrics should incorporate metrics that capture the extent to which person-centered outcomes are infused throughout the provision of clinical services. Person-centered outcomes not only recognize each person's unique individuality, systems and organizations promote health when they incorporate the person's wants and needs in the provision of services and supports.

We wish you much success as build a roadmap to patient-report outcome performance measures. Please do not hesitate to reach out if you have any questions or would like further clarification.

Respectfully,

Mary Kay Rizzolo, PhD
President & CEO
CQL | The Council on Quality and Leadership

RESPONSE

Thank you for your comment. We agree that person-centered measures should incorporate people's priorities, goals, desires, and informed choices. We have added relevant language to the "Covers Desired PROs From Patient and/or Caregiver Perspective" portion of the Attributes and Definitions section of the paper. We have also made edits throughout the paper to emphasize the importance of person-

centeredness to the patient-reported outcome process, and we will ensure this lens remains in place as we develop the Technical Guidance Report later in 2021.

American Association on Health and Disability

COMMENT

The American Association on Health and Disability and the Lakeshore Foundation appreciate the opportunity to provide comments. We write to endorse the June 3 submitted comments of the Council on Quality and Leadership (CQL). In addition to the CQL submitted comments, we also strongly encourage the NQF to fully integrate and reference in some detail two NQF previously published reports – Person-Centeredness (July 2020) and Home and Community-Based Services (September 2016).

The June draft PRO-PMs report, page 5, raises the question: "Whether a PROM is well suited for performance measurement"? The CQL comments answer yes and explain the importance to beneficiaries and recipients of services and supports (including patients of health services programs).

The American Association on Health and Disability (AAHD) (www.aahd.us) is a national, nonprofit organization of public health professionals, both practitioners and academics, with a primary concern for persons with disabilities. The AAHD mission is to advance health promotion and wellness initiatives for persons with disabilities. AAHD is specifically dedicated to integrating public health and disability into the overall public health agenda.

The Lakeshore Foundation (www.lakeshore.org) mission is to enable people with physical disability and chronic health conditions to lead healthy, active, and independent lifestyles through physical activity, sport, recreation and research. Lakeshore is a U.S. Olympic and Paralympic Training Site; the UAB/Lakeshore Research Collaborative is a world-class research program in physical activity, health promotion, and disability linking Lakeshore's programs with the University of Alabama, Birmingham's research expertise.

AAHD & Lakeshore Foundation endorse June 3 Council on Quality and Leadership (CQL) submitted comments.

RESPONSE

Thank you for your support of the comment submitted by the Council on Quality and Leadership (CQL) and your support of PROMs in the use of PRO-PMs. Please see our response to CQL's comment.

American Association on Health and Disability

COMMENT

AAHD and Lakeshore Foundation endorse the CQL observations and sharing:

A quality service system necessitates a truly person-centered outcome performance measurement process. The person-centered measures should be designed to assist the person to achieve personally defined outcomes, while also contributing to the assurance of their health and welfare. To truly be person-centered, outcome measurement must reflect people's meaningful priorities/goals, reflect their desires, and include informed choice.

It is important that quality metrics are not utilized simply because they are easy to measure. In addition, whenever possible, it is important to avoid utilizing process-based assessments and metrics. We believe

that quality assurance metrics should incorporate metrics that capture the extent to which person-centered outcomes are infused throughout the provision of clinical services. Person-centered outcomes not only recognize each person's unique individuality; systems and organizations promote health when they incorporate the person's wants and needs in the provision of services and supports.

RESPONSE

Thank you for your support of the comment submitted by the Council on Quality and Leadership (CQL). Please see our response to CQL's comment.

American College of Physicians

COMMENT

The American College of Physicians (ACP) appreciates the opportunity to comment on NQF's Building a Roadmap From Patient-Reported Outcome Measures to Patient-Reported Outcome-based Performance Measures - Interim Draft Report. ACP recognizes that the development of sound patient-reported outcome-based performance measures (PROM-PMs) is a necessary step towards achieving patient-centered care, and we applaud the work of the committee in getting us closer to that goal. We have some comments and concerns for the committee's consideration.

1. We value the committee's efforts to identify the desirable attributes for high quality patient-reported outcome measures (PROMs) that could be used in developing patient-reported outcome–based performance measures (PRO-PMs). We believe that this attribute list will help measure developers immensely in identifying and selecting the appropriate PROM for a PRO-PM. While we can appreciate the committee's perspective to make this list non-prescriptive and thus not make these attributes a requirement, all of these attributes are extremely important. As a result, this list of attributes, at a minimum, should be vetted for every PROM that is being considered for the development of a PRO-PM. Certainly, developers can add additional attributes that they believe are important.

2. The ACP understands the priority to move to digital measures and the call for interoperability. The committee's recommendation to use computer-assisted technology (CAT) to aid in that and to help to streamline unnecessary questions is correct. However, not all practices are ready for it. Thus, there is a critical need to have PRO-PMs that can accept scores from multiple PROMs with validated crosswalks. It is critical for a physician/practice to select the PROM that is most meaningful for that patient population and most appropriate for their setting. Understanding the unique challenges of certain care settings, and validating a diversity-focused, patient-centered assessment of the available measures will be key to success of PRO-PM.

3. As we highlighted in ACP's Recommending Caution in Patient-Reported Outcome–Based Performance Measurement paper (Ann Intern Med. doi:10.7326/M19-3603), there are several important considerations for PRO-PMs. Many of these also affect the PROM that make up the foundation of the PRO-PM. Also, as with any other outcome measure, the attribution of the PROM needs to be evidence-based. The developers should validate that the PROM targets the intended patient population and that the outcome measured can be impacted by the attributed entity. The ACP also strongly believes that, while evaluating PROMs, the developers need to consider any risk adjustment that may be needed to be incorporated later in the PRO-PM. Additionally, the seamless collection of PROM data is not yet part of the routine practice of medical care in the United States. This relates to many different factors including the burden associated with setting up systems to collect the data, the lack of engagement of patients in

reporting outcomes, and limitations on clinical utility or ability to act on results, among others. PRO-PMs represent only a fraction of the measures used in federal reporting programs and endorsed by the NQF. In 2021 CMS' QPP program, out of 209 Clinical Quality Measures listed on the QPP site, only 17 were PRO-PMs. These data further underscore the practical challenges that practices face in incorporating a PROM meaningfully in routine clinical care.

RESPONSE

Thank you for your comment. We agree that all of the attributes are extremely important. Given the subjective nature of some of the attributes (as we discuss in the "Objectivity and Subjectivity" subsection of the report), however, it is not feasible at this time to mandate use of the Attribute Grid in development or endorsement of PRO-PMs. The Technical Expert Panel has strengthened the language in the "Flexibility of the Attribute Grid" section report, though, to strongly encourage the use of the Attribute Grid. Additionally, NQF has shared your suggestion with the leadership of the Consensus Development Process. Regarding your second comment, while we do acknowledge the advantages of both CAT and crosswalks, we recognize that different clinical settings and patient populations have different needs, and it is important to use PROMs that best fit those needs. The "Relationship Between PROMs and PRO-PMs" subsection of the report discusses the TEP's recommendation that PROM-based performance measures be specified to include a choice of different validated PROMs that are relevant to the measure, and we invite you to review the 2020 NQF report titled *Patient-Reported Outcomes: Best Practices on Selection and Data Collection* for best and promising practices on the selection and implementation of PROs and PROMs.

We appreciate your third point and thank you for your reference to the article, which includes our Co-Chair, Dr. Catherine MacLean, as an author. Your comment aligns well with CMS' aim to increase the development of PRO-PMs and prepare developers for the NQF endorsement process. The section titled "Brief NQF History With PROs" discusses why NQF does not endorse PROMs and also offers a high-level explanation of how NQF's Scientific Methods Panel reviews PROMs that are used to collect data for PRO-PMs that are considered for endorsement.

Partners Health Management

COMMENT

Given the complexity of this this topic and to help assure the reader has a good grasp of the differences between PROs, PROMs, and PRO-PMs early on as they begin reading the document, it would be helpful to move the "Terminology" section that starts on page 4 to much earlier in the document. This section touches on the glossary and guides the reader to the glossary, provides a valuable table (Table 1) with a user-friendly distinction and example of the differences; however, the reader isn't exposed to this until page 4. In my opinion, it would help the audience to have this section much sooner in the document.

Thank you to the Committee members and NQF for the work put into creating this draft and guiding the future work ahead!

Selenna Moss, MHA, BS, RHIT, CHC, CHP

Partners Health Management

RESPONSE

Thank you for your comment. We have added a "reader's note" at the begin of the document with links to the "Terminology" section and the Glossary.

American Geriatrics Society

COMMENT

For meaningful use of PRO-PMs, there should be greater effort at patient engagement in the completion of PROMs.

RESPONSE

Thank you for this comment. This is a well-documented challenge and we fully agree with your statement. We invite you to review the 2020 NQF report titled *Patient-Reported Outcomes: Best Practices on Selection and Data Collection*, which discusses some best and promising practices around PROs and PROMs, including engagement of patients and caregivers in the completion of PROMs.

American Medical Association

COMMENT

The American Medical Association (AMA) appreciates the opportunity to comment on this draft report. While we make every attempt to provide substantive comments and recommend improvements to the content and recommendations, we were unable to do that this time.

On review of the report, we found that its purpose, definitions, and information on the attributes of patient-reported outcome measures (PROMs) were not clearly described. For example, the definition of a high quality PROM lacks sufficient detail to enable any measure developer to distinguish a good versus poor PROM and on page 9 of the report, a reference is made to criteria and attributes but no further mention of criteria can be found and only attributes are discussed. Perhaps more importantly when we compare the attributes list to PROM characteristics discussed in the NQF Patient-Reported Outcomes in Performance Measurement report released in 2013, they are very duplicative and we believe that the previous report provides more substantive and useful guidance than what is included in this current draft report.

The AMA encourages NQF to consider whether this report contributes any new information to the field and if not, perhaps reaffirm the report from 2013. In addition, since the next phase in this project is to produce technical guidance for developers as they develop patient-reported outcome performance measures, we believe that that report and any supplemental materials must be sufficiently detailed and clear to further advance measurement in this important area.

RESPONSE

Thank you for your comment. We have clarified language around criteria and attributes, and we appreciate you calling these to our attention. The 2013 NQF report is foundational to the current work. Advancements have occurred in PROMs and PRO-PMs since 2013 and this report provides an opportunity to review, confirm, and expand upon the findings from 2013. While we recognize and applaud the similarities between the two reports and are pleased by the ongoing relevance of the 2013 report, there are new findings and recommendations in this Interim Report and the upcoming Technical Guidance Report that will serve measure developers who are selecting PROMs for use in PRO-PMs. We have added additional language in the section titled "Brief NQF History with PROs" and we have added a

new section titled "Commonalities and Differences Between the Interim Report and Past Reports" that describes the similarities and differences between this report and the 2013 report.

In the section titled "Attributes of High Quality PROMs for Use in Performance Measures" (pages 9-12), which, if any, attributes are missing or need updated information?

American Association on Health and Disability

COMMENT

AAHD & Lakeshore Foundation also strongly encourage the NQF to fully integrate and reference in some detail two NQF previously published reports – Person-Centeredness (July 2020) and Home and Community-Based Services (September 2016).

NQF projects and reports typically are silo focused, reflecting the expertise of the particular technical expert panel and committee engaged on a project. NQF excels at multi-stakeholder engagement and analysis and protype public transparency and engagement; however, many products still reflect a particular silo. We strongly encourage the NQF PRO-PMs report to fully integrate and reference in some detail two previously published reports; these reinforce and are totally consistent with the CQL comments. The two reports:

"Person-Centered Planning and Practice" (July 31, 2020)

"Quality in Home-and-Community-Based Services To Support Community Living: Addressing Gaps in Performance Measurement" (September 2016)

RESPONSE

Thank you for your comment. We are aware of the two reports you reference and aim to build on the work in those reports. These reports, contain relevant findings that are applicable to both the Interim Report and the Technical Guidance Report that will be published in late 2021 as part of this project. We will use these findings to strengthen the person-centeredness of both "Building a Roadmap" reports.

American Medical Association

COMMENT

The definition of a high quality PROM lacks sufficient detail to enable any measure developer to distinguish a good versus poor PROM and on page 9 of the report, a reference is made to criteria and attributes but no further mention of criteria can be found and only attributes are discussed. Perhaps more importantly when we compare the attributes list to PROM characteristics discussed in the NQF Patient-Reported Outcomes in Performance Measurement report released in 2013, they are very duplicative and we believe that the previous report provides more substantive and useful guidance than what is included in this current draft report.

RESPONSE

Thank you for your comment. We have clarified language around criteria and attributes, and we appreciate you calling these to our attention. The 2013 NQF report is foundational to the current work. Advancements have occurred in PROMs and PRO-PMs since 2013 and this report provides an opportunity to review, confirm, and expand upon the findings from 2013. While we recognize and applaud the similarities between the two reports and are pleased by the ongoing relevance of the 2013 report, there are new findings and recommendations in this Interim Report and the upcoming Technical

Guidance Report that will serve measure developers who are selecting PROMs for use in PRO-PMs. We have added additional language in the section titled "Brief NQF History with PROs", and we have added a new section titled "Commonalities and Differences Between the Interim Report and Past Reports " that describes the similarities and differences between this report and the 2013 report.

American Geriatrics Society

COMMENT

The assessment of whether a particular measure meets the criterion that the PRO-PM "covers desired PROs from patients and/or caregiver perspectives" is missing. The example case study of PHQ-9 is a patient-reported outcome, but it does not necessarily prove that depression is important to patients/caregivers. We can assume that it does, which is reasonable, but that is not actually directly shown, and we can see how this may become a slippery slope in other settings. For example, we may have a shared decision-making scale that is validated and assume that patients/caregivers would want more shared decision-making, but literature suggests that some older adults may prefer a more passive role in decisions. We would like to see that attribute – how would one know if the attribute is adequately met – more clearly stated considering what type of evidence is needed, how strong does this level of evidence need to be, and how we can ensure that the evidence includes diverse patients/caregivers in terms of background/race/ethnicity/culture.

We are also concerned that clinically meaningful improvement in PRO-PMs does not necessarily have face validity to providers at ground level which is why the measurement group is perhaps the most important decision. For non-patient-reported outcome measures, the evidence base generally points to thresholds and targets that identify ideal situations. However, when these outcome measures are made into performance measures, they can highlight systems/providers caring for more challenging patients as underperforming. For example, A1C <7% is an outcome measure with the greatest evidence, which when translated to a performance measure was problematic. Outcome measures usually identify idealized thresholds ("the ceiling") while performance measures are best at identifying "the floor." There should be critical examination before using the same thresholds for both.

RESPONSE

Thank you for your comment. We agree that the first attribute, "Covers desired PROs from patient and/or caregiver perspective" is challenging to objectively prove. Our intent with the attributes and the Attribute Grid is to not be prescriptive but rather to guide measure developers in critically evaluating whether a PROM is high quality for use in a PRO-PM. The section titled "Additional Considerations About the Attributes" contains a subsection that discusses the objectivity or subjectivity of different attributes. We have refined the language in that subsection based on your comment. To your second point, we recognize the difficulty of finding the right outcome measure for the PRO-PM. We realize this initiative will not completely solve that problem, but the attributes discussed in this report and the roadmap that will be offered in the upcoming Technical Guidance Report will offer tools to measure developers to assist with navigating these challenges.

What information should be edited or added to help measure developers understand what defines a high quality PROM for use in performance measures?

American Medical Association

COMMENT

The American Medical Association (AMA) appreciates the opportunity to comment on this draft report. While we make every attempt to provide substantive comments and recommend improvements to the content and recommendations, we were unable to do that this time.

On review of the report, we found that its purpose, definitions, and information on the attributes of patient-reported outcome measures (PROMs) were not clearly described. For example, the definition of a high quality PROM lacks sufficient detail to enable any measure developer to distinguish a good versus poor PROM and on page 9 of the report, a reference is made to criteria and attributes but no further mention of criteria can be found and only attributes are discussed. Perhaps more importantly when we compare the attributes list to PROM characteristics discussed in the NQF Patient-Reported Outcomes in Performance Measurement report released in 2013, they are very duplicative and we believe that the previous report provides more substantive and useful guidance than what is included in this current draft report.

The AMA encourages NQF to consider whether this report contributes any new information to the field and if not, perhaps reaffirm the report from 2013. In addition, since the next phase in this project is to produce technical guidance for developers as they develop patient-reported outcome performance measures, we believe that that report and any supplemental materials must be sufficiently detailed and clear to further advance measurement in this important area.

RESPONSE

Thank you for your comment. We have clarified language around criteria and attributes, and we appreciate you calling these to our attention. The 2013 NQF report is foundational to the current work. Advancements have occurred in PROMs and PRO-PMs since 2013, and this report provides an opportunity to review, confirm, and expand upon the findings from 2013. While we recognize and applaud the similarities between the two reports and are pleased by the ongoing relevance of the 2013 report, there are new findings and recommendations in this Interim Report and the upcoming Technical Guidance Report that will serve measure developers who are selecting PROMs for use in PRO-PMs. We have added additional language in the section titled "Brief NQF History With PROs", and we have added a new section titled "Commonalities and Differences Between the Interim Report and Past Reports" that describes the similarities and differences between this report and the 2013 report.

Received via Direct Outreach

Human Services Research Institute

COMMENT

Thank you for the opportunity to provide comment as this important work is developed.

The Human Services Research Institute (HSRI) is a nonprofit dedicated to the advancement of person-centered, outcomes-oriented service models. For 45 years, HSRI has led targeted evaluation efforts that have amplified the voices and priorities of people with disabilities in driving systems improvements. We focus on collaborative and inclusive research, integrated services, and continuous quality improvement

across the human services sectors, including those that address the needs of people with intellectual and developmental disabilities, people experiencing behavioral health disorders, children youth and families, seniors and people with physical disabilities, and people experiencing housing instability or homelessness. We also assist states and communities in efforts to promote population health.

The National Core Indicators (NCI), developed as a partnership between HSRI and the National Association of State Directors of Developmental Disabilities Services (NASDDDS), is a validated survey tool and protocol that allows for state-level monitoring of quality person-centered outcomes. NCI began in 1997 and is currently in use in 47 states and the District of Columbia. Through state-level and national reports, states participating in NCI benchmark outcome domains for people with intellectual and developmental disabilities. All of the reports are made publicly available at www.nationalcoreindicators.org. In 2014, the National Core Indicators for Aging and Disability (NCI-AD) survey tool was added, in collaboration with ADvancing States. (See www.nci-ad.org for public reports focused on the population with age-related support needs and physical disabilities.) These tools have been developed and validated. At present, a set of NCI measures are under consideration for endorsement by NQF.

We have worked with states and sub-state entities to collect and use quality of life and quality outcomes data to enhance supports and services. Based on this experience, we offer these comments:

Community-based supports, such as those provided from Home and Community-Based Services (HCBS) waivers are broadly defined and should be tailored to be person-centered, taking into account the person's individual support needs and circumstances. Quality person-centered community-based supports are based in a person's true community, and may include multiple paid and unpaid supports over different ranges of time for the person to be able to live a good life of their choosing.

There are several key life domains included in NCI that parallel many of the quality domains outlined in NQF's Quality in HCBS to Support Community Living: Addressing Gaps in Performance Measurement (September 2016). This report, as well as the priorities outlined in the NQF Person Centered Planning and Practice Final Report (2018) point to important quality elements for person centered practices, including skills required for supporters and facilitators of person-centered practice.

National Core Indicators surveys yield person-reported outcomes, providing a statewide perspective on people's experiences and outcomes. States have used results of the surveys to create targeted quality improvement initiatives in areas such as medication review, employment supports, and supports for developing and maintain relationships. Each of these areas are important factors in supporting good health and quality outcomes; however, the attribution or direct cause of the outcome is difficult to establish in a community-based setting. As the NQF report Improving Attribution Models (2018) pointed out, "Understanding who is responsible is essential to driving improvements in care as well as for securing long-term buy in from providers and facilitating the ability of value-based purchasing and alternative payment models to influence provider behavior. However, accurate attribution presents a significant challenge when a patient sees numerous providers in multiple settings for several conditions" (p. 3).

Another project at HSRI, the National Center on Advancing Person Centered Practices and Systems (NCAPPS) has begun to develop core competencies for person centered practice facilitation to enhance

person-reported outcomes. These have been identified in the report Five Competency Domains for Staff who Facilitate Person Centered Planning (NCAPPS, 2020), however consistent measures of competency, and attribution to outcomes is still in development.

On the question of whether a person-centered outcome (PRO) is well suited for performance measurement (PRO-PM), we advise caution for the following reasons.

1. Priority measures include those such as relationships and employment; however, the attribution of responsibility to any one facilitator, agency, or process (and therefore target for PRO-PM) in a community-based setting becomes increasingly difficult the more the person is fully engaged in their community.

2. Person-reported outcome measures can require significant effort to collect and report in a valid manner, particularly across population types. The emphasis on electronic measures demonstrates a priority on capturing most available data, perhaps at the expense of most important data for quality monitoring. Systems must anticipate that survey data, or data collection other than readily available e-measures will be essential to capture in an ongoing manner in order to effectively align with true quality.

3. There is a need to continue to test the ability of PROs to accurately reflect person-centered priorities for people from communities typically underrepresented including people of color, LGBTQ people, and for people with lived experience and non-English speakers. Until this work is further developed, any resulting performance measures may not fully represent the population eligible and in receipt of supports.

We applaud NQF for pursuing this important work to drive quality improvement using the voice of people who are receiving supports and are hopeful that this work will serve as another step toward consistent, equitable, and effective supports.

RESPONSE

Thank you for your comment. The "Building a Roadmap" Interim Report is not focused on any specific setting or discipline and is intended to discuss PROMs and PRO-PMs at an agnostic level. We agree that attribution is an important element of performance measurement. We have added language that explains this report uses the word "attributes" not in the context of attribution but rather to describe the attributes or characteristics of a PROM that is suitable for use in a PRO-PM. The upcoming Technical Guidance Report will discuss attribution at a high level, but attribution is largely out-of-scope for this project. We recommend that you follow the work of the ongoing NQF project, "Attribution for Critical Illness and Injury," as it might contain information you find useful. We agree with the challenges you raise around person-centeredness and health equity. The attributes "Cultural Appropriateness, Language, and Translations With Culturally Appropriate Items" and "Guidance on Standardized Data Collection (Including Modes and Methods)" attempt to look at issues around language and digital divide, and we will explore these issues in the upcoming Technical Guidance Report. We have revised language in the report to better reflect patient-centeredness. You also might be interested in reading NQF's 2020 report *Patient-Reported Outcomes: Best Practices on Selection and Data Collection*.

National Association of State Directors of Developmental Disabilities Services

COMMENT

NASDDDS would like to take this opportunity to thank the National Quality Forum on the opportunity to provide comment on the work to recognize patient-reported outcome performance measures.

The National Association of State Directors of Developmental Disabilities Services (NASDDDS) is a national nonprofit representing the nation's agencies in 50 states and the District of Columbia providing services to children and adults with intellectual and developmental disabilities and their families. NASDDDS promotes visionary leadership, systems innovation, and the development of national policies that support home and community-based services (HCBS) for individuals with disabilities and their families. All of the work of NASDDDS is built on a foundational goal of supporting member states to provide high quality person-centered services in which people with IDD are supported identify and live the lives they desire with self-determination, dignity, and caring relationships.

The National Core Indicators (NCI), developed as a partnership between NASDDDS and Human Services Research Institute (HSRI), is a validated tool and protocol that allows for state-level data collection and reporting of quality person-centered outcomes. NCI began in 1997 and is currently in use in 48 states including 21 regional entities and the District of Columbia. Through state level and national reporting and data analysis, NCI member states have the ability to benchmark outcome measures performance, compare quality performance to other states and national averages, identify quality improvement projects and assure compliance with HCBS program requirements for services provided to people with IDD. All of the NCI reports are available publicly at www.nationalcoreindicators.org. At present, a set of NCI measures are under consideration for endorsement by NQF.

It is from the experience of our work with State DD Systems to use quality of life and quality outcomes data to enhance services and supports that we offer these comments that support and are very similar to our NCI partner, HSRI:

We recognize the need to have service population differences and awareness in any PRO-PM development. Long Term Services and Supports (LTSS) and HCBS are not monolithic across populations and are a complex labyrinth of services, supports, quality expectations and provider networks. The LTSS population includes people of all ages, races and ethnicities, household composition, and geographic regions. It includes people with a wide variety of health conditions, including acute, chronic, and behavioral health conditions. It also includes people with a range of functional support needs, from those who need support with most life activities, to those with very few but critical support and service needs. In addition, the LTSS population is growing more racially and ethnically diverse, which has implications for ensuring cultural competency and language access in service delivery policies and practices including ensuring quality.

Community based supports, such Home and Community-Based (HCBS), are broadly defined in order to be tailored to be person-centered, ensuring the person’s individual support preferences, needs and circumstances be taken into account. Each individual person’s person-centered supports are based on true community and can include a variety of supports including multiple paid and unpaid supports over different ranges of time for the person to be able to live their personally defined “good life.”

There are several key life domains in NCI that parallel several of the quality domains outlined in NQF's Quality in HCBS to Support Community Living: Addressing Gaps in Performance Measurement (September 2016). This report, as well as the priorities outlined in the NQF Person Centered Planning and Practice Final Report (2018) point to important quality elements for person centered practices, including skills required for supporters and facilitators of person-centered practice.

National Core Indicators surveys yield person-reported outcomes, providing a statewide and national perspective of people with IDD using the LTSS delivery system. NCI states have used this person-reported outcome measure data to examine quality in key system areas such as health, employment supports, and supports for developing and maintain relationships.

On the question of whether a person-centered outcome (PRO) is well suited for performance measurement (PRO-PM), we offer the following:

1. Quality measures include those such as relationships and employment; however, the attribution of responsibility to any one facilitator, service provider, or process (and therefore target for PRO-PM) in a community-based setting becomes increasingly challenging the more the person is fully engaged in their community.

2. Person-reported outcome measures can require significant effort to collect and report in a valid manner, particularly across population types. The provider network is diverse and the deployment design would require an additional understanding of the network, including that people usually have multiple community based service providers as well as a wide array of potential services.

3. In order to represent the LTSS population, the PROs must be truly tested with great vigor for individuals with a wide range of disabilities, communication styles/strategies and support needs. In addition, there is a need to continue to test the ability of PROs to accurately reflect person-centered priorities for people from diverse communities and for people with lived experience.

We wish you much success and applaud NQF for pursuing this important work to drive quality improvement from the perspectives of people with IDD and are hopeful that this work will serve as another step toward consistent, equitable, and high quality supports. Please do not hesitate to reach out if you have any questions or desire further clarification for our comments.

RESPONSE

Thank you for your comment. The "Building a Roadmap" Interim Report is not focused on any specific setting or discipline and is intended to discuss PROMs and PRO-PMs at an agnostic level. We agree that attribution is an important element of performance measurement. We have added language that explains this report uses the word "attributes" not in the context of attribution but rather to describe the attributes or characteristics of a PROM that is suitable for use in a PRO-PM. The upcoming Technical Guidance Report will discuss attribution at a high level, but attribution is largely out-of-scope for this project. We recommend that you follow the work of the ongoing NQF project, "Attribution for Critical Illness and Injury," as it might contain information you find useful. We agree with the challenges you raise around person-centeredness and health equity. The attributes "Cultural Appropriateness, Language, and Translations With Culturally Appropriate Items" and "Guidance on Standardized Data Collection (Including Modes and Methods)" attempt to look at issues around language and digital divide, and we will explore these issues in the upcoming Technical Guidance Report. We have revised language

in the report to better reflect patient-centeredness. You also might be interested in reading NQF's 2020 report, "Patient-Reported Outcomes: Best Practices on Selection and Data Collection."