Selecting patient-reported outcome measures: "what" and "for whom"

Jason B. Liu[1,2,3] , **Nan E. Rothrock**[4] , **Maria O. Edelen**[1,3],*

[1]Patient-Reported Outcomes, Value, and Experience (PROVE) Center, Brigham and Women's Hospital, Boston, MA 02115, United States
[2]Division of Surgical Oncology, Department of Surgery, Brigham and Women's Hospital, Boston, MA 02115, United States
[3]Harvard Medical School, Boston, MA 02115, United States
[4]Department of Medical Social Sciences, Northwestern University Feinberg School of Medicine, Chicago, IL 60611, United States
*Corresponding author: Department of Surgery, Brigham and Women's Hospital, Boston, MA 02115, United States. Email: medelen@bwh.harvard.edu

Abstract

Patient-reported outcomes (PROMs) are becoming more widely implemented across health care for important reasons. However, with thousands of PROMs available and the science of psychometrics becoming more widely applied in health measurement, choosing the right ones to implement can be puzzling. This article provides a framework of the different types of PROMs by organizing them into 4 categories based upon "what" is being measured and "from whom" the questions are asked: (1) condition-specific and domain-specific, (2) condition-specific and global, (3) universal and global, and (4) universal and domain-specific. We delve deeper into each category with clinical examples. This framework can empower health care leaders and policymakers to make more informed decisions when selecting the best PROMs to implement, ensuring PROMs deliver on their potential to promote high quality, patient-centered care.

Key words: patient-reported outcomes; PROMs; health outcomes; quality of care; value-based health care; implementation; patient-centered care.

Introduction

The fundamental goals of health care are to save and improve lives. Whether the latter has been successfully accomplished can be challenging to measure because these patient-centered outcomes often concern health-related quality of life (HRQL), defined as an individual's perception of how an illness and its treatment affect the physical, mental, and social aspects of their life.[1] Important aspects of health contributing to HRQL, including symptoms, functional status, and health perceptions, are often unobservable and known only to patients. Patients must, therefore, tell us whether they have experienced these outcomes.[2]

Based in psychometric science, patient-reported outcome measures (PROMs) are standardized questionnaires that contain multiple questions, or items, patients answer on their own to generate numerical scores measuring symptoms, function, perceived health status, and other important unobservable aspects of HRQL. At the individual level, PROMs can support patient-centered clinical care and research.[3] Aggregated at higher levels, PROMs can spur performance improvement, inform health policymaking, and drive value-based health care.[4,5] Unsurprisingly, PROMs are increasingly incorporated into many health care ventures.[6,7]

A critical step to implementing PROMs is choosing what to implement, which requires an understanding of what outcomes PROMs can measure and from whom. The chosen PROMs should measure the most relevant outcomes and support programmatic goals. Implementing inappropriate PROMs will yield unusable data, wasting time, energy, and health care resources. To help health care leaders, clinicians, and policymakers make more informed decisions, we provide, in what follows, a framework of the different types of PROMs available, with recent example clinical applications.

Four categories of PROMs along 2 dimensions

Patient-reported outcome measures can be conceptualized along 2 dimensions: specificity of outcome, or "what" is being measured, and specificity of patient group, or "from whom" the questions are asked.

For the "what" dimension, PROMs that ask questions about a specific aspect of health, such as fatigue, physical function, or the interference of pain on a patient's life, are called "domain-specific." In contrast, PROMs that ask questions about either general or overall health (eg, In general, how would you rate your physical health?) or generate scores summarizing multiple separate domains together are called "global."

A single PROM can be psychometrically designed to be both domain-specific by generating separate scores for multiple specific domains and global by combining multiple domains together to generate summary scores. These PROMs are sometimes called profiles or batteries. For example, the Patient Reported Outcome Measurement Information System (PROMIS)-29[8] and the Medical Outcomes Study 36-item Short Form Health Survey (SF-36)[9] are profile measures that generate scores for individual domains, such as physical function, anxiety, and sleep disturbance, but can also produce summary scores for overall physical and mental health. For patients with hip problems, the Hip
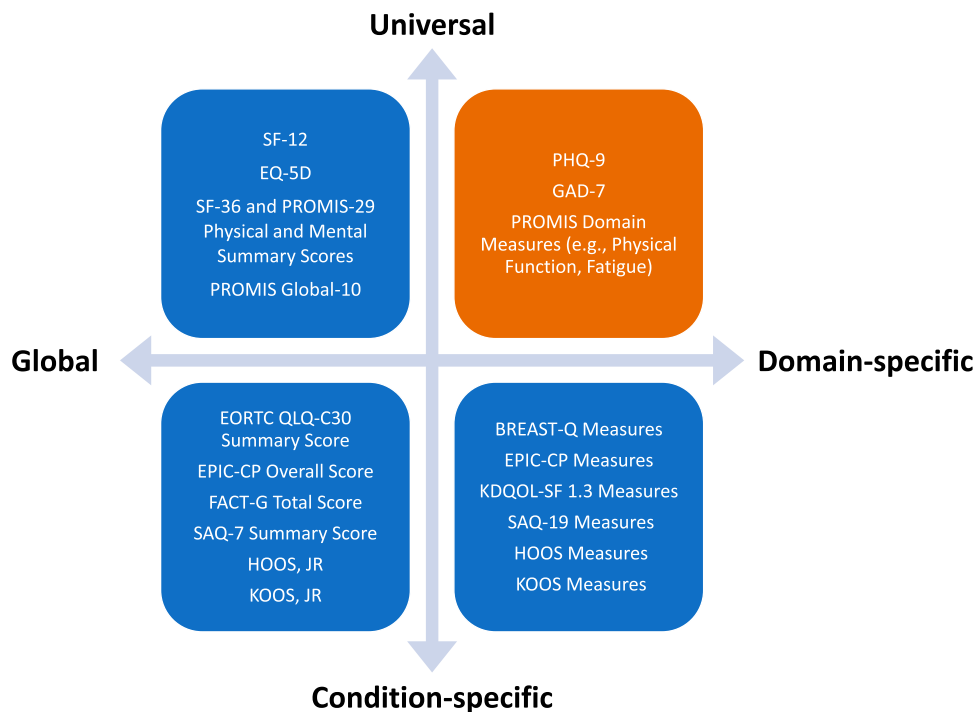
Figure 1. Four categories of PROMs based on what outcomes are measured and from which patients. Because a PROM can be psychometrically designed to generate scores for individual domains and summary scores by combining domains, it may appear to belong in multiple categories. More precisely, each score that a PROM generates belongs to only 1 of these 4 categories. Knowing how the PROM is scored can help clarify what is being measured. Abbreviations: EORTC QLQ-C30, European Organization for Research and Treatment of Cancer Quality of Life Group Core Questionnaire; EPIC-CP, Expanded Prostate Cancer Index Composite for Clinical Practice; FACT-G, Functional Assessment of Cancer Therapy–General; GAD-7, Generalized Anxiety Disorder-7; HOOS, JR, Hip Dysfunction and Osteoarthritis Outcome Score for Joint Replacement; KDQOL-SF, Kidney Disease Quality of Life–Short Form; KOOS, JR, Knee Injury and Osteoarthritis Outcome Score for Joint Replacement; PHQ-9, 9-item Patient Health Questionnaire; PROM, patient-reported outcome measure; PROMIS, Patient-Reported Outcomes Measurement Information System; SAQ-7, Seattle Angina Questionnaire—7 items; SAQ-19, Seattle Angina Questionnaire—19 items; SF-12, 12-item Short Form Health Survey; SF-36, 36-item Short Form Survey.

Disability and Osteoarthritis Outcome Score (HOOS)[10] is another PROM that generates both domain-specific (eg, pain, symptoms and stiffness, function in sports and recreational activities) and global (ie, total hip disability) scores.

The "from whom" dimension similarly has 2 categories: PROMs intended to measure aspects of health from the general population or across patient groups regardless of their condition are called "universal" or "generic," whereas those intended to measure aspects of health from specific patient groups with a particular condition or disease, such as dementia or breast cancer, are called "condition-" or "disease-specific." Nearly every HRQL domain can be evaluated from either a universal or condition-specific perspective depending on the wording of the items. For example, patients can answer questions about universal (or generic) pain interference (eg, pain interfered with my ability to sleep) or condition-specific pain interference (eg, low back pain interfered with my ability to sleep). Reading the items of a PROM can help clarify what is being measured and from whom.

Based on the cross-classification of these 2 dimensions, PROM scores fall into 1 of 4 categories described below (Figure 1).

Category 1: condition-specific and domain-specific
Most currently available PROMs used in clinical care are condition-specific and domain-specific. These PROMs are designed to measure relevant domains (ie, symptoms, functional status, and health perspectives) for a particular patient population.

For example, in 2007, the United Kingdom conducted the National Mastectomy and Breast Reconstruction Audit to understand determinants and outcomes of care for women with breast cancer having a mastectomy with or without breast reconstruction.[11,12] They chose the BREAST-Q, a collection of PROMs designed to measure domains relevant to patients with breast cancer undergoing reconstructive breast surgery, such as Satisfaction with Breasts and Sexual Well-being, using separate measures for each domain. The results of the audit not only provided evidence for the positive effects of breast reconstruction on patients' HRQL but also that variation in hospital performance could be detected, opening avenues for performance improvement. Continued research with the BREAST-Q PROMs has defined reference values for scores to identify which patients may benefit at the point of care from additional intervention,[13] much like how laboratory tests have reference ranges for normal.

Clinicians view PROMs in this category to be the most relevant for use in clinical care because they address specific issues for patients of interest to them. However, the quality of their development and the potential loss of comparability are major considerations. Many condition- and domain-specific PROMs exist, and more are constantly developed, each with slight modifications. Because these PROMs are often tailored to satisfy a particular patient population and research agenda, they can be hastily developed and thus suffer from less psychometric rigor. Moreover, they are, by definition, limited in scope, hindering broad clinical use and comparisons outside of the specific patient group.[14] With more than one-third of US

adults living with multiple chronic conditions,[15] the clinical desire for specificity can backfire—patients may find it difficult to attribute what they are experiencing with the specificity expected by clinicians.

Category 2: condition-specific and global

In contrast to the BREAST-Q, which produces scores for multiple domains, some PROMs provide a total score reflecting overall HRQL for a particular condition or disease. These condition-specific and global PROMs represent a 30 000-foot view of HRQL for a particular patient population and often do so by combining scores from multiple related domains.

The recently enacted Centers for Medicare and Medicaid Services (CMS) PRO-based performance measure (PRO-PM) for total hip and knee replacement surgeries[16] uses 2 condition-specific and global PROMs: the Hip Dysfunction and Osteoarthritis Outcome Score for Joint Replacement (HOOS, JR) and the Knee Injury and Osteoarthritis Outcome Score for Joint Replacement (KOOS, JR). The HOOS, JR uses 6 questions about pain and physical function that are summed to provide a global score, ranging from 0 to 100, where 0 means total hip disability and 100 perfect hip health. Similarly, the KOOS, JR uses 7 questions to provide a global score representing total knee disability to perfect knee health. The CMS will use risk-adjusted, hospital-level changes in HOOS, JR and KOOS, JR scores from before to 1 year after hip or knee replacement surgery to adjust payment and publicly report results.

A condition-specific and global approach can ensure the PROM content is important to patients and clinicians while maintaining brevity and practicality for implementation. However, a single global score is analogous to a composite quality measure, which can be challenging to inform action without understanding the contribution of its constituent components.[17,18] A hospital seeking to improve on the CMS PRO-PM for total joint replacement surgery, for instance, may find it difficult to discern whether to invest in more pain-management services or strengthen their physical therapy referral network.

Category 3: universal and global

Patient-reported outcome measures in this category provide a high-level view of HRQL for a wide range of patients.

The Veterans RAND 12-item Health Survey (VR-12), a variant of the Medical Outcomes Study 12-item Short Form Health Survey (SF-12), is a universal and global PROM that generates scores for overall physical and mental health. The CMS has administered the VR-12 as part of the Medicare Health Outcomes Survey (HOS) since 1998.[19] Each year, a new cohort of Medicare Advantage beneficiaries complete the HOS, and each cohort is resurveyed in 2 years. The data are used for public reporting, to assess care quality, to target beneficiaries at heightened risk for poor health outcomes, and to inform policymaking.[6,20] The rich HOS data have also been linked with the National Cancer Institute's Surveillance, Epidemiology, and End Results (SEER) cancer registry data to foster HRQL-related research in patients with cancer.[21]

Another popular universal and global PROM is the PROMIS Global-10 measure.[22] With only 10 questions, it provides a summary score for patients' physical and mental

HRQL regardless of their medical conditions and is understandably quite attractive because of its brevity. Although useful for some purposes,[23,24] like surveillance, the global view on any (universal) patient's physical or mental HRQL cannot directly inform clinically actionable targets. Instead, more domain specificity is needed, which can, in fact, be obtained without substantially increasing the number of questions that patients must answer if a PROM built with modern measurement theory is selected.[25,26]

Category 4: universal and domain-specific

Universal and domain-specific PROMs are intended for all patients regardless of their specific condition by measuring aspects of health that are shared among them. The measurement capabilities of these PROMs allow them to overcome many of the shortcomings noted for the other 3 PROM types.

One of the most popular universal and domain-specific PROMs used in clinical care, particularly primary care, is the 9-item Patient Health Questionnaire (PHQ-9).[27] The PHQ-9 is based on the 9 DSM-IV (*Diagnostic and Statistical Manual of Mental Disorders, fourth edition*) criteria to diagnose depressive disorder and can be used to screen for likely depressive disorder. It is scored from 0 to 27, with higher scores reflecting more severe depressive symptoms. Clinical thresholds for depression severity occur every 5 points (ie, ≥ 5 mild, ≥ 10 moderate, etc) and a meaningful clinical improvement is a decrease of 5 points.[28] The PHQ-9 has also been incorporated into quality measures[29] and used in innovative interventions to address social determinants of health.[30]

As depression is relevant to all patients, many other domains are also relevant to many different patients. For example, the interference of pain on patients' daily activities is an important domain universally relevant to all patients experiencing pain. Fatigue is another domain relevant to all patients regardless of their medical condition or number of different conditions.

Universal and domain-specific PROMs are cross-cutting and allow for valid comparisons of disease burden and treatment impact across different patient populations and conditions by generating scores on the same "ruler." When designed using modern measurement theory (eg, PROMIS),[31] these PROMs can be efficient and flexible. They can be readily applied to specific conditions and maintain responsiveness to condition-specific treatments.[32,33] Most important, the domain specificity provides clinically meaningful and actionable information.

Conclusion

Implementing the routine collection of PROMs can be challenging. With thousands of PROMs available, 1 major challenge is identifying the most relevant PROMs to implement. In this article, we provided a framework of how PROMs are categorized by understanding "what" they measure and "from whom."

Of course, identifying the type of PROM, what to measure, and from whom, is only 1 step. It is also critical to ensure that the PROMs chosen are of high quality, measuring what they are intended to measure and can do so repeatedly without error. Numerous reference resources are available to help determine whether the measurement properties of a given PROM—that is, its validity, reliability, and responsiveness—are of sufficient quality for the intended purpose of the

measurement.[34-36] The Consensus-based Standards for the Selection of Health Measurement Instruments (COSMIN), for example, provides guidance on how to rigorously evaluate PROM quality and provides a freely available repository of such evaluations for future reference.[37] Next, considerations for implementation must be determined, such as measurement cadence, mode of collection, score interpretation, and workflow integration. These challenges are surmountable with thoughtful planning and perseverance.[34,38-42] Certainly, inviting a psychometrician to join the implementation team can also be incredibly valuable.

Patient-reported outcome measures are not merely a health care trend—they stand to be further embedded in health policy as stakeholder momentum builds. Advances in health information technology and psychometric rigor have now made the ability to measure HRQL with PROMs not only possible but remarkably accurate and meaningful. However, health care leaders and other stakeholders seeking to enhance patient-centered care must implement PROMs with well-informed decisions. Only with thoughtful selection and integration can PROMs unlock the potential to align health care delivery more closely with what matters most to patients.

Supplementary material

Supplementary material is available at *Health Affairs Scholar* online.

Conflicts of interest

Please see ICMJE form(s) for author conflicts of interest. These have been provided as supplementary materials.

Notes

1. Wilson IB, Cleary PD. Linking clinical variables with health-related quality of life. A conceptual model of patient outcomes. *JAMA*. 1995;273(1):59-65.
2. Kaplan RS, Jehi L, Ko CY, Pusic A, Witkowski M. Health care measurements that improve patient outcomes. *NEJM Catal*. 2021;2(2):CAT.20.0527. https://doi.org/10.1056/CAT.20.0527
3. Lavallee DC, Chenok KE, Love RM, et al. Incorporating patient-reported outcomes into health care to engage patients and enhance care. *Health Aff (Millwood)*. 2016;35(4):575-582. https://doi.org/10.1377/hlthaff.2015.1362
4. Van Der Wees PJ, Nijhuis-Van Der Sanden Mg, Ayanian JZ, Black N, Westert GP, Schneider EC. Integrating the use of patient-reported outcomes for both clinical practice and performance measurement: views of experts from 3 countries. *Milbank Q*. 2014;92(4):754-775. https://doi.org/10.1111/1468-0009.12091
5. Calvert M, Kyte D, Price G, Valderas JM, Hjollund NH. Maximising the impact of patient reported outcome assessment for patients and society. *BMJ*. 2019;364:k5267. https://doi.org/10.1136/bmj.k5267
6. Hanmer J, Cizik AM, Gulek BG, et al. A scoping review of US insurers' use of patient-reported outcomes. *Am J Manag Care*. 2022;28(6):e232-e238. https://doi.org/10.37765/ajmc.2022.89162
7. Jacobs DB, Schreiber M, Seshamani M, Tsai D, Fowler E, Fleisher LA. Aligning quality measures across CMS — the universal foundation. *N Engl J Med*. 2023;388(9):776-779. https://doi.org/10.1056/NEJMp2215539
8. Hays RD, Herman PM, Rodriguez A, Edelen MO. Comparison of Patient-Reported Outcomes Measurement Information System (PROMIS®)-29 and PROMIS global physical and mental health scores. *Qual Life Res*. 2024;33(3):735-744. https://doi.org/10.1007/s11136-023-03559-y
9. Stewart A, Ware J. *Measuring Functioning and Well-Being: The Medical Outcomes Study Approach*. Duke University Press; 1992. https://doi.org/10.7249/CB361
10. Nilsdotter AK, Lohmander LS, Klässbo M, Roos EM. Hip disability and osteoarthritis outcome score (HOOS)—validity and responsiveness in total hip replacement. *BMC Musculoskelet Disord*. 2003;4:10. https://doi.org/10.1186/1471-2474-4-10
11. Royal College of Surgeons of England. National Mastectomy and Breast Reconstruction Audit—fourth report. Accessed January 25, 2024. https://www.rcseng.ac.uk/library-and-publications/rcs-publications/docs/mastectomy-breast-4/
12. Jeevan R, Cromwell DA, Browne JP, et al. Findings of a national comparative audit of mastectomy and breast reconstruction surgery in England. *J Plast Reconstr Aesthetic Surg*. 2014;67(10):1333-1344. https://doi.org/10.1016/j.bjps.2014.04.022
13. Nelson JA, Chu JJ, McCarthy CM, et al. BREAST-Q REACT: clinical reference values for the BREAST-Q in post-mastectomy breast reconstruction patients. *Ann Surg Oncol*. 2022;29(8):5280-5293. https://doi.org/10.1245/s10434-022-11521-4
14. Reeve BB, Burke LB, Chiang Y-P, et al. Enhancing measurement in health outcomes research supported by agencies within the US Department of Health and Human Services. *Qual Life Res*. 2007;16(Suppl 1):175-186. https://doi.org/10.1007/s11136-007-9190-8
15. Bierman AS, Wang J, O'Malley PG, Moss DK. Transforming care for people with multiple chronic conditions: Agency for Healthcare Research and Quality's research agenda. *Health Serv Res*. 2021;56(Suppl 1):973-979. https://doi.org/10.1111/1475-6773.13863
16. Centers for Medicare and Medicaid Services. THA/TKA PRO-PM overview. Accessed January 25, 2024. https://qualitynet.cms.gov/inpatient/measures/THA_TKA
17. Centers for Medicare and Medicaid Services. Measures Management System Hub: composite measures for accountability programs. Published online August 2023. Accessed January 23, 2024. https://mmshub.cms.gov/sites/default/files/Composite-Measures-Accountability.pdf
18. Kara P, Valentin JB, Mainz J, Johnsen SP. Composite measures of quality of health care: evidence mapping of methodology and reporting. *PLoS One*. 2022;17(5):e0268320. https://doi.org/10.1371/journal.pone.0268320
19. Bierman AS, Lawrence WF, Haffer SC, Clancy CM. Functional health outcomes as a measure of health care quality for Medicare beneficiaries. *Health Serv Res*. 2001;36(6 Pt 2):90-109.
20. Rose AJ, Bayliss E, Baseman L, Butcher E, Huang W, Edelen MO. Feasibility of distinguishing performance among provider groups using patient-reported outcome measures in older adults with multiple chronic conditions. *Med Care*. 2019;57(3):180-186. https://doi.org/10.1097/MLR.0000000000001013
21. Ambs A, Warren JL, Bellizzi KM, Topor M, Haffer SC, Clauser SB. Overview of the SEER–Medicare Health Outcomes Survey linked dataset. *Health Care Financ Rev*. 2008;29(4):5-21.
22. Hays RD, Bjorner JB, Revicki DA, Spritzer KL, Cella D. Development of physical and mental health summary scores from the Patient-Reported Outcomes Measurement Information System (PROMIS) global items. *Qual Life Res*. 2009;18(7):873-880. https://doi.org/10.1007/s11136-009-9496-9
23. Barile JP, Reeve BB, Smith AW, et al. Monitoring population health for healthy people 2020: evaluation of the NIH PROMIS® global health, CDC healthy days, and satisfaction with life instruments. *Qual Life Res*. 2013;22(6):1201-1211. https://doi.org/10.1007/s11136-012-0246-z
24. Fayers P, ed. *Assessing Quality of Life in Clinical Trials: Methods and Practice*. 2nd ed. Oxford Univeristy Press; 2007.
25. Edelen MO, Reeve BB. Applying item response theory (IRT) modeling to questionnaire development, evaluation, and refinement. *Qual Life Res*. 2007;16(S1):5-18. https://doi.org/10.1007/s11136-007-9198-0

26. Nolla K, Rasmussen LV, Rothrock N, et al. Seamless integration of computer-adaptive patient reported outcomes into an electronic health record. *Appl Clin Inform*. 2024;15(1):145-154. https://doi.org/10.1055/a-2235-9557

27. Kroenke K, Spitzer RL, Williams JBW. The PHQ-9. *J Gen Intern Med*. 2001;16(9):606-613. https://doi.org/10.1046/j.1525-1497.2001.016009606.x

28. Choi SW, Schalet B, Cook KF, Cella D. Establishing a common metric for depressive symptoms: linking the BDI-II, CES-D, and PHQ-9 to PROMIS depression. *Psychol Assess*. 2014;26(2):513-527. https://doi.org/10.1037/a0035768

29. National Committee for Quality Assurance (NCQA). Utilization of the PHQ-9 to monitor depression symptoms for adolescents and adults. Accessed January 30, 2024. https://www.ncqa.org/hedis/measures/utilization-of-the-phq-9-to-monitor-depression-symptoms-for-adolescents-and-adults/

30. Byhoff E, Guardado R, Zack RM, Nokes K, Tripodis Y. A community partnership to reduce food insecurity and improve patient-reported depression. *NEJM Catal*. 2023;4(2). https://doi.org/10.1056/CAT.22.0330

31. Reeve BB, Hays RD, Bjorner JB, et al. Psychometric evaluation and calibration of health-related quality of life item banks: plans for the Patient-Reported Outcomes Measurement Information System (PROMIS). *Med Care*. 2007;45(5 Suppl 1):S22-S31. https://doi.org/10.1097/01.mlr.0000250483.85507.04

32. Bartlett SJ, Gutierrez AK, Andersen KM, et al. Identifying minimal and meaningful change in a patient-reported outcomes measurement information system for rheumatoid arthritis: use of multiple methods and perspectives. *Arthritis Care Res*. 2022;74(4):588-597. https://doi.org/10.1002/acr.24501

33. Ahmad FS, Jackson KL, Yount SE, et al. The development and initial validation of the PROMIS®+HF-27 and PROMIS + HF-10 profiles. *ESC Heart Fail*. 2022;9(5):3380-3392. https://doi.org/10.1002/ehf2.14061

34. The PROTEUS Consortium. Home page. Accessed October 26, 2023. https://theproteusconsortium.org/

35. About PROQOLID—ePROVIDE™. ePROVIDE—Mapi Research Trust. Accessed January 30, 2024. https://eprovide.mapi-trust.org/about/about-proqolid

36. International Society for Quality of Life Research (ISOQOL). Resource center. Accessed January 30, 2024. https://www.isoqol.org/resource-center/

37. COSMIN. Improving the selection of outcome measurement instruments. COSMIN. Accessed January 25, 2024. https://www.cosmin.nl/

38. Nelson TA, Anderson B, Bian J, et al. Planning for patient-reported outcome implementation: development of decision tools and practical experience across four clinics. *J Clin Transl Sci*. 2020;4(6):498-507. https://doi.org/10.1017/cts.2020.37

39. Sisodia RC, Dankers C, Orav J, et al. Factors associated with increased collection of patient-reported outcomes within a large health care system. *JAMA Netw Open*. 2020;3(4):e202764. https://doi.org/10.1001/jamanetworkopen.2020.2764

40. MacLean CH, Antao VC, Fontana MA, Sandhu HS, McLawhorn AS. PROMs: opportunities, challenges, and unfinished business. *NEJM Catal*. 2021;2(11). https://doi.org/10.1056/CAT.21.0280

41. Stover AM, Haverman L, van Oers HA, et al. Using an implementation science approach to implement and evaluate patient-reported outcome measures (PROM) initiatives in routine care settings. *Qual Life Res*. 2021;30(11):3015-3033. https://doi.org/10.1007/s11136-020-02564-9

42. Franklin PD, Bond CP, Rothrock NE, Cella D. Strategies for effective implementation of patient-reported outcome measures in arthroplasty practice. *J Bone Joint Surg Am*. 2021;103(24):e97. https://doi.org/10.2106/JBJS.20.02072