Health conditions may cause patients to feel ill and have impaired functioning in their daily lives. Thus, it is important to assess how patients are feeling and functioning when evaluating the effects of interventions to prevent or treat health conditions. Aspects of health that patients can report on directly, such as the severity of pain or limitations in physical functioning, are patient-reported outcomes.

The recommended quantitative approach to measure these aspects of health status is to ask patients directly using a standardized questionnaire. Patient-reported outcome measures are reports of “the status of a patient’s health condition that comes directly from the patient without interpretation of the patient’s response by a clinician or anyone else.” An example of a patient-reported outcome measure, the Short Form 36 (SF-36), was used in a randomized clinical trial conducted by Ghogawala et al to compare 2 surgical approaches (ventral or dorsal) for the treatment of cervical spondylotic myelopathy, which is a condition that can cause significant impairments in physical functioning.

Explanation of the Concept

What Is a Patient-Reported Outcome Measure?

A patient-reported outcome measure is a questionnaire designed to assess a patient-reported outcome; the outcome being assessed is referred to as the concept, construct, or domain; and the individual questions of the patient-reported outcome measure are called items. Some patient-reported outcome measures are designed to measure only a single concept (eg, a 4-item measure of fatigue), whereas others are designed to measure multiple concepts (eg, the SF-36, which measures 8 health concepts and includes 36 items). A patient’s score on the concept is estimated by their answers to 1 or more questions (or items), using standardized response options for the patient-reported outcome measure.

For example, an item from the Patient-Reported Outcomes Measurement Information System SexFS intends to measure interest in sexual activity asks, “In the past 30 days, how interested have you been in sexual activity?” The response options are not at all, a little bit, somewhat, quite a bit, or very much. Ordered sets of response options such as these that are expressed as words are typically converted to numbers to compute a score (eg, not at all = 0; a little bit = 1). Some response options integrate numbers in the scale, such as the 11-point pain intensity numeric rating scale in which 0 = no pain and 10 = worst pain imaginable. Based on responses to the items, a patient-reported outcome measure generates a numeric score that represents the patient’s level on the health concept of interest.

Designing a new patient-reported outcome measure and collecting validity evidence involves a rigorous multistep process that includes both qualitative and quantitative methods. First, qualitative studies using individual interviews or focus groups are used to understand patients’ experiences with the health concept of interest so that the developers of the patient-reported outcome measure know what to assess. Next, items are written, following best practices for questionnaire design to assess the health concept. The items usually reference a recall period (eg, “in the past 7 days, …” or “in the past 24 hours, …”), which developers select based on the variability of the symptom or functioning being measured and the expected accuracy of recall. Individual cognitive interviews are then conducted with patients to evaluate the comprehensibility and appropriateness of the items and content.

Next, the new patient-reported outcome measure is administered to a sample of patients and quantitative studies are conducted to see if the items perform as well as the developers expected (eg, all item responses related to physical function are correlated with one another). Developers collect evidence of the reliability of the scores (eg, test-retest reliability; seeing how scores remain stable over time in a clinically stable population) and evidence of validity, including relationships between the scores on the new patient-reported outcome measure and other associated variables (eg, clinical markers, other established patient-reported outcome measures, or other outcomes). The type of validity evidence collected might differ depending on the type of patient-reported outcome measure (eg, a single item measuring 1 concept, multiple items measuring 1 concept, or multiple items measuring multiple concepts) and the context in which the patient-reported outcome measure will be used (eg, to screen for eligible patients in a trial, as a basis for a primary end point, or as a basis for an exploratory end point).

Why Are Patient-Reported Outcome Measures Important?

Patient-reported outcome measures often capture information about patients’ experiences with their disease or treatment that cannot be captured any other way. For example, patients can directly report on their pain, fatigue, or anxiety, whereas outside observers have only indirect access to these types of experiences through observing behaviors such as grimacing or appearing tired or nervous. Patients are typically the best reporters of their daily health experiences because they are living those experiences, whereas clinicians and others are unlikely to have full access to the patients in their lived environments. Another advantage of patient-reported outcome measures is that they are usually quicker, less burdensome, and less costly to collect than clinical measures.

Limitations or Alternatives to Using a Patient-Reported Outcome Measure

Patient-reported outcome measures cannot be used for patients who cannot reliably self-report their own health experiences, such as someone who is too young, too ill, or has cognitive impairment. In such cases, measures based on caregiver or clinician reports can be considered, but should not be considered equivalent to self-report. Memory error might influence scores when the survey items ask patients to recollect their experiences over some extended period (eg, “during the past 4 weeks, …”). When this is a concern, developers of patient-reported outcome measures can either use shorter recall periods or conduct an empirical study to evaluate the accuracy of recall.

Scores on patient-reported outcome measures may be less familiar than other types of outcomes to clinicians. To interpret study findings in terms of such scores, additional empirical studies might...
considered meaningful to patients, representing the minimal clinically important difference.6

Subjective evaluations such as patient-reported outcome measures (as well as clinician- and caregiver-reported measures) might be influenced by respondents’ expectations, which might bias estimates of treatment effects in open-label studies. There is often no gold standard for assessing how patients feel or function, which means the choice of patient-reported outcome measure cannot be based on which measure best relates to a criterion standard. (This is also the case for some clinician- and observer-reported measures.)

Incorporating patient-reported outcome measures into clinical research requires careful consideration of many issues, including the mode of administration (eg, tablet computer), prevention of missing data, cultural or linguistic relevance of patient-reported outcome measures in multinational studies, and construction of well-defined end points. Scores for patient-reported outcome measures can be on an ordinal or interval scale; thus, the appropriate statistical analyses should be used to evaluate between-group differences or changes over time. There are international guidelines for high-quality trial protocols involving patient-reported outcome measures7 and for reporting the results of patient-reported outcome measure-based end points in clinical trials.8 A guideline for the ethical considerations in the use of patient-reported outcomes has been published.9

How Was a Patient-Reported Outcome Measure Applied in This Case?

In the trial reported by Ghogawala et al,3 the researchers wished to understand how different surgical approaches altered patients’ functioning and bodily pain in their daily lives. They assessed patients before and after surgery using the SF-36 that generates 8 subscale scores reflecting different aspects of health: vitality, physical functioning, bodily pain, general health perceptions, physical role functioning, emotional role functioning, social role functioning, and mental health. Because the investigators wished to assess the effects of surgery across a broad range of physical health effects, they used the SF-36 physical component summary (PCS) score, which is a weighted combination of each patient’s scores from the 8 subscales, with greater weights applied to the physical health-related domains.

A weighted combination of domains such as the PCS score may avoid problems caused by having multiple primary end points (ie, increased likelihood of a type I error10 and difficulty interpreting study results when the treatment effects are inconsistent across multiple primary end points). The PCS scores are normalized, with a score of 50 corresponding to the average score in the US general adult population and an SD of 10. This normalization can make interpretation of scores easier. For example, someone with a PCS score of 45 could be said to have overall physical health that is 0.5-SD lower than the average US adult.

How Should the Patient-Reported Outcome Measure Be Interpreted?

The surgery trial did not find significant differences between the surgical approaches in terms of PCS scores (estimated mean difference, 0.3 [95% CI, −2.6 to 3.1]).2 Even though a composite score such as the PCS has the advantage of reflecting a broad range of health effects, it might not be sensitive to small but meaningful improvements in specific aspects of health, such as pain. Still, the use of patient-reported outcome measures in this trial2 helped researchers to understand the effects of surgery on patients’ daily lives directly from the perspective of the patients.

ARTICLE INFORMATION

Author Affiliations: Center for Health Measurement, Department of Population Health Sciences, Duke University School of Medicine, Durham, North Carolina.

Corresponding Author: Kevin P. Weinfurt, PhD, Department of Population Health Sciences, Duke University School of Medicine, 215 Morris St, Durham, NC 27701 (kevin.weinfurt@duke.edu).

Section Editors: Robert Golub, MD, Executive Deputy Editor; and Roger J. Lewis, MD, PhD, Department of Emergency Medicine, Harbor-UCLA Medical Center and David Geffen School of Medicine at UCLA.

Published Online: July 15, 2022. doi:10.1001/jama.2022.11238

Conflict of Interest Disclosures: Dr Weinfurt reported receiving personal fees from Regenener Pharmaceuticals and being a co-developer of the National Institutes of Health-funded Patient-Reported Outcomes Measurement Information System (PROMIS) and the Comprehensive Assessment of Self-Reported Urinary Symptoms. Dr Reeve reported receiving personal fees from Johns Hopkins University; receiving nonfinancial support from the European Organisation for Research and Treatment of Cancer; and being a co-developer of PROMIS, the Pediatric Patient-Reported Outcomes version of the Common Terminology Criteria for Adverse Events, the Observer-Reported Communication Ability, the Piper Fatigue Scale-12, the Patient-Centered Communication in Cancer Care, the Comprehensive Heart Disease Knowledge Questionnaire, the 2005 Healthy Eating Index, and the Everyday Discrimination Scale.

REFERENCES


