

Conceptual framework: A conceptual framework ensures that a researcher identifies and defines: what he or she intends to measure (i.e., the PRO concept); how he or she will measure it (i.e., the PRO instrument); and why he or she is measuring it (i.e., what is the label claim to be supported by the PRO data).

PRO Instrument Selection Factors

The importance of a thoroughly specified conceptual framework cannot be understated, since it is the basis of all the other evaluations of the instrument and for evaluating its relevance to product labeling. In other words, stating clearly what the researcher intends to measure and relating it clearly to the proposed claim, to the condition being treated, and to the product's mechanism of action sets the stage for justifying why Questionnaire A is well-suited or better suited for the task than Questionnaire B.

Most PRO assessments evaluate either signs and symptoms of a condition or health-related quality of life (HRQOL) domains. These two areas have substantial conceptual and practical differences that need to be addressed in the conceptual framework. In general, researchers should be aware that measures of signs and symptoms whose relationship to the underlying disease is well known and which are measured unidimensionally will tend to generate very simple and very straightforward conceptual framework models. On the other hand, the conceptual framework is more complex and elaborate when researchers propose to measure HRQOL domains, which are measured multidimensionally and whose relationship with disease processes are indirect and mediated, for example, by symptom improvements. Table 2 broadly characterizes the typical properties of symptom vs. HRQOL assessment.⁴

Characteristic	Signs and Symptoms	HRQOL Assessment
Measurement	Unidimensional	Multidimensional
Relationship to disease	Direct	Indirect and mediated
Conceptual framework	Simple and straightforward	Complex and elaborate
Measurement error	Low	High
Response options	Simple	Complex
Scoring	Simple	Complex
Interpretation	Simple	Complex
Validation	Simple	Complex
Reliability	High	Low
Validity	High	Low
Stability	High	Low
Responsiveness	Low	High
Feasibility	High	Low
Acceptability	High	Low
Cost	Low	High
Time	Low	High
Complexity	Low	High
Interpretability	High	Low
Communication	High	Low
Documentation	High	Low
Regulatory	High	Low
Commercial	High	Low

Properties of Symptoms Versus HRQOL Assessment

Administration characteristics: Documenting the administration characteristics of a PRO instrument outlines what the PRO instrument will look like when administered, for example, the test format, patient instructions, test items, response options, data collection method, and scoring procedures. When using an existing and unmodified PRO instrument, this information can often be gleaned from the actual instrument. However, when creating a new instrument or when using a modified tool, it is the researcher's responsibility to provide a sound logical and empirical rationale for why an instrument looks as it does (e.g., how did you generate a pool of items? how did you select the final test items?) or why a particular modification was made (e.g., why did you add or omit items from the original version?).

Researchers should be prepared to document their expectations that the instrument can be read and understood by the intended population. Parameters such as reading level and length are relevant here, but cognitive debriefing (having a small number of subjects use the instrument and then interview them about the experience) is often useful in ensuring the respondents understand and interpret the items and response options as intended.

Performance characteristics: Reliability and validity are the principal performance characteristics that the FDA will evaluate. Broadly, reliability is the extent to which measurements are stable and repeatable (i.e., free from measurement error or random "noise"), and validity is the extent to which a test is measuring what it says it will measure. There are several types of reliability and validity estimates. While defining all of them is beyond our scope, it is important to recognize that it is the researcher's burden to determine which reliability and validity estimates are most relevant to their PRO instrument.

When using existing PRO instruments, reliability and validity data can often be obtained from previous reports (e.g., published studies, test manuals, etc.). However, it is not enough to say that "PRO Instrument X" is a reliable and valid tool because it has been previously administered. Clinical researchers need to show that the previous reliability and validity estimates apply to the intended use—i.e., that they were obtained under conditions of administration resembling those planned for the new trial (e.g., similar patient population, same instruction set, items, recall interval, etc.).

Researchers using PRO tools should also be aware that the ability to detect patient change over time is another performance criterion that the FDA will evaluate. This is important because not all PRO instruments are designed to assess change. For example, some PRO instruments are used as study screening tools to assess a PRO concept at a single point in time, but may not be sensitive to change (e.g., one assessing history of a condition). Sensitivity to change is demonstrated when scores on an instrument change in the theoretically proposed direction following the introduction of a known-effective treatment (i.e., they reflect clinically meaningful improvement).

Study design: A variety of study design features can systematically alter how patients respond to PRO

questions and, thus, can systematically alter conclusions drawn from PRO data.⁵ For example, patient diary studies using paper-and-pencil for off-site data collection can generate far less valid and trustworthy PRO data because these diaries are often back-filled and even forward-filled (i.e., completing a diary entry before its designated time⁶).

Researchers can easily defend off-site data collection study designs to the FDA, however, by documenting the use of more modern methods such as electronic diaries, which can time- and date-tag data entry. The PRO Guidance suggests that researchers discuss specifically how the PRO assessment strategy within the study design adequately and accurately evaluates the primary hypothesis for the study.

Instrument modifications

Clinical researchers are challenged by the FDA's intentions to evaluate modified PRO tools as if they were newly developed instruments. Many researchers are wondering if this means that any and all PRO instrument changes must be revalidated by additional and extensive psychometric revalidation studies. The FDA actually takes a more flexible and nonprescriptive tone, which is reflected in the PRO Guidance: "The extent of additional validation recommended depends on the type of modification made" (Reference 2, lines 582–583).

This flexible point of view is logical and consistent with decades of psychometric and applied research, which shows that small changes have little effect on an instrument and larger changes can undermine its validity. For example, plain, simple changes in presentation such as in font, color, or size of the paper used for pen-and-paper PRO instruments would not require any empirical revalidation of any sort, unless they directly affected readability. Conversely, substantial changes in the content or meaning of items or in the content or number of response options can significantly influence the performance of an assessment^{7,8} and would, therefore, require considerable empirical support to validate that change.

Corresponding to the range of modifications, the PRO Guidance recognizes a variety of validation procedures, ranging from less to more resource intensive. For example, validation evidence could be obtained via existing studies, cognitive debriefing, equivalence testing, and/or full psychometric validation studies. Yet, conceptualizing the potentially unlimited number of PRO instrument modifications and deciding exactly which procedures to use to validate those changes, poses a challenge for some clinical researchers.

Fortunately, it is not an insurmountable challenge. We offer a "Validation Hierarchy" that proposes a hierarchy of instrument modifications and a corresponding level of evidence required to support each level of modification (see Table 3). By determining the level of change made in the instrument, the researcher can identify the appropriate level of revalidation that may be necessary.⁹ If a researcher decides that a PRO instrument modification is necessary, our Validation Hierarchy prompts him or her to categorize that change among the four levels of PRO instrument modifications, which range from small to substantial. This decision, in turn, suggests a particular validation procedure (ranging from none to psychometric revalidation) to validate that change.

Modification	Evidence	Validation Procedure	Supporting Evidence
No change in content or meaning	None	None	None
Small change in content or meaning	None	None	None
Medium change in content or meaning	None	None	None
Large change in content or meaning	None	None	None
Substantial change in content or meaning	None	None	None

To categorize the level of change made to an instrument, the researcher should first ask whether that change alters the content and/or meaning of the original instrument. If the modification does not alter the content and/or meaning of the original instrument, it is a "small" or "medium" change and requires either no validation or minor validation. If, however, the modification does alter the content and/or meaning of the original instrument, then more validation may be necessary.

Next, the researcher should determine whether the proposed modification of the PRO instrument has already been empirically validated. In some cases, prior empirical data may establish the impact of a particular change. For example, in a recent review of 41 studies evaluating 235 scales,¹⁰ we found that porting an instrument from paper to electronic administration on a PC or palmtop platform yielded a psychometrically equivalent instrument, thus mitigating the need to test (i.e., revalidate) every time a new instrument is migrated from paper to electronic administration.

Importantly, this example highlights the notion that instrument validation procedures, when interpreted cumulatively and thoughtfully, can produce results that generalize and are applicable beyond a particular PRO instrument. The practical advantage of this is that researchers porting their paper PRO tools to electronic text

platforms, regardless of the tool, can be confident that the migration will produce equivalent results and that they will not have to invest their resources in more expensive and time consuming revalidation procedures.

Summary

The recently released PRO Guidance carries with it tremendous promise as well as certain challenges. Researchers planning clinical trials must adapt to new regulatory standards regarding the selection, use, and interpretation of data generated by PRO instruments. We believe, however, that these challenges are manageable. Our intent in this article was to inform researchers as to what information is needed to support their selection of a particular PRO instrument. Additionally, we explained how researchers can categorize modifications they make to PRO instruments to inform them of the validation procedures needed to support that modification.

With appropriate planning, researchers can enhance the utility of their PRO assessments and their acceptance by FDA in support of label claims. If the FDA, as anticipated, begins to apply the standards laid out in the PRO Guidance to the assessment of other clinical endpoints (e.g., clinician-reported outcomes assessments), then researchers will be well prepared to easily manage these expectations as well.

References

1. R.J. Willke, L.B. Burke, P. Erickson, "Measuring Treatment Impact: A Review of Patient-Reported Outcomes and Other Efficacy Endpoints in Approved Product Labels," *Controlled Clinical Trials*, 25 (6) 535–552 (2004).
2. Federal Register, Vol. 71, No. 23; Friday, February 3, 2006.
3. PRO Consulting (Unpublished White Paper), "Documentation of PRO Instruments to Meet Contemporary FDA Standards," www.patientreported.com (2006).
4. PRO Consulting (Unpublished White Paper), "Distinguishing Among Symptom vs. Health Related Quality of Life PRO Concepts: Developing a Conceptual Framework," www.patientreported.com (2006).
5. K.F. Schulz, I. Chalmers, R.J. Hayes, "Empirical Evidence of Bias: Dimensions of Methodological Quality Associated with Estimates of Treatment Effects in Controlled Trials," *JAMA*, 273, 508–512 (1995).
6. AA. Stone, S. Shiffman, J.E. Schwartz, J.E. Broderick, M.R. Hufford, "Patient Non-Compliance with Paper Diaries," *British Medical Journal*, 324, 1193–1194 (2002).
7. G. Menon and E.A. Yorkston, "The Use of Memory and Contextual Cues in the Formation of Behavioral Frequency Judgements." In A.A. Stone, J.S. Turkkan, C.A. Bachrach, J.B. Jobe, H.S. Kurtzman, V.S. Cain, eds., *The Science of Self-Report: Implications for Research and Practice* (Erlbaum, Mahwah, NJ, 2000).
8. N. Schwarz, "Self-Reports. How the Questions Shape the Answers," *American Psychologist*, 54, 93–105 (1999).
9. PRO Consulting (Unpublished White Paper), "Validating and Revalidating PRO Instruments for Use in Specified Clinical Trials," www.patientreported.com (2006).
10. PRO Consulting (Unpublished White Paper), "Equivalence of Electronic and Paper-and-Pencil Administration of Patient Reported Outcomes: A Meta-analysis," www.patientreported.com (2006).

Alan Shields , * PhD, is a scientific consultant with PRO Consulting, a division of invivodata inc., 2100 Wharton Drive, Suite 505, Pittsburgh, PA 15203, (412) 697-6390, email: ashields@patientreported.com and an assistant professor of Psychology, East Tennessee State University. **Chad Gwaltney** , PhD, is a scientific consultant with PRO Consulting and an assistant professor (research) in the Department of Community Health, Brown University. **Brian Tiplady** , PhD, is a senior scientific consultant with PRO Consulting. **Jean Paty** , PhD, is founder and senior vice president of scientific, quality, and regulatory affairs with invivodata inc. **Saul Shiffman** , PhD, is founder and chief science officer with invivodata, inc.

**To whom all correspondence should be addressed.*