



**NATIONAL
QUALITY FORUM**

Driving measurable health
improvements together

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

Final Report

May 18, 2021

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

Contents

- Executive Summary 1
- Introduction 2
- Background 2
 - Previous Work in Patient-Reported Outcomes in Performance Measurement 2
 - Terminology 4
- Environmental Scan Goals and Objectives 5
- Environmental Scan Methodology..... 5
 - Literature Review 5
 - Measure Scan of Existing PROMs and PRO-PMs 6
 - Discussions With Experts 6
- Environmental Scan Findings 7
 - Role of PROMs and PRO-PMs in Quality-Based Models 7
 - Potential Resources for Identifying Candidate PROMs for Performance Measurement 8
 - Currently Available Guidance for Developing PRO-PMs 11
 - Review of NQF PRO-PM Endorsement Process 12
 - Challenges and Barriers of Developing PRO-PMs 15
- Limitations of the Environmental Scan Report 20
- Conclusion..... 20
- References 21
- Appendices..... 26
 - Appendix A: Attribute Grid Example 26
 - Appendix B: Search Terms 27
 - Appendix C: PROMs Reviewed for This Report 28
 - Appendix D: Reference Tables 28
 - Appendix E: Committee Members, Federal Liaisons, and NQF Staff 32
 - Appendix F: Public Comments 36

Executive Summary

The National Quality Forum (NQF) and the Centers for Medicare & Medicaid Services (CMS) have a long-standing partnership to advance patient-reported outcomes (PROs). Despite this shared commitment, progress towards the widespread development and use of patient-reported outcome measures (PROMs) and patient-reported outcome performance measures (PRO-PMs) has been slow: As of April 1, 2021, only 29 PRO-PMs were endorsed by NQF.

With the CMS-funded initiative titled [Building a Roadmap From Patient-Reported Outcome Measures to Patient-Reported Outcome Performance Measures](#), NQF aims to identify and reduce the barriers to the development of digital PRO-PMs (i.e., performance measures that automatically pull data generated during the normal course of clinical care and calculate scores within an electronic health record [EHR] system). NQF intends to achieve this goal by providing guidance for the development of digital PRO-PMs that utilize data from high quality PROMs, are suitable for use in CMS' Value-Based Purchasing (VBP) programs and alternative payment models (APMs), and can be calculated and transmitted via EHR systems and other technologies.

The environmental scan report is the first of three reports related to this initiative that NQF will publish in 2021, and it assesses the current state of PRO-PM development. It begins with an overview of the trend away from fee-for-service reimbursement and toward quality-based payment. The report reviews the CMS Meaningful Measures initiatives, including the goal to amplify patients' voices through the use of PROMs and CMS' aim to have 100 percent of digital measures fully interoperable by 2025.^{1,2}

The scan presents a brief overview of several PROMs that are widely embraced in different healthcare settings (e.g., health systems, payers, and federal agencies). This overview lays the groundwork for the interim report, the second report in the initiative, in which the Technical Expert Panel (TEP) will review multiple PROMs and identify the attributes of a high quality PROM that is suitable to be the basis for a PRO-PM. While the TEP will not prepare a list of high quality PROMs, the environmental scan report includes resources (e.g., NQF's CMS-funded 2020 report titled [Patient-Reported Outcomes: Best Practices on Selection and Data Collection](#), the International Consortium for Health Outcomes [ICHOM] [Standard Sets](#), and recommendations from professional societies) that measure developers and other stakeholders can use to identify PROMs that might serve as the base for future PRO-PMs.³ The scan also includes a discussion of the limited number of existing resources to guide measure developers through the development of digital PRO-PMs that are suitable for VBP programs and APMs.

Given the small number of endorsed PRO-PMs, the scan offers an overview of NQF's endorsement process for PRO-PMs. This includes the Scientific Methods Panel's (SMP) review of PRO-PMs as complex measures and the assessment of each PRO-PM against the five criteria of the Consensus Development Process (CDP). (NQF does not endorse PROMs alone, but any PROM used within a PRO-PM must be identified in that PRO-PM's measure specification.⁴) The report then describes challenges related to the endorsement of PRO-PMs, including the reliability and validity, feasibility, and usability and use criteria within the CDP.

Additional challenges beyond the endorsement process are highlighted, which include balancing the advantages and disadvantages of using one PROM as the base for a PRO-PM as opposed to multiple

PROMs; overcoming technical considerations related to interoperability, data standardization, and coding; and addressing data challenges faced by patients and caregivers.

NQF will describe attributes of high quality PROMs in a second report, to be published in summer 2021.

The third report, which NQF will publish in autumn 2021, will comprise a step-by-step “roadmap” for measure developers to use when developing digital PRO-PMs.

Introduction

An opportunity exists for measure developers to create digital PRO-PMs—in which EHR systems not only collect data but also calculate and submit aggregate scores for regulatory and reimbursement purposes—that are based on high quality PROMs. For this to occur, measure developers need access to a list of attributes of high quality PROMs for use in performance measures and a roadmap to follow when creating digital PRO-PMs for accountability purposes.

This initiative supports the development of digital PRO-PMs that are based in high quality PROMs and that may be appropriate for CMSVBP programs or APMs. The purpose of this environmental scan report is to document currently available guidance on best practices for identifying high quality PROMs for use in digital PRO-PMs. NQF will develop two additional reports after the completion of the environmental scan report. The interim report will highlight attributes of high quality PROMs, and the final technical guidance will comprise step-by-step guidance to measure developers who are using these PROMs as the basis for developing PRO-PMs that are suitable for CMS accountability programs.

Background

Previous Work in Patient-Reported Outcomes in Performance Measurement

Over the past decade, NQF has actively participated in the development of numerous reports 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.

NQF endorses performance metrics but does not endorse instruments or scales (including PROMs) on their own.⁴ If a PROM is explicitly identified in the specification of a PRO-PM, that PROM is reviewed for reliability and validity as part of the endorsement process. However, NQF remains agnostic to the specific instrument and reviews it only to the extent that it meets an acceptable scientific standard as an element of the PRO-PM.

NQF currently has 29 endorsed PRO-PMs that 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 pain), and settings (e.g., ambulatory, inpatient, long-term care, and hospice).⁵ An overview of each of NQF’s projects to further PROs, PROMs, and PRO-PMs follows below.

In 2012, with funding from CMS, NQF launched the PROs in Performance Measurement project. The project included two commissioned background reports. The first report, [Methodological Issues in the Selection, Administration and Use of Patient-Reported Outcomes in Performance Measurement in Health Care Settings](#), focused on selecting PROMs for use in performance measurement and was

updated in 2015 by David Cella and his colleagues.^{6,7} The second report, [PRO-Based Performance Measures for Healthcare Accountable Entities](#), focused on the reliability and validity of PRO-PMs.⁸ As part of the project, NQF also convened two meetings of an Expert Panel who contributed to the development of the 2013 report titled [Patient-Reported Outcomes in Performance Measurement](#).⁴ This work brought together diverse experts to lay the groundwork for developing, testing, endorsing, and implementing PRO-PMs.

In 2017, NQF partnered with PatientsLikeMe, and with funding from the Robert Wood Johnson Foundation, it developed and published [Measuring What Matters to Patients: Innovations in Integrating the Patient Experience into Development of Meaningful Performance Measures](#).⁹ In this report, the authors reiterated the importance of patient-centered quality measurement and demonstrated the value that online patient communities could offer to measure developers and other stakeholders involved with PROs.

Many challenges have become clearer since the previous reports were published, such as clinician resistance to PROs, burdensome workflows related to data collection, and unclear funding sources to support the use of PROMs. In 2019, CMS funded the first of two new projects with NQF, and in September 2020, NQF published the final report from the first initiative. In *Patient-Reported Outcomes: Best Practices on Selection and Data Collection*, NQF identified best practices to help clinicians and administrators select and implement PROs and PROMs in care settings.¹⁰ This report presented solutions to common challenges, such as increasing clinician support by securing physician buy-in before launching a PRO program, engaging staff in developing feasible and effective clinical workflows, and collaborating with leadership to identify funding sources to offset the costs of collecting and using PRO data.

The Expert Panelists for the 2020 initiative designed a PROM Attribute Grid that guides the selection of high quality PROMs within a clinic or health system ([Appendix A](#)). While this grid was intended for clinicians who were selecting PROs and implementing PROMs, it may also provide insight into the attributes of high quality PROM-based performance measures. As a result, this grid directly informs the current initiative, which is also funded by CMS. In *Building a Roadmap From Patient-Reported Outcome Measures to Patient-Reported Outcome Performance Measures*, NQF will revisit the work from 2012, 2013, and 2015 to understand what has—and has not—worked well in using high quality PROMs as the basis for PRO-PMs that can be used for accountability. NQF will also provide guidance to advance the development of digital PRO-PMs based on high quality PROMs.

The TEP for the current initiative comprises multistakeholder experts who represent measure developers, health information technology (IT) professionals, payers, researchers, clinicians, and other healthcare perspectives. Importantly, the TEP also includes patients and patient advocates, both of whom are critical voices for this effort. Because of the highly technical nature of this topic and the focus on PROMs and PRO-PMs that are used by federal agencies, NQF intentionally included individuals and organizations that are involved in the development and/or stewardship of PROMs, including the following:

- Patient-Reported Outcomes Measurement Information System (PROMIS)
- Kansas City Cardiomyopathy Questionnaire – 12 item (KCCQ-12)
- Hip disability and Osteoarthritis Outcome Score, Joint Replacement (HOOS, JR)

- Knee injury and Osteoarthritis Outcome Score, Joint Replacement (KOOS, JR)
- PRO-PMs related to depression and hip/knee replacement outcomes

Terminology

In this paper, NQF will continue to use established terminology from the 2013 report to distinguish between PROs, PROMs, and PRO-PMs (Table 1). Additionally, this scan uses the following terminology related to measurement:

- **Digital quality measures (dQMs):** These measures automatically pull 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.¹
- **Electronic clinical quality measures (eCQMs):** These are the most recognizable of the digital quality 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.¹¹
- **Performance measures:** 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.¹²

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. ⁴	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 health care 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. ⁴	Patient Health Questionnaire 9 (PHQ-9) [®] , a standardized tool to assess depression

Concept	Definition	Example
PRO-Based Performance Measure (PRO-PM)	A performance measure that is based on patient-reported outcomes assessed through data often collected through a PROM and then aggregated for an accountable healthcare entity. ⁴	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)

Environmental Scan Goals and Objectives

The goal of the environmental scan report is to provide the TEP with a clear summary of the current state of using high quality PROMs as the basis for PRO-PMs in accountability programs. The scan contains existing guidance that measure developers can use when developing PRO-PMs. It also incorporates discussion of how well that guidance serves people at various stages in their careers, from new staff members to experienced measure developers. The scan also contains an overview of current approaches and requirements for testing PRO-PMs, including evaluations of reliability, validity, usability, and feasibility. The scan includes a review of resources that can help measure developers identify high quality PROMs, as well as gaps that could hinder the identification of these PROMs.

Although the environmental scan report (as well as the subsequent two reports that will emerge from this initiative) is relevant to the development of PROM-based PRO-PMs, this project focuses on digital PRO-PMs. Digital PRO-PMs collect outcome data from patients with minimal burden, maximize response rates to PROMs to increase representativeness, and leverage EHRs for data collection, storage, and measure calculations.

Because the phrase “high quality” is open to interpretation, the scan also consists of materials to help the TEP clarify a working definition of *high quality PROMs*. The scan includes potential examples of high quality PROMs that are currently in use as part of CMS VBP programs or APMs or ones that serve as the basis for NQF-endorsed PRO-PMs. Given the focus on digital PRO-PMs, the scan includes a brief review of issues related to interoperability, as well as issues related to both patient and health system burden in completing PROMs and how EHRs and digital quality measures can reduce that burden.

Environmental Scan Methodology

NQF conducted the environmental scan using three interrelated approaches. First, NQF conducted a literature review to assess the body of literature related to PROMs and PRO-PMs and identify those articles most relevant to this initiative. Second, NQF conducted a closely related scan of existing PROMs and PRO-PMs, as well as the organizational bodies that assess the quality of these measures. Third, NQF held discussions with experts in fields related to PROMs and PRO-PMs, including targeted discussions during the web meetings of the TEP. Each of these approaches is outlined in more detail below.

Literature Review

To support the goals and objectives, NQF conducted a literature review to provide the TEP with an overview of the current landscape of PROMs and PRO-PMs. The literature review included a search for information sources that detail attributes of high quality PROMs that can be used in CMS’ VBP programs or APMs.

Methods

Databases for the literature review included PubMed/Medline and Google Scholar. NQF conducted a targeted search within these databases using various combinations of keywords that were derived terms related to guidance on developing PRO-PMs (including digital PRO-PMs) as well as general terms to capture broader work that may include relevant information ([Appendix B](#)). In order to maintain focus on current recommendations and practices, NQF confined the search to English-language work published between 2015 and present day, unless an older source remains an important part of the body of literature (i.e., it is noted as important by the TEP, it is widely recognized or cited by experts in the field, and/or its conclusions or recommendations remain relevant and have not been significantly revised or disproven). The findings from the literature review informed identification of current measure gaps and challenges in developing and implementing PRO-PMs. Collectively, the information gathered will support the development of a roadmap for the creation of digital PRO-PMs from PROMs.

NQF also included grey literature in the literature review and considered papers and websites from government, not-for-profit, and corporate organizations for the environmental scan. The project team conducted additional searches using Google, with the intent of identifying grey literature that did not appear in the database searches. The CMS Innovation Models website is one source that NQF extensively reviewed for the scan in order to accurately represent the current state of PROMs and PRO-PMs in VBP programs and APMs.¹³

NQF also reviewed and listed the websites related to each PROM in the [PROMs in Use With CMS VBP Programs or APMs](#) section ([Appendix C](#)). The project team located these websites via Google searches focused on the copyright, licensing, and/or developer information for each PROM.

Within the environmental scan report, unless a fact or recommendation is explicitly attributed to a specific source, NQF gathered the information from the TEP and synthesized it.

Measure Scan of Existing PROMs and PRO-PMs

NQF conducted a measure scan for NQF-endorsed PRO-PMs ([Appendix D, Table 2](#)). Because this initiative aims to create a roadmap that will guide measure developers to the beginning of the NQF endorsement process, the scan centered on PRO-PMs that are currently endorsed by NQF. Additionally, the scan identified potential high quality PROMs and primarily focused on those that are used in federal programs and/or have widespread adoption across a range of clinical settings. NQF's scan for PROMs and PRO-PMs included repositories such as NQF's Quality Positioning System (QPS), the CMS Measure Inventory Tool (CMIT), and ICHOM's Standard Sets.

Discussions With Experts

NQF selected 25 experts to serve on the TEP. These experts bring diverse perspectives on developing PROM-based PRO-PMs for use in VBP programs and APMs, including viewpoints of measure developers and patients. Because the literature review and measure scan did not reveal extensive information on identifying a high quality PROM as the basis of a performance measure, the information presented in this report is partially based upon discussion that occurred during the first three meetings of the TEP. These discussions were moderated by NQF staff and facilitated by the co-chairs of the TEP. Information elicited during the discussions included anecdotal experiences that were common to multiple TEP members, as well as professional activities related to performance measurement that are not

represented in the literature. The two documents that follow the environmental scan report will also utilize key informant interviews and/or focus groups with multistakeholder experts.

Environmental Scan Findings

Role of PROMs and PRO-PMs in Quality-Based Models

The environmental scan report confirms the importance placed on PROMs and PRO-PMs by a broad range of healthcare stakeholders, including federal agencies, payers, health systems, professional societies, patient advocacy organizations, and quality improvement organizations.

CMS and Industry-Wide Perspective

There has been an industry-wide shift away from fee-for-service reimbursement to value-based payment models, which includes discussions about the role of PROMs in value-based payment. The use of PRO-PMs in accountability and value-based payment initiatives has the potential to promote patient-centeredness, improve care, and lower cost.^{14,15} The industry has identified important aspects of successful PRO-PM development, including a clear analysis plan and framework for interpreting results, appropriate measurement scales, and actionable results.¹⁶

CMS has supported the shift toward value-based care through a variety of programs and initiatives that encourage the use of PROMs and PRO-PMs in quality measurement and improvement. In 2017, CMS launched the [Meaningful Measures Initiative](#), which identifies and prioritizes areas for quality measurement and improvement.² This initiative also helps to identify and close important measurement gaps, align measures across both the continuum of care and payers, and spur innovation in new types of measures, such as patient-reported measures and electronic measures.¹⁷ CMS identified PRO-PMs in this initiative as a way of unleashing the patient voice, to drive measures toward patient-centered, value-based care through the development, selection, and implementation of quality measurement.¹⁷ In addition to the Meaningful Measures Initiative, CMS also sets priorities based on input from the [National Impact Assessment](#) of CMS Quality Measure Reports, further emphasizing the importance of prioritizing PROMs and measures using patient-generated data.¹⁷

Although there is increasing discussion and attention given to PROM and PRO-PM adoption, there is a small number of examples of payer implementation available to date. One payer example of a large-scale implementation of PROMs is the effort by Blue Cross Blue Shield of Massachusetts (BCBSMA) to incorporate PROMs into clinical care.¹⁵ The implementation started with a phased adoption of PHQ-9 for depression or HOOS, JR / KOOS, JR, utilized for orthopedic joint replacement, and ended with the goal of using PRO-PMs to create outcome accountability.¹⁵ The BCBSMA case demonstrates potential ways in which PROM adoption can improve diagnosis and treatment, such as improved diagnosis and longitudinal tracking of depression as well as accurate prediction of outcomes from baseline functioning scores for hip and knee replacement patients.¹⁵ However, a number of challenges in key areas remain, such as implementation and endorsement, and the integration of PRO-PMs in quality-based models remains largely an empirically and operationally intimidating endeavor.¹⁵ Therefore, as measure developers create digital PRO-PMs that are based on high quality PROMs and as EHR systems collect data, calculate, and submit aggregate scores for regulatory and reimbursement purposes, measure developers need access to a list of attributes of high quality PROMs for use in performance measures and a roadmap to follow when creating digital PRO-PMs for accountability purposes.

CMS Goals Related to PRO-PMs and Digital PRO-PMs

One aim of CMS' Meaningful Measures 2.0 is to amplify patients' voices through the use of PROMs. CMS aims to have 100 percent of digital measures that are fully interoperable (i.e., allowing for data entry, storage, integration, calculation, and reporting to be conducted by EHRs and enabling submitted information to be used in multiple ways) by 2025 to promote important patient-centered goals, such as increasing the alignment across the quality measurement enterprise and improving care coordination.¹ The ambitious goal of modernizing and digitizing quality measures and programs includes key steps, such as finalizing a digital measure strategy and advancing the electronic data infrastructure.¹ Moreover, CMS aims to improve the collection and integration of patient voices across programs by increasing the use of PROMs and improving their integration into the EHR workflow.¹⁸ One of the ways in which CMS plans to integrate PROMs into the EHR workflow is by aligning the EHR certification process with other CMS reporting requirements.¹

Potential Resources for Identifying Candidate PROMs for Performance Measurement

PROMs in Use With CMS VBP Programs or APMs

With the passing of the Affordable Care Act of 2010, the U.S. healthcare system has shifted towards improving and rewarding value.¹⁹ CMS designed the VBP Program to increase the quality of care and experience for patients.²⁰ There are several VBP programs that can apply to various provider settings, such as hospitals and outpatient centers.

In addition to VBPs, one of the tracks of CMS' Quality Payment Program (QPP) is the APM, which provides incentive to eligible participants to ensure high quality and cost-efficient care is provided.²¹ VBPs and APMs are likely to interact during this shift towards improving value, given that incentives linked to APMs may be in similar form to VBP programs for some providers.¹⁹

Given that the patient experience is so critical to quality measurement, payers will need to leverage the instruments that measure and account for the patient voice. To help improve value, the patient voice will need to be included within VBP programs and when an appropriate APM has been selected by an organization.¹⁹

In 2015, Cella et al shared that attributes of PROMs include reliability, validity, interoperability of scores, minimal patient or caregiver burden, alternative methods of administration, adaptability to different cultures and languages, and the ability to be incorporated into EHRs.⁷ In addition to these attributes, measure precision and sensitivity to change are shared among most PROMs. The following list highlights a few PROMs that may demonstrate shared attributes of high quality PROMs. While this list identifies some specific PROMs that CMS has selected for use in accountability programs, this is *not* an endorsement of any individual PROM. Rather, it provides examples of some PROMs that may be candidates for performance measures. Newer PROMs that have not been reviewed for use by federal agencies are not included in this list but should be considered by developers as potential PROMs for performance measurement.

- [National Cancer Institute's \(NCI\) Patient-Reported Outcomes version of the Common Terminology Criteria for Adverse Events \(PRO-CTCAE\)](#): This tool was designed to measure and evaluate symptoms and adverse events for participants in cancer clinical trials.²² The PRO-CTCAE Measurement System should be utilized with the Common Terminology Criteria for Adverse

Events (CTCAE) due to the supplemental information that the PRO-CTCAE can provide clinicians.²² The PRO-CTCAE functions to enhance the precision of adverse-event reporting in cancer clinical trials, to provide useful data for clinicians, and to ensure that the patient perspective related to experiencing an adverse event is collected. Given the favorable test-retest reliability (median ICC 0.77) in a sample of 975 patients, as well as being linguistically validated, PRO-CTCAE demonstrates strong validity, reliability, and responsiveness.²³

- National Institutes of Health (NIH)-funded initiative Patient-Reported Outcomes Measurement Information System® (PROMIS)**: This initiative was established in 2004 with the goal of standardizing measures to allow for different PRO domains to be assessed.²⁴ The set of standards for the development and validation of item banks and instruments within PROMIS provides a useful tool for developers. PROMIS offers short forms, computerized adaptive testing (CAT), and profiles (i.e., fixed collection of short forms from multiple domains), as well as appropriate use across a range of patient populations. In using PROMIS measures with CAT, measures usually only require four to six items for precise measurement of health-related constructs, thereby reducing respondent burden.
- Patient Health Questionnaire 9 (PHQ-9)**: This instrument has been in use since 1999 after it was developed through a grant from Pfizer. As a nine-question instrument, the PHQ-9 is shorter than its PHQ predecessor and assesses the presence and intensity of depression and depression symptoms. The PHQ-9 is defined in the denominator of four NQF-endorsed PRO-PMs that are stewarded by MN Community Measurement and are related to depression remission and depression response at six and 12 months. There are also PRO-PMs related to the utilization of the PHQ-9 tool with patients diagnosed with depression and bipolar disorder. Given its use in various medical specialty areas, the PHQ-9 and the related performance measures are one example of how a widely adopted PROM can be used as the basis for NQF-endorsed PRO-PMs.
- Kansas City Cardiomyopathy Questionnaire – 12 item (KCCQ-12)**: This PROM is a sensitive and specific health-related quality of life measure for patients with heart failure (HF).²⁵ Similar to PHQ-9, the KCCQ-12 is a truncated version of the KCCQ. Although the KCCQ has been shown to be valid, reliable, and sensitive, its length (23 questions) has been a barrier to gaining insight on the patient experience. An additional instrument that has been utilized to assess quality of life among HF patients is the Minnesota Living with Heart Failure Questionnaire (MLHFQ). The MLHFQ has two domains—physical and emotional—and is a self-administered instrument.²⁶ Both instruments are commonly used in clinical research and have the potential to predict outcomes important to clinicians of HF patients.²⁷
- Hip disability and Osteoarthritis Outcome Score, Joint Replacement (HOOS, JR) and Knee injury and Osteoarthritis Outcome Score, Joint Replacement (KOOS, JR)**: Within orthopedics, PROMs are utilized for several conditions, including ligament injuries and joint replacements.²⁸ Two examples of validated and commonly used PROMs for knee injury and joint replacement include the KOOS, JR and the HOOS, JR. Both the KOOS, JR and HOOS, JR PROMs assess outcomes after total knee arthroplasty (TKA) and total hip arthroplasty (THA), respectively, and are short-form measures that help to reduce patient survey fatigue.^{29,30} In separate validation studies, high internal consistency and high responsiveness were seen with a Pearson Separation Index of 0.84 for KOOS, JR and 0.86 for HOOS, JR.^{31,32} An additional attribute of both measures is the output of a single score that clearly relays knee and hip “health” to the clinician. The KOOS,

JR and HOOS, JR have crosswalks that allow scores from these PROMs to be converted to Oxford Knee Scores and Oxford Hip Scores (other PROMs that are widely used after TKA and THA), and vice versa.³⁰ The KOOS, JR and HOOS, JR are also both included as acceptable PROMs to meet the reporting requirements for CMS' Comprehensive Care for Joint Replacement Model.³³

NQF Resources

While NQF does not endorse PROMs or other data collection instruments for PROs, the organization does endorse PRO-PMs. When a PRO-PM is based on a specific PROM, the reliability and validity of that PROM is considered during the endorsement process. Measure developers should thoroughly review the process used to validate PROMs, including sample size when identifying candidate instruments for PROM-based performance measures.³⁴ As mentioned in the [Background](#) section, the September 2020 NQF report titled *Patient-Reported Outcomes: Best Practices on Selection and Data Collection* offers guidance to clinicians and health systems that are choosing PROMs.¹⁰ This report features guidance that developers may consider when determining if a PROM is of high quality.

ICHOM Standard Sets

ICHOM was founded in 2012 with the intent of creating “critical foundations for value-based healthcare.”³⁵ Part of the organization’s work has focused on convening clinical experts and patients to develop Standard Sets of outcomes, measurement tools, time points, and risk adjustment factors for specific conditions.¹³ As of 2020, ICHOM has published 39 Standard Sets, each of which is a pragmatic measurement recommendation based on a working group’s comprehensive review of relevant PROMs and other measures and data sources. As an example, the Hip and Knee Osteoarthritis Standard Set identifies a minimum data set of case-mix variables, treatment variables, and outcomes, then recommends three potential HRQoL PROMs, a visual pain scale, and a hip- and knee-specific physical function PROM.³⁶ Because of ICHOM’s focus on outcomes that matter most to patients and a vetting process that involves clinical experts and consumers, the Standard Sets are a potential source of high quality PROMs. The Standard Sets include common chronic diseases, such as diabetes; population-specific sets, including older person primary and preventive care and hypertension in low- and middle-income countries; and behavioral health conditions, including dementia, depression, and anxiety.

PROMs Identified by Professional Societies

Professional societies can be a valuable resource for identifying PROMs that may be strong foundations for future PRO-PMs. Many professional societies have convened working groups to evaluate PROMs and recommend those that meet certain criteria, such as patient-centeredness, cost, modality (e.g., digital entry, paper questionnaires), completion time, clinical meaningfulness, and widespread clinical adoption. Approaches and recommendations from three societies are listed below, but there are numerous associations and societies that have published comparable recommendations on their websites, in white papers, journals, and other media.

- **American Academy of Orthopaedic Surgeons (AAOS):** This academy established the Quality Outcomes Data (QOD) Work Group in 2015 whose charge was, among other tasks, to evaluate PROMs. This workgroup evaluated instruments for PROs in orthopedics against the criteria of free use, inclusion of only patient-reported data, multiple modalities, number of questions, responsiveness, one generic quality of the PROM, no more than three joint or disease-specific PROMs, and availability of CAT.¹³ As a result of this work, AAOS developed a set of

recommended PROMs for upper extremities (e.g., shoulder and shoulder instability, along with elbow, wrist, and hand), lower extremities (e.g., foot and ankle, knee, and hip), spine, and disease-agnostic quality of life.^{37–40}

- **Society of Gynecologic Oncology (SGO):** This society convened a daylong meeting of its Policy, Quality, and Outcomes Taskforce in 2018 that resulted in disease-specific recommendations for PROs data collection using the Functional Assessment of Cancer Therapy (FACT)-G7 as a general HRQoL questionnaire; disease-specific PROMs for ovarian, uterine, cervical, vulvar, and vaginal cancers; and instruments that specifically address sexual health in women with cancer.^{41,42}
- **American Academy of Neurology (AAN):** Some societies opt to list PROMs that are common in their field. While these lists may not be as rigorously vetted as those from societies that assign dedicated working groups to recommend PROMs, they can still be useful in identifying PROMs that may provide meaningful bases for PRO-PMs. AAN provides a brief list of common PROMs used in neurology, including cross-cutting instruments, such as PROMIS and PHQ-9 as well as condition-specific scales and tools for dementia, headache, epilepsy, and multiple sclerosis.⁴³

Regardless of how societies assemble a list of preferred PROMs, their research and recommendations can be useful for identifying high quality PROMs that may be suitable bases for digital PRO-PMs targeted to accountability programs.

Currently Available Guidance for Developing PRO-PMs

Availability and Quality of Detailed, Step-by-Step Instructions on PRO-PM Development

Currently, there is limited guidance on the development of PRO-PMs from PROMs and how to adequately address the challenges of understanding and processing PRO data.¹⁴ As a result, practical, step-by-step guidance is needed to assist measure developers with the identification of high quality PROMs and the development of related PRO-PMs.¹⁴

As part of its Measures Management System (MMS) Blueprint supplement, CMS outlines the following steps for developing PRO-PMs: “Choose and define a PRO, determine the appropriate way to collect the PRO using a PROM (tool), and determine the appropriate performance measure.”⁴⁴ While these steps are important, this guidance lacks the level of detail to be useful to measure developers who are navigating the challenges of PRO-PM development.

A TEP assembled by the American Medical Association (AMA) prepared a guidance document that provides more detailed recommendations. The paper written by Basch et al identified nine best practices for developing PRO-PMs from 13 PRO programs and 10 guidance documents ([Appendix D, Table 3](#)).¹⁴ Several of the best practices correspond to NQF’s measure evaluation criteria on reliability, validity, usability, and feasibility.¹⁴ The best practices were developed with the goal of supporting future development of robust approaches to better understand the impact of care on the patient experience.¹⁴ However, the TEP also noted that although best practices, use cases, and guidance documents are available for PRO-PM development, the evidence and experience in this area is limited and best practices will most likely evolve with more evidence.¹⁴ Neither the Basch et al paper nor the 10 guidance documents reviewed by its authors specifically addressed the development of digital PRO-PMs that are

based on high quality PROMs, which emphasizes the need for the guidance that will be provided by this new initiative.

Ease of Use for Novice and Experienced PRO-PM Developers

Since limited guidance is available regarding PRO-PM development, there is a pressing need to create more detailed instructions for guiding measure developers, regardless of experience, through the PRO-PM development process. In addition to further guidance on development, a roadmap that helps developers navigate NQF's endorsement process for PRO-PMs is also useful in ensuring that clear information is provided to developers that are inclined to gain input and consensus from an array of healthcare stakeholders.

Promising Strategies for Patient Burden Reduction

There are several common patient burden factors related to PROMs, including limited patient understanding of the importance of PROM completion, excessive time to complete a questionnaire, questions that may be perceived as intrusive or irrelevant, selection bias, and low participation from certain vulnerable populations.¹⁰ Therefore, to support the reduction of patient burden, developers should consider the frequency of when patients are asked to complete questionnaires, the number of questions on instruments, and the types of questions included in PROMs.¹⁰ One strategy to minimize response burden for patients includes using brief items and questionnaires written in plain language with a clear purpose that are valid, reliable, and sensitive to change over time.^{14,45} Other strategies include:

- providing multiple modalities (e.g., paper, phone, email, and patient portal); note, however, that modalities such as paper can be counterproductive for digital PRO-PMs and can have data implications on the use of performance measures, as discussed in the “Challenges” section;
- offering PROMs in multiple settings (e.g., home, clinic) for completion;
- using yes/no and multiple-choice questions instead of open-ended questions; and
- incorporating technologies, such as CAT.¹⁰

CAT algorithms can help to assess PROs by tailoring the questions to the patient's health status.^{10,46} With an advanced method of assessment, CAT selects subsequent items in the PROM based on the patient's response to the first item, which can help improve the quantity and quality of questions and reduce patient burden.^{10,46} Further developing strategies to help reduce patient burden is essential to the goal of increasing the precision and participation of PROMs and the quality of PRO-PMs.¹⁰

Review of NQF PRO-PM Endorsement Process

Themes Related to NQF Analysis

NQF has 29 currently endorsed PRO-PMs ([Appendix D, Table 3](#)). These measures vary in scope but have similar goals of gathering and quantifying PROs. These measures include various topic areas, such as functional status, experience with care (e.g., Consumer Assessment of Healthcare Providers and Systems [CAHPS], CoreQ [i.e., patient, resident, and family satisfaction for skilled nursing care centers and assisted living communities]), depression response or remission, shared decision making/patient activation, transitions of care, pain management, quality of life, and contraception.

NQF-endorsed PRO-PMs are based on PRO data aggregated for an entity deemed as accountable for PROM implementation, collection, and action. PRO-PMs are specifically considered complex measures because of the stringent requirements against which they are evaluated. The criteria unique to PRO-PM consideration include an evaluation of the psychometric properties and testing of the survey instrument as the basis of the measure, which requires statistical methodologies to ensure a reliable and valid survey instrument. Additionally, reliability must be demonstrated at the performance score-level to address precision of measurement challenges related to the survey instruments. These are the main factors qualifying PRO-PMs as complex measures, requiring additional scientific review focused on validity and reliability. NQF considers the following types of measures complex; therefore, they require an evaluation by the SMP:

- Instrument-based measures (e.g., PRO-PMs)
- Outcome measures, including intermediate clinical outcomes or PRO-PMs
- Cost/resource use measures
- Efficiency measures (combining concepts of resource use and quality)
- Composite measures⁴⁷

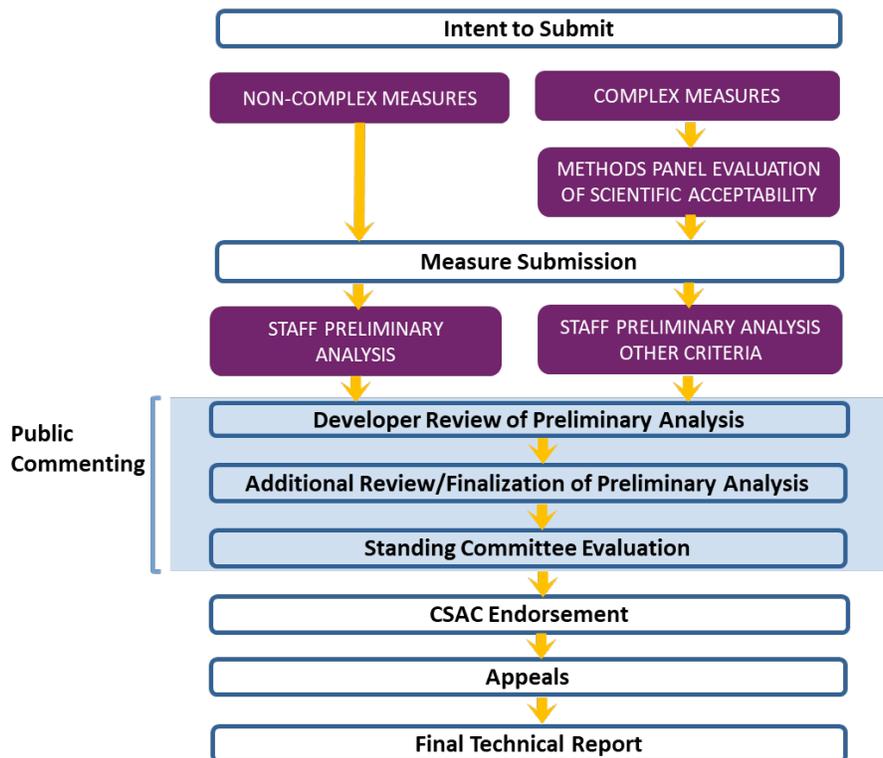


Figure a. This graphic represents a high-level measure workflow through NQF’s CDP.

Additionally, the Trial Use Program may have applicability to a limited set of digital PRO-PMs that also meet eCQM criteria.⁴⁸ The Trial Use Program is specifically for eMeasures that are ready for implementation but cannot yet be adequately tested to meet endorsement criteria.⁴⁸ Requested by a developer, Trial Use designation expires three years after the initial approval date if full endorsement is not achieved.⁴⁸

Scientific Methods Panel

The SMP is currently composed of 28 individuals with methodological expertise who provide NQF Standing Committees with evaluations of measures' scientific acceptability. They use NQF's standard measure evaluation criteria for new and maintenance measures.⁴⁹ Measures rated by the SMP as "low" or "insufficient" for reliability or validity will usually be removed from the current cycle, allowing time for additional testing, clarification, and NQF technical support prior to consideration in a future cycle.⁴⁷ The SMP's feedback is critical for endorsement recommendations by the Standing Committees and for endorsement decisions by the Consensus Standards Approval Committee (CSAC). Although the number of PRO-PMs coming through the SMP process is relatively low compared with other types of measures, the SMP recognizes the inherent complexity of PRO-PMs.

Consensus Development Process and Standing Committee Reviews

The CDP is NQF's formal process to evaluate and endorse measures and is designed to allow input and discussion from stakeholder groups across the industry. The CDP or measure endorsement process, including maintenance of previously endorsed measures, is standardized in a regular, twice-per-year cycle of topic-based measure evaluation across 14 areas, such as cardiology, primary care and chronic illness, cancer, and prevention and population health.⁵⁰ The CDP involves [six principal steps](#):

Step 1: Intent to Submit

Intent to submit notifies NQF at least three months prior to the designated cycle's measurement submission deadline, allowing for adequate opportunity for technical assistance prior to submitting measures for evaluation.⁵¹ Intent to submit includes the planned submission date, submission type, measure title, measure description, measure type, level of analysis, data source, numerator and denominator statements, and testing information.⁵¹

Step 2: Call for Nominations

The purpose of the call for nominations process is to seat a Standing Committee, to help shape a project, develop specific plans for the project, offer expert advice, ensure input is obtained from relevant stakeholders, and make recommendations to NQF membership about standards that are proposed for endorsement.⁵²

Step 3: Measure Review

Measure review is then subsequently conducted by the Standing Committee, as well as a technical advisory panel if applicable. The Standing Committee and/or panel will meet several times to review and discuss the submitted measures. During the measure evaluation, the Standing Committee is expected to reach consensus, as defined in the Office of Management and Budget (OMB) Circular A-119.⁵³

Recommendations of a measure review can be either:

- a candidate measure continues through the CDP toward possible endorsement by NQF; or
- a candidate measure is returned to the standard steward and/or developer for further development and/or refinement.⁵⁴

Step 4: Public Commenting With Member Support

After a recommendation of the Standing Committee is included in a draft report, all recommendations are opened to the public for commenting.⁵⁵ NQF members and interested members of the public can electronically submit comments to express their support or nonsupport of the Committee's

recommendation. The Committee reviews all public and member comments and may choose to revise a recommendation.

Step 5: Measure Endorsement

The CSAC is an advisory Committee whose members are appointed by the NQF Board of Directors. The CSAC reviews the submitted measure for the strategic importance of measures within the portfolio, cross-cutting issues concerning measure properties, and CDP concerns. The CSAC makes a measure endorsement decision and advises of any changes in the CDP.⁵⁶ The CSAC may uphold a Standing Committee's recommendation to endorse or not to endorse a measure, or they can send the measure back for reconsideration.⁵⁷

Step 6: Measure Appeals

After the CSAC's decisions are made public, a 30-day appeals period begins. Any party may file an appeal with the Appeals Board during this period. Measure endorsement decisions that are eligible for appeal must be those attributable to procedural errors reasonably likely to affect the outcome of the original endorsement decision (e.g., a failure to follow NQF's CDP) or new information or evidence that was unavailable at the time the CSAC made its endorsement decision and is reasonably likely to affect the outcome of that decision.⁵⁸

CDP criteria for all measures considered throughout the process are listed below:

- **Criterion 1: Importance to Measure and Report:** Extent to which the specific measure focus is evidence-based, important to making significant gains in healthcare quality, and improving health outcomes for a specific high-impact aspect of healthcare in which there is variation in or overall less-than-optimal performance.
- **Criterion 2: Reliability and Validity – Scientific Acceptability of Measure Properties:** Extent to which the measure produces consistent and credible results about the quality of care when implemented.
- **Criterion 3: Feasibility:** Extent to which the specifications, including measure logic, required data that are readily available or could be captured without undue burden and can be implemented for performance measurement.
- **Criterion 4: Usability and Use:** Extent to which potential audiences (e.g., consumers, purchasers, providers, and policymakers) are using or could use performance results for both accountability and performance improvement to achieve the goal of high quality, efficient healthcare for individuals or populations.
- **Criterion 5: Comparison to Related or Competing Measures:** If a measure meets all criteria and there are endorsed or new related or competing measures, the measures are compared to address harmonization and/or selection of the best measure.⁵⁹

Challenges and Barriers of Developing PRO-PMs

Some challenges of developing PRO-PMs include navigating the NQF endorsement process; identifying whether a PRO-PM should utilize data from one PROM or multiple PROMs; technical issues regarding PROMs and PRO-PMs; implications on patients and caregivers, including low response rates, under-detection of poor performance when sicker patients might not be able to self-report, data fidelity, patient burden, and patient integration with clinical workflow; and achieving stakeholder buy-in.^{14,60–62}

Challenges With NQF Endorsement

In addition to being resource intensive, unique challenges arise for PRO-PMs undergoing the endorsement process, specific to three of the CDP criteria. Additionally, developing an endorsement-ready, PROM-based performance measure requires a PROM that has been used extensively for data collection, which can eliminate the consideration of recently developed PROMs or those that have not been widely adopted within the clinical or research communities.

Criterion 2: Reliability and Validity – Scientific Acceptability of Measure Properties

According to [NQF criteria](#), for instrument-based measures (including PRO-PMs), reliability *and* validity must be demonstrated for the data element level (i.e., the PROM) as well as for the computed performance score (i.e., the PRO-PM). For validity, the related threats to validity must also be addressed (i.e., exclusions, risk adjustment, discriminating performance comparability if multiple PROMs are used). Generally, for other types of measures, empirical testing at the data-element or measure score level is sufficient. This requirement for PRO-PMs adds to the complexity of the endorsement process for these measures.

Reliability of the measure score refers to the proportion of variation in the performance scores due to systematic differences across the measured entities in relation to random variation or noise using beta-binomial signal-to-noise or other statistical testing methods. Although it is critical to demonstrate that PRO-PMs can appropriately tell apart providers' performance, developers are often confused by the two-level testing requirements, thus complicating the confidence in rating providers appropriately.

Criterion 3: Feasibility

An important piece of feasibility is to ensure there is an achievable and implementable plan for data collection, with data or information being retrievable without undue burden. Measures or PRO-PMs tend to be more feasible if they occur during the normal process of care, such as an intake survey to check functional status on an iPad in a waiting room. Electronic measures are generally preferred, as it is easier to collect large amounts of data.

Examples of considerations for undue burden for both clinicians and patients include the financial impact of having to hire an external vendor (e.g., CAHPS) or survey length and timing when requesting feedback from a patient.

Criterion 4: Usability and Use

NQF-endorsed measures should be used in accountability programs and publicly reported to ensure measures remain in use. Measures that are not used in programs or are not publicly reported may not be ideal for endorsement, as they may not be maintained or updated over time.

Importance of Relationships Between PROMs and Performance Measures

Some existing PRO-PMs rely upon data from a single PROM (i.e., a 1:1 relationship between a PROM and a PRO-PM) while others are designed to accommodate multiple PROMs that address a specific disease or condition (i.e., a many:1 relationship). There are advantages and disadvantages to both approaches, and not understanding their respective benefits and drawbacks can present a challenge to measure developers who are considering candidate instruments as data collection tools for performance measures.

The [Technical Challenges in Digital PRO-PM Development](#) section of this report discusses challenges related to the lack of standardized structured data fields across both PROMs and EHRs. This underlines one argument in favor of the 1:1 relationship between PROMs and PRO-PMs: The measure developer can tailor the specification for the performance measure to the unique data structure of a single PROM. Because different instruments use different scores to measure change, a performance measure based on a single PROM only needs to consider one score; this eliminates the need to identify and build scores from different instruments into the measure specification. Measure developers and stewards can more easily maintain a PROM-based performance measure that depends on a single instrument. Lastly, measure developers might want to focus a performance measure on one PROM that exhibits more attributes of a high quality PROM than other similar instruments.

One example of an NQF-endorsed PRO-PM that is based on a single PROM is NQF #0711 *Depression Remission at Six Months*. This measure, stewarded by MN Community Measurement, assesses improvement in depression scores on the PHQ-9 by measuring the number of adults with a diagnosis of major depression and an initial PHQ-9 score greater than nine who have achieved a six-month PHQ-9 score of less than five.⁵ According to its measure specification, MN Community Measurement selected the PHQ-9 as the PROM for this measure because it is “(1) validated with a sensitivity of 0.080 and a specificity of 0.92 with substantial heterogeneity I2 = 82%, (2) widely accepted and utilized in Minnesota, (3) available for clinical use, (4) translated into many languages, and (5) easy for the patient to complete and the provider to score.”⁵ This PRO-PM was initially endorsed in January 2011 and updated in 2015, making it a relevant and time-tested example of a single-PROM performance measure.⁵

Widespread support also exists for a many-to-1 relationship between PROMs and PRO-PMs (i.e., a PRO-PM based on multiple PROMs), in part because it provides clinicians with the flexibility to choose instruments based on appropriateness for their setting (e.g., language translations, licensing costs, and brevity of instrument) rather than requiring them to use a specific questionnaire. While the lack of standardized data fields across PROMs does create a challenge for development of a PRO-PM, the strategies discussed in the *Technical Considerations in Digital PRO-PM Development* section of this report identifies potential avenues for developers to map disparate instruments to a single performance measure within the measure specification. While different instruments use different scoring systems and cut points (i.e., markers in PROMs that indicate the need to screen for a diagnosis or provide treatment), opportunities exist to combine these different approaches into a single measure.⁸

Regardless of whether a 1 PROM-to-1 PRO-PM or a many PROMs-to-1 PRO-PM relationship exists, the quality of PROMs shapes the effectiveness of PRO-PMs. If a PROM suffers from poor design or inaccurate data collection, the PRO-PM will suffer as well. The September 2020 NQF report titled *Patient-Reported Outcomes: Best Practices on Selection and Data Collection* provides extensive guidance on selecting PROs and selecting and implementing high quality PROMs in clinical settings.

Technical Challenges in Digital PRO-PM Development

In 2011, CMS established the Medicare and Medicaid EHR Incentive Programs to encourage providers (including skilled nursing facilities, dialysis facilities, and hospitals) to adopt, implement, upgrade, and demonstrate meaningful use of certified electronic health record technology (CEHRT). These programs are now known as the Promoting Interoperability Programs, moving beyond the requirements of meaningful use to a new phase of EHR measurement with an increased focus on interoperability and

improving patient access to health information.⁶³ One such opportunity is implementing universal standards for data.

Health Level Seven International (HL7) is a nonprofit institute serving the healthcare industry with a focus on clinical and administrative data.⁶⁴ HL7 has been addressing challenges within the healthcare data exchange system, aiming to ensure EHRs are available, discoverable, and understandable in an increasingly digitized system.⁶⁴ Fast Health Interoperability Resources (FHIR) is an HL7 standard for exchanging healthcare information electronically through EHRs and other health IT systems.⁶⁵ FHIR works as a part of HL7 standards to simplify implementation of these standards without sacrificing information integrity by leveraging existing logical and theoretical models to provide a consistent, easy-to-implement, and rigorous mechanism for exchanging data between healthcare applications.⁶⁵

Interpreting clinical laboratory test results depends on an accurate and reliable understanding of units of measure.⁶⁶ To account for differences in reported units, Logical Observation Identifiers Names and Codes (LOINC) provides a separate code for each unit of measure.⁶⁷ Similar to clinical laboratory test results, results of completed PROMs cannot be easily shared with EHRs and other health IT systems unless there are accepted vocabulary standards.⁶⁸ LOINC supports the structure and content of assessment surveys by creating a model to capture the essential aspects of assessments. This model represents the hierarchical panel structure, global item attributes, panel-specific item attributes, and structured answer lists.⁶⁸ LOINC has embraced adapting standardized scales, such as the Glasgow Coma Score and the Apgar Score⁶⁸, and evolving to PROMs, such as the PHQ-2 and PHQ-9, Confusion Assessment Method (CAM), PROMIS, and Outcome and Assessment Information Set (OASIS) assessments.⁶⁸ Today, LOINC supports more than 500 survey instruments.⁶⁹ Despite progress within standardized code sets, such as LOINC, there remains a gap in coding and the storage of individualized codes required for each PROM.

Recommended in NQF's previous report titled *Patient-Reported Outcomes: Best Practices on Selection and Data Collection*, integration of PROMs within EHRs is imperative to successful clinical uptake, whether from the vendor or built locally. These technical challenges point to the importance of both the embedding of the instruments in the EHR and the actionability or utility of the results. While interoperability is an industry-wide initiative, the burden ends up falling on the implementor of the PROMs and PRO-PMs rather than the measure developers. There are also alternative approaches, suggesting that first focusing on the performance measure requirements and specifications, as opposed to the PROM specifications fitting into the current EHR, would facilitate the successful development of PRO-PMs.

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 LOINC.⁷⁰

Patient and Caregiver Data Challenges

Along with questionnaire developers and practitioners, patients and caregivers have long expressed concerns about the inherent burden associated with PROM use. Instrument length, layout, and cognitive load have been noted as factors that can affect the strain on patients and caregivers. While this insight has led to effective changes, such as shortened versions of existing questionnaires, it remains an obstacle to capturing what is most important to the patient as it relates to their health outcomes.

Burden on patients to complete PROMs causes downstream issues for performance measurement. Factors such as low motivation, pain or functional limitations, recall difficulties, and negative survey perception can add to response burden and affect the amount and quality of patient data collected. Additionally, issues regarding social determinants of health and health disparities, such as patients' access to digital tools and language barriers, can lead to less patient engagement in PROMs. Performance measures cannot exist if patients do not complete the questionnaires; therefore, increased focus should be placed on addressing causes of nonresponse.

Physical and cognitive impairments can also have an impact on the completion of PROMs. Patients with severe physical or cognitive impairments may require proxies (i.e., caregivers, family members, or other people who complete PROMs on a patient's behalf). While it is important to ensure all patients can complete PROMs and that caregiver voices are also captured and measured, mixing patient-reported data with proxy-reported data can create data fidelity issues that affect PRO-PMs.

Lastly, attempts to decrease patient burden and increase data collection can have unintended consequences. The [Promising Strategies for Patient Burden Reduction](#) section of this report discusses the potential benefits of offering multiple modalities (e.g., paper and patient portal) to patients, but this can affect the fidelity of the data that are used for digital PRO-PMs. Additionally, alleviating one burden can create a new burden. As an example, patients may need to access multiple websites or applications (apps) in order to complete PROMs for different providers at different health systems; a dialysis patient could need to download three separate apps to complete PROMs for their primary care physician (PCP), nephrologist, and dialysis center.

Other Challenges and Considerations

The implementation of performance measures will have an impact on clinical workflows, data flow, patient experience and satisfaction, clinician engagement, and much more. Patients are particularly affected by workflows that require the completion of lengthy PROMs at redundant intervals, and active engagement by patients is critical to the success of PRO-PMs. All stakeholders, such as measure developers, clinicians, and those within health IT, must be engaged as active partners in addressing data collection and workflows.

In addition to workflow challenges, there can be gaps between what existing PROMs measure and what is valued by patients. A theme in elevating the patient voice through PRO-PMs, or ensuring PRO-PMs are patient-directed, is to focus on what is important to patients. Not all PROMs align with patients' priorities, though, and not all patients share the same priorities. While limitations of existing PROMs are beyond the scope of this environmental scan, it is a reminder that performance measurement must be guided by what matters most to patients.

Measure development is a time-consuming and costly endeavor, and PRO-PMs are classified by NQF as complex measures. This suggests a need to consider whether PRO-PM development would benefit from incentives, such as increased funding for measure developers or less complex development requirements, that could lead to a larger range of organizations and stakeholders involved in development.

Limitations of the Environmental Scan Report

The environmental scan encompassed a measure review of PROMs that are used by federal agencies, health systems, payers, and other key stakeholders, as well as a literature review of guidance on developing PRO-PMs from existing PROMs. To note, there are two main limitations of the environmental scan. First, the PROMs discussed in this report were included as potential examples of high quality PROMs; however, the TEP's recommended attributes of high quality PROMs will not be finalized until the publication of the 2nd report in this initiative. Second, the literature search returned only a short listing of guidance for developing PRO-PMs. NQF's approach remains to continue searching peer-reviewed and grey literature while also engaging with TEP members and measure development experts to advance this area of knowledge.

Conclusion

There is a road that travels from high quality PROMs to NQF-endorsed PRO-PMs to CMS VBP programs and APMs. That road, however, is fraught with barriers and delays. Although there are many PROMs in today's healthcare environment, there is not a clear way to identify those high quality PROMs that will provide a foundation for a digital PRO-PM. The well-defined processes to NQF endorsement can be challenging for PRO-PMs, and the number of endorsed PRO-PMs remains small. While healthcare's technical infrastructure is at an unprecedented level of sophistication, developing and implementing digital PRO-PMs remains difficult. Understanding the current state of these barriers and delays is the first step in navigating the road from high quality PROMs to PRO-PMs. This environmental scan report sets the stage for the creation of clear guidance on creating PRO-PMs, including a list of attributes of high quality PROMs for use in performance measures and a roadmap that measure developers—from entry-level employees just out of school to veteran developers with decades of experience—can follow when creating digital PRO-PMs for CMS accountability programs.

References

- 1 Centers for Medicare & Medicaid Services (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.
- 2 CMS. Meaningful Measures Framework | CMS. Centers for Medicare and Medicaid Services. <https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/QualityInitiativesGenInfo/CMS-Quality-Strategy>. Last accessed January 2021.
- 3 ICHOM. Standard Sets. ICHOM – International Consortium for Health Outcomes Measurement. <https://www.ichom.org/standard-sets/>. Last accessed January 2021.
- 4 National Quality Forum (NQF). Patient reported outcomes (PROs) in performance measurement. January 2013. <https://www.qualityforum.org/WorkArea/linkit.aspx?LinkIdentifier=id&ItemID=72537>.
- 5 NQF. NQF: Quality Positioning System™. <https://www.qualityforum.org/QPS>. Last accessed January 2021.
- 6 Cella D, Hahan E, Jensen S, et al. Methodological issues in the selection, administration and use of patient-reported outcomes in performance measurement in health care settings. September 2012. <http://www.qualityforum.org/WorkArea/linkit.aspx?LinkIdentifier=id&ItemID=72156>.
- 7 Cella DF, Hahn EA, Jensen SE, et al. *Patient-Reported Outcomes in Performance Measurement*. Vol 97. RTI Press Research Triangle Park, NC; 2015.
- 8 Deutsch A, Smith L, Gage B, et al. Patient-reported outcomes in performance measurement commissioned paper on PRO-based performance measures for healthcare accountable entities. October 2012. <http://www.qualityforum.org/WorkArea/linkit.aspx?LinkIdentifier=id&ItemID=72157>.
- 9 Measuring what matters to patients: Innovations in integrating the patient experience into development of meaningful performance measures. August 2017. <http://www.qualityforum.org/WorkArea/linkit.aspx?LinkIdentifier=id&ItemID=85770>.
- 10 NQF. *Patient-Reported Outcomes: Best Practices on Selection and Data Collection - Final Technical Report.*; 2020. http://www.qualityforum.org/Publications/2020/09/Patient-Reported_Outcomes__Best_Practices_on_Selection_and_Data_Collection_-_Final_Technical_Report.aspx. Last accessed December 2020.
- 11 NQF. NQF health information technology glossary: A guide to HIT jargon. https://www.qualityforum.org/Measuring_Performance/HIT_Glossary.aspx.
- 12 NQF. Measuring performance. https://www.qualityforum.org/Measuring_Performance/Measuring_Performance.aspx. Last accessed February 2021.
- 13 Innovation Models | CMS Innovation Center. <https://innovation.cms.gov/innovation-models#views=models>. Last accessed January 2021.

- 14 Basch E, Spertus J, Dudley RA, et al. Methods for developing patient-reported outcome-based performance measures (PRO-PMs). *Value in Health*. 2015;18(4):493-504.
- 15 Gelb Safran D, Higgins A. Getting to the next generation of performance measures for value-based payment. Health Affairs Blog. <https://www.healthaffairs.org/doi/10.1377/hblog20190128.477681/full/>. Last accessed December 2020.
- 16 Squitieri L, Bozic KJ, Pusic AL. The role of patient-reported outcome measures in value-based payment reform. *Value Health*. 2017;20(6):834-836.
- 17 CMS. *CMS Measures Management System Blueprint*. ; 2020:97. <https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/MMS/Downloads/Blueprint.pdf>.
- 18 Meaningful Measures 2.0: Moving from measure reduction to modernization. CMS. <https://www.cms.gov/meaningful-measures-20-moving-measure-reduction-modernization>. Last accessed December 2020.
- 19 Chee TT, Ryan AM, Wasfy JH, et al. Current state of value-based purchasing programs. *Circulation*. 2016;133(22):2197-2205.
- 20 Hospital value-based purchasing | CMS. <https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/Value-Based-Programs/HVBP/Hospital-Value-Based-Purchasing>. Last accessed December 2020.
- 21 Advanced APMs - QPP. <https://qpp.cms.gov/apms/advanced-apms>. Last accessed December 2020.
- 22 Background of the PRO-CTCAE. <https://healthcaredelivery.cancer.gov/pro-ctcae/background.html>. Last accessed December 2020.
- 23 Development, testing, and implementation of the PRO-CTCAE Measurement System. Development, testing, and implementation. <https://healthcaredelivery.cancer.gov/pro-ctcae/testing.html>. Published April 12, 2021.
- 24 Patient-Reported Outcomes Measurement Information System (PROMIS). National Institute on Aging. <http://www.nia.nih.gov/research/resource/patient-reported-outcomes-measurement-information-system-promis>. Last accessed December 2020.
- 25 Spertus JA, Jones PG. Development and validation of a short version of the Kansas City Cardiomyopathy Questionnaire. *Circulation: Cardiovascular Quality and Outcomes*. 2015;8(5):469-476.
- 26 Bilbao A, Escobar A, García-Perez L, et al. The Minnesota Living With Heart Failure questionnaire: comparison of different factor structures. *Health Qual Life Outcomes*. 2016;14(1):23.
- 27 Yee D, Novak E, Platts A, et al. Comparison of the Kansas City Cardiomyopathy Questionnaire and Minnesota Living With Heart Failure Questionnaire in predicting heart failure outcomes. *The American Journal of Cardiology*. 2019;123(5):807-812.
- 28 Gagnier JJ. Patient reported outcomes in orthopaedics. *Journal of Orthopaedic Research*. 2017;35(10):2098-2108.

- 29 Hospital for Special Surgery (HSS). Osteoarthritis & joint replacement outcome surveys. Hospital for Special Surgery. <https://www.hss.edu/hoos-jr-koos-jr-outcomes-surveys.asp>. Last accessed December 2020.
- 30 Polascik BA, Hidaka C, Thompson MC, et al. Crosswalks Between Knee and Hip Arthroplasty Short Forms: HOOS/KOOS JR and Oxford. *J Bone Joint Surg Am*. 2020;102(11):983-990.
- 31 Lyman S, Lee Y-Y, Franklin PD, et al. Validation of the HOOS, JR: A short-form hip replacement survey. *Clin Orthop Relat Res*. 2016;474(6):1472-1482.
- 32 Lyman S, Lee Y-Y, Franklin PD, et al. Validation of the KOOS, JR: A short-form knee arthroplasty outcomes survey. *Clin Orthop Relat Res*. 2016;474(6):1461-1471.
- 33 Comprehensive Care for Joint Replacement Model | CMS Innovation Center. <https://innovation.cms.gov/innovation-models/cjr>. Last accessed January 2021.
- 34 Rose AJ, Bayliss E, Baseman L, et al. Feasibility of Distinguishing Performance Among Provider Groups Using Patient-reported Outcome Measures in Older Adults With Multiple Chronic Conditions: *Medical Care*. 2019;57(3):180-186.
- 35 ICHOM | About Us | Our Mission | History | Michael Porter. *ICHOM – International Consortium for Health Outcomes Measurement*. <https://www.ichom.org/mission/>. Last accessed December 2020.
- 36 Hip & knee osteoarthritis data collection reference guide. International Consortium for Health Outcomes Measurement (ICHOM). <https://ichom.org/files/medical-conditions/hip-knee-osteoarthritis/hip-knee-osteoarthritis-reference-guide.pdf>. Published 2015.
- 37 Patient Reported Outcome Measures. <https://www.aaos.org/quality/research-resources/patient-reported-outcome-measures/upper-extremity-proms/>. Last accessed December 2020.
- 38 Patient Reported Outcome Measures. <https://www.aaos.org/quality/research-resources/patient-reported-outcome-measures/lower-extremity-performance-measures/>. Last accessed December 2020.
- 39 Patient Reported Outcome Measures. <https://www.aaos.org/quality/research-resources/patient-reported-outcome-measures/spine/>. Last accessed December 2020.
- 40 Patient Reported Outcome Measures. <https://www.aaos.org/quality/research-resources/patient-reported-outcome-measures/treatment-outcome-quality-of-life/>. Last accessed December 2020.
- 41 SGO Calls on Clinicians to Measure Patient-Reported Outcomes Routinely. Mass General Advances in Motion. <https://advances.massgeneral.org/obgyn/journal.aspx?id=1695>. Last accessed December 2020.
- 42 Sisodia RC, Dewdney SB, Fader AN, et al. Patient reported outcomes measures in gynecologic oncology: A primer for clinical use, part I. *Gynecologic Oncology*. 2020;158(1):194-200.
- 43 Patient Reported Outcomes. <https://www.aan.com/policy-and-guidelines/quality/quality-measures2/patient-reported-outcome-pro/>. Last accessed December 2020.
- 44 CMS. *Patient Reported Outcome Measures*.; 2020:6. <https://www.cms.gov/files/document/blueprint-patient-reported-outcome-measures.pdf>.

- 45 Nelson E, Eftimovska E, Lind C, et al. Dartmouth Institute for Health Policy and Clinical Practice. Peer reviewed technical report: using patient-reported information to improve health outcomes and health care value: case studies from Dartmouth, Karolinska, and Group Health. *BMJ*. 2012;350:g7818-g7818.
- 46 Gausden EB, Levack A, Nwachukwu BU, et al. Computerized Adaptive Testing for patient reported outcomes in ankle fracture surgery. *Foot Ankle Int*. 2018;39(10):1192-1198.
- 47 NQF. 2017 Consensus Development Process redesign. National Quality Forum. https://www.qualityforum.org/docs/2017_Kaizen__CDP_Redesign_Report.aspx.
- 48 NQF. eMeasure Trial Use Program. https://www.qualityforum.org/eMeasure_Trial_Use.aspx. Last accessed February 2021.
- 49 NQF: Scientific Methods Panel. https://www.qualityforum.org/Measuring_Performance/Scientific_Methods_Panel.aspx. Last accessed December 2020.
- 50 NQF: Consensus Development Process. https://www.qualityforum.org/Measuring_Performance/Consensus_Development_Process.aspx. Last accessed December 2020.
- 51 NQF. Intent to submit. http://www.qualityforum.org/Measuring_Performance/Consensus_Development_Process_s_Principle/Intent_to_Submit.aspx. Last accessed January 2021.
- 52 NQF. Call for nominations. http://www.qualityforum.org/Measuring_Performance/Consensus_Development_Process_s_Principle/Call_for_Nominations.aspx. Last accessed January 2021.
- 53 CIRCULAR NO. A-119 Revised: Federal participation in the development and use of voluntary consensus standards and in conformity assessment activities. <https://www.whitehouse.gov/wp-content/uploads/2017/11/Circular-119-1.pdf>. Published February 10, 1998.
- 54 NQF. Measure review. http://www.qualityforum.org/Measuring_Performance/Consensus_Development_Process_s_Principle/Measure_Review.aspx. Last accessed January 2021.
- 55 NQF. Public commenting with member support. http://www.qualityforum.org/Measuring_Performance/Consensus_Development_Process_s_Principle/Public_Commenting_with_Member_Support.aspx. Last accessed January 2021.
- 56 NQF. Measure endorsement. http://www.qualityforum.org/Measuring_Performance/Consensus_Development_Process/CSAC_Decision.aspx. Last accessed January 2021.
- 57 NQF: Consensus Standards Approval Committee. https://www.qualityforum.org/About_NQF/CSAC/Consensus_Standards_Approval_Committee.aspx. Last accessed December 2020.

- 58 NQF. Measure appeals. http://www.qualityforum.org/Measuring_Performance/Consensus_Development_Process_s_Principle/Measure_Appeals.aspx. Last accessed January 2021.
- 59 NQF. Measure evaluation criteria and guidance for evaluating measures for endorsement. National Quality Forum. <https://www.qualityforum.org/WorkArea/linkit.aspx?LinkIdentifier=id&ItemID=92804>.
- 60 Fairclough DL, Peterson HF, Cella D, et al. Comparison of several model-based methods for analysing incomplete quality of life data in cancer clinical trials. *Statistics in Medicine*. 1998;17(5-7):781-796.
- 61 Brundage M, Bass B, Davidson J, et al. Patterns of reporting health-related quality of life outcomes in randomized clinical trials: implications for clinicians and quality of life researchers. *Quality of Life Research*. 2011;20(5):653-664.
- 62 NQF. *Patient-Reported Outcomes: Approaches and Challenges in Selection and Data Collection - An Environmental Scan*.; 2019. http://www.qualityforum.org/Publications/2019/12/Patient-Reported_Outcomes_Environmental_Scan.aspx. Last accessed December 2020.
- 63 CMS. Promoting Interoperability Programs | CMS. <https://www.cms.gov/Regulations-and-Guidance/Legislation/EHRIncentivePrograms>. Last accessed February 2021.
- 64 Health Level 7 International. FAQs | HL7 International. <http://www.hl7.org/about/faqs/index.cfm#A1>. Last accessed February 2021.
- 65 Health Level 7 International. Overview - FHIR v4.0.1. <https://www.hl7.org/fhir/overview.html>. Last accessed February 2021.
- 66 Hauser RG, Quine DB, Ryder A, et al. Unit conversions between LOINC codes. *Journal of the American Medical Informatics Association*. 2018;25(2):192-196.
- 67 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.
- 68 Vreeman DJ, McDonald CJ, Huff SM. Representing patient assessments in LOINC®. *AMIA Annu Symp Proc*. 2010;2010:832-836.
- 69 LOINC. Survey instruments archives. *LOINC*. <https://loinc.org/panels/category/survey-instruments/>. Last accessed February 2021.
- 70 Benson T. Measure what we want: a taxonomy of short generic person-reported outcome and experience measures (PROMs and PREMs). *BMJ Open Qual*. 2020;9(1):e000789.

Appendices

Appendix A: Attribute Grid Example

The following Attribute Grid was developed by the TEP and published in the Final Technical Report for the CMS-funded initiative titled *Patient-Reported Outcomes: Best Practices on Selection and Data Collection*. The Attribute Grid was presented as a tool to aid in the comparison and selection of PROMs for use in clinical settings.

ATTRIBUTE	PROM 1	PROM 2	PROM 3	PROM 4
Covers desired PROs:				
Covers desired PROs:				
Covers desired PROs:				
Covers desired PROs:				
Covers desired PROs:				
Contains goal attainment and goal attainment follow-up questions				
Symptoms				
Impacts				
Costs/fees				
Languages/translations available				
Length (number of items)				
Psychometric soundness: burden, including time and effort				
Psychometric soundness: clear, conceptual, and measurement models	Concepts included:	Concepts included:	Concepts included:	Concepts included:
Clinical applicability to desired population	Intended population:	Intended population:	Intended population:	Intended population:
Psychometric soundness: reliability (include sample size, various estimates if provided, and applicable population(s))	Test-retest reliability:	Test-retest reliability:	Test-retest reliability:	Test-retest reliability:
Good, better, or best reliability	Internal Consistency (Cronbach's a):			

ATTRIBUTE	PROM 1	PROM 2	PROM 3	PROM 4
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):			
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 B: Search Terms

To gain a broad understanding of literature related to PRO-PMs and existing guidance for developing PRO-PMs, various search terms were included within PubMed and Google Scholar queries. An initial PubMed search that incorporated potentially relevant Medical Subject Heading (MeSH) terms yielded zero results. Given that MeSH terms are not in existence for PROMs or PRO-PMs, general search terms and phrases were utilized. Such phrases included the following:

- “PROM”
- “PRO-PM”
- “Patient-Reported Outcome-Performance Measure” and “Patient-Reported Outcome-Performance Measures”
- “Patient-Reported Outcome Measure” and “Patient-Reported Outcome Measures”
- “Attributes of patient-reported outcome measures”
- “Development of patient-reported outcome measures”
- “PRO-PM guidance”
- “LOINC”
- “Patient assessments” and “LOINC”
- “HL7 FHIR”
- “Interoperability”

Terms also included specific searches for those PROMs referenced in Appendix C.

Appendix C: PROMs Reviewed for This Report

The following PROMs were reviewed as potential examples of high quality PROMs based on recommendations by CMS, documentation for CMS VBP programs and/or APMs, the TEP, and information identified during literature reviews. PROMs are linked to a homepage or developer site where possible.

- [Hip disability and Osteoarthritis Outcome Score, Joint Replacement](#) (HOOS, JR)
- [Kansas City Cardiomyopathy Questionnaire – 12 item](#) (KCCQ-12)
- [Knee injury and Osteoarthritis Outcome Score, Joint Replacement](#) (KOOS, JR)
- [Minnesota Living with Heart Failure Questionnaire](#) (MLHFQ)
- [Patient Health Questionnaire](#) (PHQ-9)
- [Patient-Reported Outcomes Measurement Information System](#) (PROMIS)
- [Patient Reported Outcomes version of the Common Terminology Criteria for Adverse Events](#) (PRO-CTCAE)

Appendix D: Reference Tables

Table 2: Current NQF-Endorsed PRO-PMs

Measure Title	NQF#	Measure Steward	Updated Date
CAHPS Clinician & Group Surveys (CG-CAHPS) Version 3.0-Adult, Child	0005	Agency for Healthcare Research and Quality	October 25, 2019
Consumer Assessment of Healthcare Providers and Systems (CAHPS) Health Plan Survey, Version 5.0 (Medicaid and Commercial)	0006	Agency for Healthcare Research and Quality	October 25, 2019
Experience of Care and Health Outcomes (ECHO) Survey (behavioral health, managed care versions)	0008	Agency for Healthcare Research and Quality	September 17, 2012
HCAHPS (Hospital Consumer Assessment of Healthcare Providers and Systems) Survey*	0166	Centers for Medicare & Medicaid Services	October 25, 2019
Comfortable Dying: Pain Brought to a Comfortable Level Within 48 Hours of Initial Assessment	0209	National Hospice and Palliative Care Organization	October 26, 2016
Consumer Assessment of Healthcare Providers and Systems (CAHPS) In-Center Hemodialysis Survey (ICH CAHPS)	0258	Centers for Medicare & Medicaid Services	October 25, 2019
Functional Status Change for Patients With Knee Impairments*	0422	Focus on Therapeutic Outcomes, Inc.	August 31, 2017
Functional Status Change for Patients With Hip Impairments*	0423	Focus on Therapeutic Outcomes, Inc.	August 31, 2017
Functional Status Change for Patients With Foot and Ankle Impairments*	0424	Focus on Therapeutic Outcomes, Inc.	August 31, 2017
Functional Status Change for Patients With Low Back Impairments	0425	Focus on Therapeutic Outcomes, Inc.	July 31, 2020

Measure Title	NQF#	Measure Steward	Updated Date
Functional Status Change for Patients With Shoulder Impairments*	0426	Focus on Therapeutic Outcomes, Inc.	August 31, 2017
Functional Status Change for Patients With Elbow, Wrist and Hand Impairments*	0427	Focus on Therapeutic Outcomes, Inc.	July 7, 2015
CAHPS® Home Health Care Survey (Experience With Care)	0517	Centers for Medicare & Medicaid Services	October 25, 2019
Depression Remission at 12 Months	0710e	MN Community Measurement	March 06, 2015
Depression Remission at Six Months	0711	MN Community Measurement	March 06, 2015
Consumer Assessment of Healthcare Providers and Systems (CAHPS)® Surgical Care Survey Version 2.0	1741	American College of Surgeons, Division of Advocacy and Health Policy	June 05, 2018
Depression Response at Six Months-Progress Towards Remission	1884	MN Community Measurement	February 08, 2016
Depression Response at 12 Months-Progress Towards Remission	1885	MN Community Measurement	October 26, 2016
Gains in Patient Activation (PAM) Scores at 12 Months	2483	Insignia Health	April 07, 2016
CoreQ: Short Stay Discharge Measure	2614	AHCA/NCAL	November 20, 2020
CoreQ: Long-Stay Resident Measure	2615	American Health Care Association	November 20, 2020
CoreQ: Long-Stay Family Measure	2616	AHCA/NCAL	November 20, 2020
Average Change in Functional Status Following Lumbar Spine Fusion Surgery	2643	MN Community Measurement	March 28, 2017
CAHPS® Hospice Survey (Experience With Care)	2651	Centers for Medicare & Medicaid Services	November 20, 2020
Average Change in Functional Status Following Total Knee Replacement Surgery	2653	MN Community Measurement	July 07, 2015
Adolescent Assessment of Preparation for Transition (ADAPT) to Adult-Focused Health Care	2789	Center of Excellence for Pediatric Quality Measurement	May 04, 2016
Informed, Patient-Centered (IPC) Hip and Knee Replacement Surgery	2958	Massachusetts General Hospital	October 25, 2016
Shared Decision Making Process	2962	Massachusetts General Hospital	October 25, 2016
CAHPS® Home and Community-Based Services Measures	2967	Centers for Medicare & Medicaid Services	November 14, 2017
CollaboRATE Shared Decision Making Score	3227	The Dartmouth Institute for Health Policy & Clinical Practice	October 25, 2019

Measure Title	NQF#	Measure Steward	Updated Date
CoreQ: AL Resident Satisfaction Measure	3420	American Health Care Association/National Center for Assisted Living	October 26, 2018
CoreQ: AL Family Satisfaction Measure	3422	American Health Care Association/National Center for Assisted Living	October 26, 2018
Functional Status Change for Patients With Neck Impairments	3461	Focus on Therapeutic Outcomes	October 25, 2019
Patient-Centered Contraceptive Counseling (PCCC)	3543	University of California, San Francisco	November 20, 2020
Hospital-Level, Risk-Standardized Patient-Reported Outcomes Following Elective Primary Total Hip and/or Total Knee Arthroplasty (THA/TKA)	3559	Centers for Medicare & Medicaid Services	November 20, 2020

* As of April 1, 2021, these measures are identified as NQF-endorsed PRO-PMs in the CMS Measures Inventory Tool but are listed with a measure type of “Outcome” rather than “Outcome: PRO-PM” in the NQF Quality Position System.

Table 3: Best Practices and Considerations

The table below consists of methodological best practices and associated considerations for developing and evaluating proposed PRO-PMs, as identified by Basch et al.¹⁴

Best Practice	Considerations
A rationale for measuring the outcome should be described.	<p>Is a knowledge gap described and justified?</p> <p>Is there evidence that the outcome is meaningful and important to patients, caregivers, and/or other stakeholders?</p> <p>How does patient self-reporting, in particular, address the gap?</p> <p>Are patients the most appropriate source of information?</p>
The intended context of use should be described and justified.	<p>Is the intended context of use clearly described and justified?</p> <p>How is information from the measure expected to inform change in practice to improve performance in the intended context of use?</p> <p>How will the nominated measure complement other measures to improve understanding of performance in the intended context of use?</p>

Best Practice	Considerations
	Is there variability in the outcome at the practice or practitioner level?
The measure should be adequately developed for the intended context of use (or a similar context of use), including demonstration of meaningfulness and importance to patients as well as adequate psychometric properties.	<p>Is the underlying concept to be measured clearly identified (e.g., post-chemotherapy nausea)?</p> <p>Is there prior or planned qualitative work in a patient population similar to the intended context of use that demonstrates understanding of terminology and mapping of the terminology to the underlying concept(s) of interest?</p> <p>Is there evidence of adequate psychometric properties of the measure, including construct validity and reliability, meaningfulness of score changes in a comparable population, and reasonableness of the recall period?</p>
There should be prior or planned work using the measure in the intended context of use (or a similar context of use), demonstrating that it is sensitive to change and clinically actionable.	<p>Has the measure been shown to detect changes over time or differences between known patient groups, practices, and/or procedures?</p> <p>Does the measure detect change in clinical action(s)?</p> <p>Is there evidence that there is not a floor or ceiling effect of the measure in the intended context of use?</p>
There should be a recommended implementation strategy for the measure in the intended context of use.	<p>Is there a rationale for an administration mode (e.g., paper, electronic) and schedule (e.g., timing of follow-up evaluations)?</p> <p>Is there a plan to maximize recruitment and response rates (e.g., backup data collection plan for nonrespondents)?</p> <p>Is proxy or surrogate reporting considered allowable?</p> <p>Is there a plan to accurately identify patients in the target population and calculate the denominator (i.e., number of people who were asked to complete the measure)?</p>
There should be a recommended analysis plan, including a risk adjustment strategy, missing data approach, and power calculation.	<p>Is there a well-justified, <i>a priori</i> risk adjustment or stratification strategy based on evidence?</p> <p>Is there a plan to adjust analyses for case mix, recruitment bias, and response bias?</p>

Best Practice	Considerations
	<p>Is there a plan for imputing missing data with sensitivity analyses?</p> <p>What sample sizes are necessary for planned analyses?</p>
<p>There should be a recommended framework for interpreting results, including unit(s) of analysis and meaningful score thresholds.</p>	<p>What unit of analysis is recommended (e.g., hospital system, hospital, individual practice, individual practitioner, and patient-level)?</p> <p>What metrics should be used to reflect performance (e.g., proportion of patients achieving a specific score change, proportion of providers who are outliers)?</p> <p>How are the results of different PRO measures that may not agree with each other considered?</p>
<p>There should be a recommended approach for reporting and disseminating results.</p>	<p>Is there a suggested approach for packaging and presenting reports to practices, providers, and/or patients?</p>

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

Committee Members

Catherine MacLean, MD, PhD *(Co-Chair)*

Chief Value Medical Officer, Rheumatologist, Hospital for Special Surgery
New York, New York

Sam Simon, PhD *(Co-Chair)*

Director, Health Program Improvement, Mathematica
Cambridge, Massachusetts

David Andrews, PhD

Patient Advisor
Aiken, South Carolina

Katherine Ast, MSW, LCSW

Director, Quality & Research, American Academy of Hospice and Palliative Medicine
Glenview, Illinois

Rachel Brodie, BA

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

Zahid Butt, MD, FACP

President and Chief Executive Officer (CEO), Medisolv, Inc.
Columbia, Maryland

Collette Cole, BSN, RN, CPHQ

Clinical Measurement Developer, Minnesota Community Measurement
Minneapolis, Minnesota

Paula Farrell, BSN, RN, CPHQ, LSSGB

Associate Project Director, Clinical Quality Measurement, Department of Quality Measurement, The
Joint Commission
Oakbrook Terrace, Illinois

Mark Friedberg, MD, MPP

Senior Vice President, Performance Measurement & Improvement, Blue Cross Blue Shield of
Massachusetts
Boston, Massachusetts

Debbie Gipson, MD, MS

Professor of Pediatrics, University of Michigan
Ann Arbor, Michigan

Ben Hamlin, MPH

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

Janel Hanmer, MD, PhD

Medical Director of Patient Reported Outcomes, University of Pittsburgh Medical Center
Pittsburgh, Pennsylvania

Helen Haskell, MA

Founder and President, Mothers Against Medical Error
Columbia, South Carolina

Brian Hurley, MD, MBA, DFASAM

Director of Addiction Medicine, Los Angeles County Department of Health Services
Los Angeles, California

Christine Izui, MS

Health Policy Principle, Center for Transforming Health, MITRE Corporation
McLean, Virginia

Laura Jantos, LFHIMSS

Patient Advocate, Healthcare Technology/Digital Health
Mercer Island, Washington

Kirk Munsch

Patient Advocate, Rare Patient Voice
Colorado Springs, Colorado

Deborah Paone, DrPH, MHA

Performance Evaluation Lead and Policy Consultant, SNP Alliance

Washington, District of Columbia

Brenna Rabel, MPH

Lead Health Research Scientist, Battelle Memorial Institute
Arlington, Virginia

Nan Rothrock, PhD, MA

Clinical Psychologist, Feinberg School of Medicine, Northwestern University
Chicago, Illinois

Mike Sacca, AS

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

Rachel Sisodia, MD

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

John Spertus, MD, MPH, FACC, FAHA

Clinical Director of Cardiovascular Education and Outcomes Research, Mid America Heart Institute
Kansas City, Missouri

Ruth Wetta, PhD, MSN, MPH, RN

Lead Clinical Researcher, Population Health, Cerner Corporation
Kansas City, Missouri

Albert Wu, MD, MPH, FACP

Professor, Department of Health Policy and Management and Medicine, Johns Hopkins Bloomberg
School of Public Health
Baltimore, Maryland

Federal Liaisons

Joel Andress

CMS, Center for Clinical Standards and Quality

Janis Grady, RHIT, FAC-COR III

CMS, Center for Clinical Standards and Quality

Sandra Mitchell, PhD, CRNP, FAAN

National Institutes of Health, National Cancer Institute

Ashley Wilder Smith, PhD, MPH

National Institutes of Health, National Cancer Institute

Rhona Limcango, PhD

Agency for Healthcare Research and Quality

Meghan McHugh, PhD, MPH

Substance Abuse and Mental Health Services Administration

Clifford A. Smith, PhD, ABPP-Cn

US Department of Veterans Affairs, Office of Mental Health and Suicide Prevention

David Au, MD, MS

US Department of Veterans Affairs, Center of Innovation for Veteran-Centered and Value-Driven Care

Girma Alemu, MD, MPH

Health Resources and Services Administration, Office of Planning, Analysis, and Evaluation

NQF Staff

Sheri Winsper, RN, MSN, MSHA

Senior Vice President, Quality Measurement

Maha Taylor, MHA

Managing Director

Chuck Amos, MBA

Director

Teresa Brown, MHA, MA

Senior Manager

Bejier Edwards, MBA, PMP

Project Manager

Wei Chang, MPH

Analyst

Jhamiel Prince, MA

Analyst

Juanita Rogers, MS, CHES

Analyst

CMS Staff

LaWanda Burwell

Task Order COR

Sophia Chan

Senior Research Scientist

Patrick Wynne

IDIQCOR

Maria Durham

Director, DPMS/QMVIG/CCSQ

Helen Dollar-Maples

Deputy Director, DPMS/QMVIG/CCSQ

Appendix F: Public Comments

The draft environmental scan report was posted on the project webpage for public and NQF member comment from February 25, 2021 through March 17, 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 25 total comments from six 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.

From the perspective of patients, caregivers, and patient advocacy groups: Which outcomes matter most to you when considering performance measurement? (e.g., improved function, quality of life, and pain)

American Academy of Physical Medicine and Rehabilitation

COMMENT

AAPM&R strongly encourages the NQF to reach out to patients, caregivers and patient advocacy groups that are naïve to the PRO, PROM, PRO-PM terminology and engage in conversations to understand which outcomes matter to these groups. By providing a response "from the perspective" of such groups, it would be anecdotal and with a clinician bias. We do not believe the planned 12 hours or up to 9 key informant interviews will get NQF to a comprehensive answer for this question. Specialty Societies are struggling with aligning the type of data CMS wants, physicians want to inform treatment plans or condition management, and patients/caregivers want to understand if they are improving. Anecdotally, we hear that the types of outcomes that matter to patients are things like: when can I return to work, when will I be able to do "x" without pain, when can I return to activities of daily living without discomfort, when will I be able to sleep through the night? We feel strongly that performance measurement approaches for patient reported outcomes should balance the types of outcomes physicians/clinicians need to manage a patient's symptoms and those that are important to patients. Patients have to be part of the conversation since understanding what outcomes are achievable for their specific circumstances can help with treatment adherence and a care plan that is aligned between the provider and patient.

RESPONSE

We agree that engaging patients in this discussion is critical, and we will reflect patient voices not only through Key Informant Interviews but also through TEP representation, focus groups, and other means within the project scope. As we develop the Interim Report, we will consider opportunities and current limitations in balancing information that is meaningful to patients with information important to clinicians.

American Geriatrics Society

COMMENT

The American Geriatrics Society agrees that improved function and quality of life are important outcomes, as well as symptom control, and alignment with patient goals/whether patient goals were met.

There should be an explicit recognition of the importance of caregiver perspectives. This becomes obvious in patients with advanced dementia who may not be able to report their own outcomes: If PROs are important and some patients can't report outcomes, caregivers can provide the best insight into the patient perspective. Additionally, caregiver reported outcomes are important even when the patient is able to speak for themselves, if the caregiver is a critical element of the patient's well-being.

RESPONSE

Please see the “Challenges > Patient and Caregiver Data Challenges” section for new language about the implications of gathering data from caregivers, family members, and other proxies, and the importance these perspectives bring for patients who cannot complete their own questionnaires. While we agree with the importance of Caregiver-Reported Outcomes, that is beyond the scope of this report.

Patient-Reported Outcomes, Value & Experience (PROVE) Center, Department of Surgery, Brigham and Women’s Hospital

COMMENT

The purpose of performance measures is to distinguish varying degrees of quality of care among healthcare entities. Outcomes that ‘matter most’ from the patient perspective will vary across individuals and by the clinical conditions that affect them. Our focus should thus be less on the domain or focus of the PM (e.g., why or how would one prioritize improved function over pain?), and more about identifying outcomes that are actionable, can guide quality improvement, and vary sufficiently by entity to inform patients’ choice of care provider.

RESPONSE

This recommendation is particularly applicable to the Interim Report. As we develop the Interim Report, we will consider opportunities and current limitations in balancing information that is meaningful to patients with ensuring we consider actionability as we look at attributes of high quality PROMs.

Does the information in the scan appropriately represent different perspectives, such as patients, measure developers, health systems, and payers? If no, please provide feedback in improving perspective representation.

American Academy of Physical Medicine and Rehabilitation

COMMENT

See prior question. AAPM&R also feels there should be additional perspective from practicing clinicians or broader medical specialty societies. We agree there is a need for a roadmap to go from patient outcomes to PRO-PMs, however, also feel there needs to be greater transparency on what it means to select a “high quality PROM”. The environment scan lacks detail for actual implementation, and we look forward to next publications to gain knowledge in moving our own efforts from use of PROMs and collecting PROM data to moving into PRO-PM development. As noted previously, AAPM&R encourages NQF to seek additional input directly from the patient, caregiver and patient advocacy perspectives. We also feel perspectives from successful PRO-PM measure developers would be helpful to expand. It is our understanding that one of the most significant challenges is translating PROM scales into reliable, valid measure scores.

RESPONSE

These comments are most relevant to the development of the Interim Report and the Technical Guidance Report. We will use your recommendations to guide us in developing these reports.

American Geriatrics Society

COMMENT

The American Geriatrics Society believes there could be more inclusion of patients and caregivers. They are not explicitly included in the key informant interviews and it only says “patients and caregivers ... may also be considered” so they may not be included at all.

RESPONSE

We agree that engaging patients in this discussion is critical, and we will reflect patient voices not only through Key Informant Interviews but also through TEP representation, focus groups, and other means within the project scope.

PROVE Center, Department of Surgery, Brigham and Women’s Hospital

COMMENT

The report could include more findings from the patient perspective, specifically regarding patients’ views of the importance of PROMS and PRO-PMs. Although it is critical to promote patient-centeredness, PRO-PMs cannot exist without patients supplying data. It is essential to maintain awareness of the inherent burden to patients associated with increased use of PRO-PMs. It is incumbent on measure developers and health systems to ensure that burden is as minimal as possible and has visible payoff to the patient (e.g., through contributions to shared decision-making, enhanced clinical encounters, or publicly available and readily accessible information on provider performance). Patients should want to complete PROMs.

The report attempts to be very inclusive and accessible to a wide range of experience with respect to measure developers and also refers to helping identify funding opportunities for development work. However, measure development is rather narrowly confined to those individuals and organizations that have sufficient infrastructure and access to funds to perform this work. The report could be improved by discussing this issue and considering approaches to making measure development work feasible to a larger range of stakeholders.

RESPONSE

Please see the “Challenges > Patient and Caregiver Data Challenges” section for new language about common barriers that patients face with regards to completing PROMs (e.g., length of instrument, language, digital divide) and how these have a downstream impact on PRO-PMs. Please see the “Challenges > Other Challenges and Considerations” section for new language about addressing barriers of entry for organizations wishing to develop PRO-PMs.

When considering the development and use of PROM-based PRO-PMs, what important practical information is not addressed within the scan?

American Academy of Physical Medicine and Rehabilitation

COMMENT

As noted in the prior question, the information in the scan is cursory and does not provide the level of detail to move from identification of a PROM to use and development into a PRO-PM. As a medical specialty that crosses numerous conditions, we encounter a spectrum of PROMs in clinical practice. We would welcome additional guidance on which PROM is high or best quality for a specific symptom or condition, and then should be able to use that guidance (especially if published) to help our clinical members transition to PROMs that can then be used for PRO-PM development. Clinicians and health systems need to understand the incentives to moving to standardized data collection of patient-

reported outcomes that will in turn facilitate not only evidence generation, but PRO-PMs that are important, reliable, valid, feasible and usable. AAPM&R believes that success stories, such as with the PHQ-9 and PROMIS, are important but limited in scope across specialized condition treatment. We need published information to partner with clinicians and health systems to promote wide-scale adoption of PROMs.

RESPONSE

The environmental scan report is intended to describe the current state of using PROMs as the basis for PRO-PMs. Your request for publicly available guidance will be primarily addressed in the Interim Report, which will identify the attributes of high quality PROMs that are suitable for use as the basis for digital PRO-PMs.

American College of Medical Quality

COMMENT

The Scan completely skipped over Criterion 1: Importance to Measure and Report: Extent to which the specific measure focus is evidence-based, important to making significant gains in healthcare quality, and improving health outcomes for a specific high-impact aspect of healthcare where there is variation in or overall less-than-optimal performance. This is both surprising and very concerning. The initial process circa 2011 that NQF developed and adopted for evaluating evidence related measures going through the CDP has not been updated. The current process here is loose and very unscientific and requires significant updating. Measure developers typically skirt around this step with provision of a narrative and bibliographic references, but no structured or explicit scientific approach to evaluating the effectiveness and impact on actual health outcomes related to the measure in question. In the case of PRO-PMs, the evidence that is presented usually is only related to the PRO instrument and not the effectiveness of the actual use of the related PRO-PM in the field. Change scores in PRO scores can be statistically significant but clinically insignificant. Another concern is how/if/when the results of generating PRO scores are actually used at the point of care during clinician/patient interactions. From my own experience having had several spine and hip surgeries, while I have filled out endless PRO instruments related to these procedures, I have never had a discussion or been made aware of these scores by the clinicians who have done these procedures or been involved with post procedural physical therapy. Thank you.

RESPONSE

Criterion 1 is briefly discussed in the section, “Review of NQF PRO-PM Endorsement Process.” While this project helps to prepare measure developers for the NQF endorsement process, the endorsement process itself is not within the scope of this initiative. We will share your comments with the NQF leaders who oversee the endorsement process. We agree with your concerns about the inconsistent use of PROM scores at point-of-care, and the 2020 PRO Best Practices report discusses the importance of interpretability and communication in detail.

American College of Medical Quality

COMMENT

Per my last comment, to further illustrate my points, see a good example of “evidence due diligence” for asthma from 2014 as an excellent step in determining “fit for purpose” of choosing an appropriate PROM to develop and field test a PRO-PM derived from it. Patient-reported outcome measures for

asthma: a systematic review | npj Primary Care Respiratory Medicine (nature.com)
<http://dx.doi.org/10.1038/npjpcrm.2014.20> Thank you.

RESPONSE

Thank you for the additional information supporting your previous comment.

PROVE Center, Department of Surgery, Brigham and Women's Hospital

COMMENT

The quality of a PRO-PM is so closely tied to the PROM. If there are flaws in the selection of the PROM or its collection, then the PRO-PM will be flawed. The appropriateness of the PROM and integrity of the data are critical to a successful PRO-PM. Thus, it would be helpful if this scan included a brief description of the uptake or anticipated uptake and usefulness of the 2020 report "Patient-Reported Outcomes: Best Practices on Selection and Data Collection" It would be helpful to have further discussion about barriers to successful implementation, not just in the roll-out phase, but as a sustained activity that is fully integrated into the clinical workflow. It also may be useful to include a review of the implementation evaluation literature to provide a more complete picture of implementation challenges and successes. Additionally, the issues of non-response and proxy response are major threats to the validity of PRO-PMs. Some discussion of developing and promoting guidelines for addressing these issues in the data and/or setting standards for PRO-PMs would be helpful.

RESPONSE

Please see the "Challenges > Importance of Relationships Between PROMs and Performance Measures" section for new language that frames, at a high-level, the selection and implementation of PROMs and refers readers to the 2020 PRO Best Practices report. Please see the "Challenges > Patient and Caregiver Data Challenges" section for new language about the implications of gathering data from caregivers, family members, and other proxies, and the importance these perspectives bring for patients who cannot complete their own questionnaires, as well as the implication of non-responses. We agree that there is much work to be done on systemically addressing implementation challenges and NQF aims to be involved in future work on this topic.

The scan addresses interoperability, standardized codes, and other challenges that are particularly relevant for digital PRO-PMs. What additional challenges should be included in this section?

American Academy of Physical Medicine and Rehabilitation

COMMENT

AAPM&R agrees that interoperability and standardized data elements are especially relevant for digital PRO-PMs. We also agree that when there is a 1-1 translation from a PROM to PRO-PM, there is some inherent simplicity for taking data from the PROM and using for performance measurement. However, with many conditions, there may be a variety of quality PROM's and thus we are challenged with "endorsing" a specific PROM, often with little evidence other than it being more widely used than others. We also believe modality of PROM data collection can be a challenge. We appreciate the Attribute Grid and think it is a good starting place for comparing PROMs but does not provide how to understand if something is good or bad, or considerations for modality differences. When moving toward a PRO-PM, it is important to know how attempted modality impacts data completeness, burden of collection, data standardization, etc.

While there is some discussion of the PHQ-9 outcomes measures, there is not discussion on the challenges or even the opportunities of moving from the process of collecting a PROM to having enough data and PROM experience to move into PRO-PM development. AAPM&R would like to see more discussion and guidance on how measure developers could potential skip the “process” step and move to viable, endorsable, useable PRO-PMs without 20 years of PROM data history. It took the PHQ-9 measure developers years to get from a process measure to outcomes measures, this is similar to the pathway for the Patient Activation Measure (PAM). A huge challenge in PRO-PM measure development is having the amount of data and necessary testing to meet criteria established by CMS and the NQF for measure endorsement and use. Are there potential glidepaths that could assist developers in moving quicker to important, meaningful PRO-PMs?

RESPONSE

Please see the “Challenges > Patient and Caregiver Data Challenges” section for new language about modality differences, and we will continue to consider this as we develop the Interim Report. While streamlining the development process for PRO-PMs is not the primary goal of this project, it is an important consideration; please see the introductory text in the “Challenges > Challenges with NQF Endorsement” section for new language about endorsement challenges to note the barriers that PROM data collection can present to PRO-PM development.

PROVE Center, Department of Surgery, Brigham and Women’s Hospital**COMMENT**

The heterogeneity of IT sophistication across providers is a major challenge to the success of digital PRO-PMs. Ability to effectively participate in digital PROM collection and PRO-PM reporting will be quite variable and may result in poor representation of healthcare providers in traditionally underserved communities (e.g., rural communities, low SES, high minority). Similarly, not all EHRs are created equal. The report mentions CMS’ plans to align the EHR certification process with other CMS reporting requirements. This will help, but it would also be helpful to establish and support formal EHR vendor coordination and collaboration systems to maximize efficiency and facilitate scale and spread of digital solutions across vendors. Another challenge that is not addressed in this report is the proliferation of mobile apps for PROM data collection. Many of these apps are developed independently of the EHR, by separate vendors and for specific purposes. Not only does this pose a challenge for EHR integration for PRO-PM reporting, but also adds to the complexity and burden of PROM data collection for a given patient who may be asked to access multiple apps in addition to the health system’s patient portal to complete PROMs associated with their care.

RESPONSE

The implications of interoperability on underserved populations are an important point; while outside the scope of this project, it warrants inclusion in the work and we added language to the “Challenges > Patient and Caregiver Data Challenges” section of the scan; we will also carefully consider this point as we develop the Interim Report and the Technical Guidance Report. We also expanded the language in the above section to include add-on systems used to collect PROMs and the potential burden they create.

When considering the intersection between PRO-PMs and either value-based purchasing programs or CMS Alternative Payment Models, what pertinent information is missing from this report?

American Academy of Physical Medicine and Rehabilitation

COMMENT

AAPM&R believes there needs to be more transparency from CMS on their plan to transition from traditional MIPS to MIPS Value Pathways and how PRO-PMs will be used, and how CMMI intends to use PRO-PMs and attribute scores to reimbursement. AAPM&R is concerned with how PRO-PMs will be used in value-based purchasing programs and would like to learn more information.

RESPONSE

We will share your request with CMS and encourage the agency to provide additional information.

American College of Medical Quality

COMMENT

Please see my comments for the 3rd question, which has bearing on this question. In practice, members of committees involved with the CDP for submitted measures often conflate the evidence that is (typically) presented by a measure developer for the PRO Instrument and (sometimes) the PROM in question with an expectation that this can be translated directly into a PRO-PM without additional field testing. Also, the notion of “Importance” of the PROM usually gets mixed up by committee members who are voting with that of the use of a derived PRO-PM with intended use by CMS for public reporting and differential payment determinations. This becomes especially problematic when Criterion 1 (see page 13) is presented and evaluated via the CDP. NQF should spend a lot more time in this scan fleshing these issues out in more detail and consider refining the current Criterion 1 evidence evaluation process that has not been updated for the past 10 years.

RESPONSE

We appreciate hearing your suggestions on the CDP. While this project helps to prepare measure developers for the NQF endorsement process, the endorsement process itself is not within the scope of this initiative. We will share your comments with the NQF leaders who oversee the endorsement process.

PROVE Center, Department of Surgery, Brigham and Women’s Hospital

COMMENT

The issue of accountability is complicated by use of patient reports, as so much is outside the control of the clinician or health system. This can be addressed to some extent by selecting the appropriate PROM, but even with the ‘right’ PROM in place, there are many factors that can influence the extent of change that is observed, including logistical issues such as the timing and mode of data collection as well as factors not related to the provision of care such as disruptions in patients’ personal lives. Given the high stakes, this report could be improved by including some acknowledgement and discussion of these issues. For example, this discussion could propose criteria that should be met for a PRO-PM to be used for accountability. It would also be helpful to know more about CMS’ plans to transition to use of PRO-PMs in its programs, and how entities will be held accountable in the future.

RESPONSE

We added language to the “Challenges > Patient and Caregiver Data Challenges” section to address PROM completion and its downstream impacts. We will also ensure the scan includes links and/or references to key CMS guidance.

What general comments do you have that would improve the Environmental Scan?

American Academy of Physical Medicine and Rehabilitation

COMMENT

AAPM&R believes the environmental scan is a good first step and are looking forward the next reports to provide more detailed and usable guidance.

RESPONSE

Thank you for your support. We hope you will continue to stay engaged in this initiative.

American College of Medical Quality

COMMENT

It is apparent that some listed members of the Committee are “owners” of some listed PROM derived PRO-PM (e.g. HOOS/KOOS, KCCQ, IPC/Share Decision Making and perhaps others) or vendors of systems that are designed to collect and manage the data necessary to construct the PROMs (e.g. Medisolv). In the interest of transparency, NQF should consider making conflict of interest and relationship disclosures more explicitly and publicly available in this report.

RESPONSE

Please see the “Background > Previous Work in Patient-Reported Outcomes in Performance Measurement” section for new language that discusses the intentional inclusion of PROM and PRO-PM developers and/or stewards on the TEP.

PROVE Center, Department of Surgery, Brigham and Women’s Hospital

COMMENT

The report is a great start to developing a step-by-step roadmap for PRO-PM development and implementation. The assembled TEP represents a strong and diverse set of members. We look forward to future products from this group and welcome any opportunities to contribute to this important work.

RESPONSE

Thank you for your support. We hope you will continue to stay engaged in this initiative.

American College of Physicians

COMMENT

With regards to the challenges with NQF endorsement section, the ACP strongly believes that the rigor of evidence required for a PRO-PM or any outcome measure, for that matter, be no less rigorous than that for a process measure. We understand that the type of evidence may be different but strongly recommend that it go beyond the current guidance stating that “empirical data demonstrating a relationship between the outcome and at least one healthcare structure, process, intervention, or service.” We would argue that we need empirical data that demonstrates a relationship between the PRO and at least one healthcare structure, process, intervention, or service that is actionable by the

accountable entity. Furthermore, we believe NQF needs to rethink its approach to evidence overall given that the evidence taskforce report has not been updated since 2011. The measurement arena has evolved significantly since that time which may call into question the premise of different evidence requirements for different types of measures.

RESPONSE

While this project helps to prepare measure developers for the NQF endorsement process, the endorsement process itself is not within the scope of this initiative. We will share your comments with the NQF leaders who oversee the endorsement process.

National Cancer Institute

COMMENT

The National Quality Forum (NQF) has developed an Environmental Scan Report as a first step in a larger initiative focused on creating a guide to the establishment of patient-reported outcome performance measures. There are a few notable topics that we believe need to be addressed or expanded.

In evaluating best practices for identifying or developing high quality PRO measures of clinician/health system performance, several characteristics of the PROs need to be considered and more explicitly highlighted in the report. There is substantial interplay among these characteristics:

1. PRO content domains to be measured should be chosen explicitly to reflect different aspects of clinician or health system performance. These content domains include the endpoints they are capturing including **health status** [e.g., symptoms and functional status], **satisfaction, quality of life** [e.g., social function, emotional well-being, loneliness], **experiences of care** [e.g., perceived quality of communication, shared decision-making, timeliness of care], and **technical care quality** [falls risk discussed, medication reconciliation, actions taken on distress screening]
2. Measurement properties need to have been empirically evaluated and reported, including information about the appropriate covariates to be accounted for (e.g., case mix, care delivery setting, time since procedure/event/milestone [surgery, hospital discharge]). Measurement properties include measure reliability, validity, sensitivity, and responsiveness. Strong psychometric measurement properties need to be demonstrated for the instrument both overall, and specifically as a measure of clinician or health system performance.
3. The PRO measures of clinician and health system performance must be able to be reliably interpreted at the individual level. This is critical so that as the outcomes/change scores/differences from norm are combined and presented at the group-level, that those group-level interpretations are robust and actionable.
4. In the section of the Environmental Scan that lists endorsed measures, it should also list the intended purpose of the measure articulated by measurement developer and/or purpose/use currently supported by evidence. The table should highlight those measures for which there is evidence of their sensitivity, responsiveness and interpretability as clinician performance measures, not just as health outcome measures in research studies. Any information about validity and reliability must be interpreted along with essential covariates for proper interpretation as a clinician performance measure.

5. Candidate measures of clinician or health-system performance must be fit-for-purpose. Though PROs may have strong psychometric properties in some contexts, their use as clinician or health system performance measures requires that they be able to capture information relevant to the care being delivered. Thus, their utility and interpretability as indicators of performance in specific contexts should be demonstrated empirically. Determinations that a measure is fit-for purpose include but are not limited to the following elements:
 - a. Condition-specific (e.g., for health status) or care setting-specific (e.g., for experiences of care)
 - b. Able to capture domains/concepts that are salient
 - c. Feasible (i.e., measures must be sufficiently brief so as not to be overly burdensome, they must be available in multiple languages, written at an 8th grade reading level etc)
 - d. Interpretable (estimates of change or stability available to support interpretation at the individual patient level)
 - e. Meaningful and actionable for clinicians

Second, throughout the Environmental Scan report, but especially in the Limitations section, there are some overly broad statements about measures being endorsed by organizations. It is unclear whether the endorsements represent endorsement of these measures for research, clinical care, clinician performance or something else. Some of the statements are extremely general and subtly mis-represent the state of the science around the testing of these measures as indicators of clinician performance and their state of adoption as performance measures. It is critical that any endorsements reflect PRO use as performance measures, rather than their use solely as self-report measures that capture health information.

RESPONSE

The environmental scan report is intended to describe the current state of using PROMs as the basis for PRO-PMs. Several of your suggestions align with work that will be addressed in the Interim Report and/or the Technical Guidance Report and we will consider these comments (including the importance of measuring different PRO content domains) in the upcoming stages of work. While we agree with your comment about the table of endorsed measures (#4 above), the risk of adding static information to the table is it will quickly become outdated; instead, the table includes links to additional details. Please see the “Challenges > Importance of Relationships Between PROMs and Performance Measures” section for new language that frames, at a high-level, the selection and implementation of PROMs and refers readers to the 2020 PRO Best Practices report, which addresses these topics in detail. Additionally, we have clarified instances of overly generally language, including the wording in the “Limitations” section.