Assuring the patient-centeredness of patient-reported outcomes: content validity in medical product development and comparative effectiveness research

Ethan Basch, MD, MSc[1], **Amy P. Abernethy, MD**[2], and **Bryce B. Reeve, PhD**[3]

[1]Center for Policy and Outcomes Research, Health Outcomes Research Group, Memorial Sloan-Kettering Cancer Center, New York, NY, USA

[2]Duke Cancer Care Research Program, Duke Cancer Institute, Durham, NC, USA

[3]Lineberger Comprehensive Cancer Center; Department of Health Policy and Management, Gillings School of Global Public Health, University of North Carolina, Chapel Hill, NC, USA

Keywords

Not all patient-reported outcomes are patient-centered, and not all patient-centered outcomes are patient-reported.

The essential characteristic of a patient-centered approach to outcome measurement is that it assesses concepts (i.e., health-related phenomena) that are considered most important by members of a given target population, based on direct input from representatives of that population. Concepts for measurement should not be selected based solely on convenience or interest to investigators. *Patient-centered patient-reported* outcome measures must meet this criterion and also be meaningful and comprehensible to members of a population when administered, including among those with diverse racial/cultural backgrounds and lower educational/literacy levels.

An example of a patient-reported outcome (PRO) measure which is not optimally patient-centered is the Present Pain Intensity (PPI) item of the McGill Pain Questionnaire [1]. The PPI has been used as a PRO measure in multiple phase III clinical trials in the regulatory context (including serving as the basis for US drug approval and labeling of the cancer drug mitoxantrone in 1996) [2]. This measure, however, was initially developed for clinician

Corresponding Author: Ethan Basch, Center for Policy and Outcomes Research, Health Outcomes Research Group, Memorial Sloan-Kettering Cancer Center, 1275 York Avenue, New York, NY, USA 10065.

reporting and never underwent qualitative evaluation with direct patient input. The item asks respondents:

> **_How strong is your pain?_**
>
> _1   2   3   4   5_
>
> _Mild   Discomforting   Distressing   Horrible   Excruciating_

The response options mix up the attributes of intensity ("mild") and bother ("distressing"), and the distinctions between options at each extreme of the scale are not clear ("mild" vs. "discomforting" and "horrible" vs. "excruciating"). Item development with direct patient input, and cognitive interviewing to assure patient understanding, would likely have yielded different response options. These limitations were highlighted at a meeting of the Food and Drug Administration's Oncology Drug Advisory Committee in 2007 [3], just subsequent to issuance of the FDA's 2006 draft _Guidance for industry: Patient-reported outcomes measures: Use in medical product development to support labeling claims_ (issued in final form in 2009) [4]. Subsequently, use of the "worst pain item" of the Brief Pain Inventory has been advocated by the FDA for pain intensity assessment in the regulatory context [5].

Because the PPI was not developed with a patient-centered approach, its ability to adequately assess the patient pain experience associated with disease and treatment is in question. Viewing this same concept through a patient lens, a patient-centered patient-reported outcome measure must be understandable to patients with a variety of backgrounds, which requires direct patient input during development and revisions.

In contrast, a measure can be patient-centered without being a PRO. For example, exercise capacity is an important concept to patients in selected contexts, and is best evaluated with an objective approach such as a treadmill test.

These examples highlight the importance of thoughtfulness when selecting or developing outcomes for use in clinical research. As alluded to above, this applies both to the _identification of concepts_ to be evaluated as outcomes in a given context, and to the _development of outcome measures_ for assessing these concepts.

The two-part paper on content validity in this issue of _Value in Health_ [7,8] provides an important contribution to the methodological literature and complements the FDA's PRO Guidance towards standardizing methods for development of endpoint models and PRO instruments which are optimally patient-centered. The key message of this paper is the importance of directly eliciting the patient perspective through qualitative research during identification of concepts to be evaluated (whether patient-reported or not), and development/refinement of PRO measures.

There have been critiques of the FDA's PRO Guidance as setting too high a methodological bar for sponsors to attain and of its approach to content validity as being overly focused on qualitative over quantitative methods [8]. While there is some truth to these assertions, the overall impact of the FDA's emphasis on qualitative methods has been positive, leading

investigators in both the regulatory and non-regulatory spaces to focus on the patient perspective – thus creating a need for a blueprint as provided by the two-part paper in this issue.

A potential limitation of the methods described in this two-part paper is reliance on the good faith and judgment of investigators who are interpreting qualitative data to decide which concepts are most appropriate to measure in a given context. There is a risk that investigators will select concepts for measurement which cast a particular product in the most positive light while ignoring other concepts which could appear less favorable. For example, if a product reduces pain but increases fatigue, investigators could choose to evaluate only the former although both are important to patients. This misuse of the regulatory tenet of "fit for use" endpoint design risks conveying an incomplete picture of the impact of treatment on the patient subjective experience. In fact, in part one of this paper, in the discussion titled "Understand the Disease or Condition in the Target Population" it is acknowledged that "The selection of outcomes appropriate for a given trial program is often informed by consultation with clinical, trial design, and measurement experts as well as an extensive literature review."

Therefore, a key item for investigators who are developing a new PRO measure, or selecting an existing measure for use in a new study, to 1) describe all of the concepts reported as important by patients in the target population or in a closely related population, and 2) provide a rationale for which concepts were included or excluded.

How are the recommendations of this two-part paper, which largely apply to the regulatory context (and largely to the US regulatory context), apply to comparative effectiveness research (CER)? For prospective CER controlled clinical studies, the recommendations should be taken virtually intact, with a particular emphasis on developing conceptual frameworks for the relationships of various outcomes - as elegantly shown for psoriasis in Figure 1 of part one of the paper. But the "fit for purpose" focus of the regulatory setting is less applicable to other CER approaches such as registries or longitudinal observational studies which are often more exploratory in nature. In such instances, inclusion of a broader selection of outcomes, some intended for signal generation beyond initial qualitative work, is merited. For example, a multi-item symptom or HRQL battery is appropriate, in addition to measurement of context-specific concepts of interest. Regardless, up-front qualitative research to identify salient concepts prior to conducting any type of prospective clinical CER is strongly recommended. It is critical to remember that research to inform care of patients – and to be understood and interpreted by patients – is one of the targets of CER; hence patient-centered PRO measures need to be consistently understandable and meaningful to patients themselves, and this generally requires patient input up front. Notably, additional qualitative evaluations with patients after completing quantitative psychometric studies can also be informative. For example, in psychometric analysis of a patient-reported measure, an item response bias could be detected (i.e., differential item functioning, DIF) between Hispanic and non-Hispanic patients. Statistical tools can identify DIF, but it requires qualitative work to illuminate underlying drivers of these differences. Qualitative methods can also be incorporated into prospective clinical research to provide insights about patient perspectives at key time points, or about the relationships of outcomes

to each other or with interventions. Once a measure is developed, it should be periodically re-evaluated as treatment paradigms and patient perspectives shift over time, to assure it remains appropriate and representative of meaningful concepts.

As noted in the two-part paper, central to the importance of conducting qualitative evaluations for establishing content validity in the regulatory and CER environments is inclusion of a heterogeneous sample within the target population. In addition to including representatives of various racial/cultural or applicable linguistic backgrounds, those with diverse educational/literacy levels should be included. Moreover, methodological expertise to analyze these data and adjust verbiage accordingly is recommended. Patients with higher educational levels are easier to identify and recruit in clinical research, and therefore efforts must be made to include hard-to-reach individuals. For example, it is a requirement in the development work of two US National Institutes of Health (NIH) initiatives, the Patient-Reported Outcomes Measurement Information System (PROMIS) and Patient-Reported Outcomes version of the Common Terminology Criteria for Adverse Events (PRO-CTCAE), to include cognitive interviews among patients with low education levels (e.g., < 12 years of education or measured reading level less than 9th grade using the Wide Range Achievement Test-3 Reading subtest) [9,10]. In summary, qualitative research is essential for identifying concepts for measurement in a given target population, for refining measures which assess these concepts, and for continued assurance that the concepts and measures remain appropriate and meaningful over time. This work should include hard-to-reach patients, particularly those of diverse racial/cultural and educational/literacy levels. This approach applies both to trials in the regulatory context and to prospective clinical CER. Beginning this process as early as possible in a given research program will afford an opportunity to develop or select appropriate concepts and measures that are optimally patient centered.

Acknowledgments

References

1. Graham C, Bond S, Gerkovich M, Cook M. Use of the McGill Pain Questionnaire in the assessment of cancer pain: replicability and consistency. Pain. 1980; 8:377–87. [PubMed: 7402695]

2. U.S. Food and Drug Administration. [Accessed March 20, 2011] Novantrone Drug Label. Available from: http://www.accessdata.fda.gov/drugsatfda_docs/label/2009/019297s030s031lbl.pdf

3. FDA Office of Oncology Drug Products, Center for Drug Evaluation and Research. [Accessed October 3, 2011] FDA Expectations for Endpoint Adequacy. Presentation at Oncology Drugs Advisory Committee (ODAC). Jul 24. 2007 Available from: www.fda.gov/ohrms/dockets/ac/07/slides/2007-4309s1-14-FDA-Basch.ppt

4. U.S. Food and Drug Administration. [Accessed September 28, 2011] Guidance for industry: Patient-reported outcomes measures: Use in medical product development to support labeling claims. Issued December 2009. Available from: http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM193282.pdf

5. Atkinson TM, Mendoza TR, Sit L, et al. The Brief Pain Inventory and its "pain at its worst in the last 24 hours" item: clinical trial endpoint considerations. Pain Med. 2010; 11:337–46. [PubMed: 20030743]

6. Patrick DL, Burke LB, Gwaltney CJ, et al. Content validity - establishing and reporting the evidence in newly-developed patient-reported outcomes (PRO) instruments for medical product evaluation: ISPOR PRO good research practices task force report: Part 1 – Eliciting Concepts for a New PRO Instrument. Value Health. 2011; 14:XX–XX.

7. Patrick DL, Burke LB, Gwaltney CJ, et al. Content validity - establishing and reporting the evidence in newly-developed patient-reported outcomes (PRO) instruments for medical product evaluation: ispor pro good research practices task force report: Part II - Assessing Respondent Understanding. Value Health. 2011; 14:XX–XX.

8. Magasi S, Ryan G, Revicki D, et al. Content validity of patient-reported outcome measures: perspectives from a PROMIS meeting. Qual Life Res. 2011 August 25, 2011.

9. DeWalt DA, Rothrock N, Yount S, Stone A. on behalf of the PROMIS Cooperative Group. Evaluation of Item Candidates - The PROMIS Qualitative Item Review. Med Care. 2007; 45(Suppl):S12–S21. [PubMed: 17443114]

10. Hay J, Atkinson TM, Mendoza TR, et al. Refinement of the patient-reported outcomes version of the common terminology criteria for adverse events (PRO-CTCAE) via cognitive interviewing. J Clin Oncol. 2010; 28(Suppl):15s. abstr 9060.