

Validation of Patient-Reported Outcomes: A Low Bar

Andrew J. Vickers, PhD¹

It used to be that if we wanted to assess a patient's back pain, we would evaluate how far the patient could bend by measuring the distance between the patient's fingers and the floor. Adverse effects of a drug were evaluated by collating clinicians' impressions. The outcomes of many trials of mental health were obtained by counting hospital readmissions. We went into medicine to help patients feel better, but rarely asked how anyone was feeling at all.

Patient-reported outcomes (PROs) first started to displace physician assessments in clinical research in the 1970s and 1980s. Instead of asking patients on a back pain trial to bend and twist, we now give them questionnaires that ask about their pain, symptoms, and function. In mental health trials, patients complete questionnaires about mood. In recent years, we have seen a similar shift in routine clinical practice. At my institution, Memorial Sloan Kettering Cancer Center (New York, NY), a patient returning for follow-up after radical prostatectomy in the early 2000s would have had his erectile function graded on a 1 to 5 scale by his surgeon (eg, grade 3 indicated partial erections satisfactory for intercourse; grade 1 indicated normal erections).¹ Today, such patients are given an online version of the International Index of Erectile Function (IIEF), a questionnaire that asks questions such as "How often were you able to get an erection during sexual activity?" and "How do you rate your confidence that you could get and keep an erection?"² This questionnaire is automatically scored, and the results are presented to the surgeon before the consultation.

There are numerous advantages to PROs. Using the back pain example, a patient may have improved pain despite experiencing no change in objectively measured back flexibility or might have better flexibility but no change in pain. Patients care about pain, and that is what should concern us. In the case of erectile dysfunction after prostate surgery, use of PROs avoids wishful thinking on the part of the surgeon, embarrassment on the part of the patient, or any number of other communication-related barriers to accurate assessment.³⁻⁶ PROs also allow a structured and systematic approach to evaluation. A surgeon might ask a patient about his sex life in general, or if asking more specifically about erectile function, the surgeon might ask whether the patient's erection was hard or firm enough for sex, intercourse, or penetration. The

IIEF asks the same questions in the same way each and every time.

However, there is no room for complacency about the current state of the PRO literature. It is said that PROs give us the patient's voice,⁷ but this is only partially true. Yes, we are hearing from patients, rather than from their doctors, but the words that they are using are not their own. The words were chosen by the researchers who designed the PRO instrument. This leads us to think about how such instruments are developed and, in particular, the issue of validation. A common view of PRO instruments is that essentially the only thing that matters is validation, that the world of PROs can be divided into validated versus non-validated instruments, that once we know an instrument to be valid there is not much more we need to know about it, other than that even a minor change to the questionnaire wording would likely render it invalid.

In the article that accompanies this editorial, Agochukwu et al⁸ report a validation study of a sexual medicine PRO, the Patient-Reported Outcome Measurement Information System (PROMIS) Global Satisfaction With Sex Life and Interest in Sexual Activity single-item measures. The study is large and carefully follows the methodology developed by psychometricians for instrument validation—that is, that the measures of sexual interest and satisfaction have a high correlation with factors with which they should be correlated (eg, nerve sparing, time since surgery, and other measures of erectile function) and have a low correlation with factors with which they should not be correlated (eg, bowel function). The authors conclude that the measures are valid and, indeed, "fundamental," in cancer survivorship. These positive findings seem to give a green light to any researcher wishing to use the PROMIS measures in a study or a clinician wishing to implement them in routine practice. However, a study with a positive result can only be a useful guide to action if the result could have been negative. A methodology that only gives one type of result is not much of a methodology at all.

Unfortunately, this is exactly what we see with studies validating PRO instruments. I searched PubMed using the phrase "quality of life validation," and here are the conclusions of the first 10 pertinent validation studies I

ASSOCIATED CONTENT

See accompanying article on page 2017

Author affiliations and support information (if applicable) appear at the end of this article.

Accepted on May 8, 2019 and published at jco.org on June 26, 2019; DOI <https://doi.org/10.1200/JCO.19.01126>

© 2019 by American Society of Clinical Oncology

retrieved (references available on request): “high values of reliability and validity”; “reliable and valid”; “reliable and valid”; “appropriate for assessing . . . function”; “acceptable validity and reliability”; “a reliable scale”; “easy to use, valid and effective”; “valid and reliable”; “a reliable tool”; and “a promising tool for assess[ment].” Validation studies are almost uniformly positive because they are non-comparative, they have no formal decision criteria for success, and moreover, correlations must fall in a middle range, not too high and not too low. We can illustrate each of these points using the work of Agochukwu et al.⁸ First, the study did not compare different possible versions of the sexual satisfaction or interest items. The investigators recommend the item “When you have had sexual activity, how satisfying has it been?” with five possible answers ranging from “not at all” to “very much.” This is because the correlations for this item were reasonable, not because those correlations were superior to those for alternative items (eg, “How fulfilling do you find your sex life?” or “How much pleasure do you get from sex?”) or answer options (eg, a 7-point scale). This is not unlike deeming a drug to be effective because patients on the drug seemed to do well, rather than because the drug was demonstrated to be superior to a placebo or existing standard of care. Second, unlike, for instance, a phase II study of a chemotherapy agent, there were no formal criteria to decide that the PROs under study were acceptable. The authors did not pre-specify that, for instance, the lower bound of a 95% CI for the correlation between item scores and the IIEF sexual satisfaction score had to be greater than 0.5 or, conversely, that correlations with bowel function had to be less than 0.2. Third, complicating the question of a decision rule is the problem of the Goldilocks effect. We know that the IIEF sexual satisfaction score is a good measure, so if the correlation between PROMIS satisfaction and IIEF satisfaction was too low, that would mean the PROMIS measure was not measuring sexual satisfaction. However, if the correlation between the two measures was too high, that would be problematic because then we might as well just use the existing IIEF instrument.

AFFILIATION

¹Memorial Sloan Kettering Cancer Center, New York, NY

CORRESPONDING AUTHOR

Andrew J. Vickers, PhD, Memorial Sloan Kettering Cancer Center, 485 Lexington Ave, New York, NY 10017; e-mail: vickersa@mskcc.org.

AUTHOR'S DISCLOSURES OF POTENTIAL CONFLICTS OF INTEREST AND DATA AVAILABILITY STATEMENT

Disclosures provided by the author and data availability statement (if applicable) are available with this article at DOI <https://doi.org/10.1200/JCO.19.01126>.

The standard psychometric approach to questionnaire validation was used to create the IIEF erectile function measure that we use in our daily work at Memorial Sloan Kettering Cancer Center, and the drawbacks are apparent in our clinics. For instance, the IIEF includes three questions that make reference to intercourse (eg, “When you attempted intercourse, how often were you able to penetrate [enter] your partner?”). These items are scored such that a man who answers “did not attempt intercourse” gets 0 points and automatically falls into the category of having erectile dysfunction. This means that the IIEF underestimates function in gay men, those without a willing and able partner, and those who prefer to have sex without intercourse. When tracking IIEF in men after radical prostatectomy, we have sometimes seen dramatic decreases in scores after an initial recovery but then might have a patient respond to a concerned surgeon merely that “my wife was away on business.” Gay men have complained to us that the IIEF questions are not appropriate for them. The IIEF also makes repeated reference to erections, even though this is not a word used in certain groups, such as in parts of the African American community.⁹

These sorts of problems also have implications for research. Agochukwu et al⁸ conclude that “[p]atients are interested in sex despite functional losses and can salvage satisfaction.” My guess is that this conclusion is sound. However, it should be noted that, for example, a gay man might have good erections and have high levels of desire and sexual satisfaction but may respond on the IIEF that he does not have intercourse and, therefore, get a low erectile function score. This man would be classed by the authors as having “salvage[d] satisfaction” despite “functional losses,” although, in fact, he is perfectly functional.

Studies like those of Agochukwu et al⁸ do provide tools that have reasonable properties at the group level. And that is not a bad start. However, we have to go on to ask whether these tools could be modified to have even better properties or be adapted for clinical use to help a clinician better evaluate an individual patient.

ACKNOWLEDGMENT

Supported in part by funds from the Sidney Kimmel Center for Prostate and Urologic Cancers; Specialized Programs of Research Excellence Grant No. P50-CA92629 from the National Cancer Institute to Howard Scher, MD; and National Cancer Institute Cancer Center Support Grant No. P30-CA008748 to Memorial Sloan Kettering Cancer Center.

REFERENCES

1. Vickers A, Savage C, Bianco F, et al: Cancer control and functional outcomes after radical prostatectomy as markers of surgical quality: Analysis of heterogeneity between surgeons at a single cancer center. *Eur Urol* 59:317-322, 2011
2. Vickers AJ, Savage CJ, Shouery M, et al: Validation study of a web-based assessment of functional recovery after radical prostatectomy. *Health Qual Life Outcomes* 8:82, 2010
3. Passik SD, Kirsh KL, Donaghy K, et al: Patient-related barriers to fatigue communication: Initial validation of the fatigue management barriers questionnaire. *J Pain Symptom Manage* 24:481-493, 2002
4. Peters S, Rogers A, Salmon P, et al: What do patients choose to tell their doctors? Qualitative analysis of potential barriers to reattributing medically unexplained symptoms. *J Gen Intern Med* 24:443-449, 2009
5. Pakhomov SV, Jacobsen SJ, Chute CG, et al: Agreement between patient-reported symptoms and their documentation in the medical record. *Am J Manag Care* 14:530-539, 2008
6. Hartmann U, Burkart M: Erectile dysfunctions in patient-physician communication: Optimized strategies for addressing sexual issues and the benefit of using a patient questionnaire. *J Sex Med* 4:38-46, 2007
7. Basch E: Patient-reported outcomes: Harnessing patients' voices to improve clinical care. *N Engl J Med* 376:105-108, 2017
8. Agochukwu NQ, Wittmann D, Boileau NR, et al: Validity of the Patient-Reported Outcome Measurement Information System (PROMIS) Sexual Interest and Satisfaction measures in men following radical prostatectomy. *J Clin Oncol* 37:2017-2027, 2019
9. Kilbridge KL, Fraser G, Krahn M, et al: Lack of comprehension of common prostate cancer terms in an underserved population. *J Clin Oncol* 27:2015-2021, 2009



The Authoritative Resource for Practicing Oncology

Journal of Oncology Practice (JOP) is a peer-reviewed medical journal published monthly that provides oncologists with information and tools to enhance practice efficiency and promote a high standard of quality patient care.

Whether in an office, hospital, or academic setting, *JOP* provides evidence- and data-driven strategies to help you:

- Allocate resources and improve systems
- Deliver efficient, high-quality care
- Control costs and maintain value
- Understand key issues in health policy
- Improve patient flow, implement guidelines, and more

Plus clinical content featuring clinical reviews, expert commentaries, and case reports!

Subscribe today at jop.ascopubs.org



ASCO
AMERICAN SOCIETY OF CLINICAL ONCOLOGY

AUTHOR'S DISCLOSURES OF POTENTIAL CONFLICTS OF INTEREST

Validation of Patient-Reported Outcomes: A Low Bar

The following represents disclosure information provided by the author of this manuscript. All relationships are considered compensated. Relationships are self-held unless noted. I = Immediate Family Member, Inst = My Institution. Relationships may not relate to the subject matter of this manuscript. For more information about ASCO's conflict of interest policy, please refer to www.asco.org/rwc or ascopubs.org/jco/site/fic.

Andrew J. Vickers

Stock and Other Ownership Interests: OPKO Health

Consulting or Advisory Role: OPKO Diagnostics

Patents, Royalties, Other Intellectual Property: I am named on a patent for a statistical method to detect prostate cancer. This method has been commercialized by OPKO as the 4Kscore. I receive royalties from sales of the 4Kscore.

Travel, Accommodations, Expenses: OPKO Health

No other potential conflicts of interest were reported.