

Standards for Evaluation and Documenting Psychometric Qualities of PRO Instruments

ISOQOL Patient-Reported Outcomes and Regulatory Guidance Meeting June 29, 2006 Washington, DC



Contributors

- Ron Hays
- Neil Aaronson,
- ✤ Diane Fairclough
- Dennis Revicki
- Bill Lenderking
- Jeff Sloan

Patient Reported Outcomes and FDA Regulatory Guidance Meeting



Session Objectives

- Summarize methods of assessment of reliability and validity of measures
- Provide guidance on reliability and validity
- Level of evidence that indicate PRO measure has sufficient reliability and validity



Reliability

- Extent to which measure yields same score when the outcome has not changed
- Estimation approaches
 - ✤ Internal consistency
 - ✤ Test-retest
 - ✤ Inter-rater or inter-interviewer
 - ✤ Scale information
- Minimum standard
 - ✤ 0.70 for group comparisons
- Reliable measures can detect differences efficiently (smaller sample size)



Draft Guidance Document p. 18

* "Test-retest reliability is the most important type of reliability for PRO constructs used in clinical trials."

Precision and Accuracy



The difference between precision and accuracy.

Want both precision and accuracy in measuring endpointsIncreasing sample size increases precision but not accuracy



Validity

- Extent to which measure yields score it should—measures what it is intended to measure
- Validity exists along a continuum
- Two main flavors of validity
 - Content—extent to which PRO measures appropriate content and represents variety of attributes that define the concept. Construct—extent to which measure "behaves" in a way consistent with theoretical hypotheses



Content Validity

- * Recommend "triangulation" of input from previous literature, patients, health care providers and, in some cases, informal caregivers (e.g., parents, spouses, teachers, etc.).
- Focus groups and other qualitative methods can help suggest important content and identify content gaps
- Experts can judge appropriateness of content

Assessing Construct Validity

- "Hypothesize expected relationships among concepts" (Fig 1, p. 7).
- Evaluate covariation of PRO scores with other measures to see whether patterns are consistent with hypotheses
 - Item-scale correlations for hypothesized scales exceed correlations of items with other scales
 - Correlations among measures of different concepts indicate sufficient unique variance
 - Cross-sectional associations (older age is correlated with lower physical function)
 - Longitudinal associations (raising hematocrit to normal levels leads to increases in energy)

Evaluation of Conceptual Framework







Evaluating Hypothesized Associations

Scale	Hypotheses	Results
Near vision	++	0.71
Driving	+	0.43
Ocular pain	\sim	0.07

++=0.50 or above; +=0.20-0.49, $\sim = < 0.20$



Interpretation of Scores

- Construct validity evaluation helps identify meaningful differences
- Responsiveness to change means the measure changes in accordance with the underlying continuum of change
 - Minimal important difference is a subset of responsiveness to change
 - Difference associated with smallest underlying change that is important

Measurement Equivalence

* "Extent to which PRO instrument' s ability to detect change varies by important patient subgroups (e.g., sex, race, age, or ethnicity) can affect clinical trial results. It is important to identify any important subgroup differences in ability to detect change so that these differences can be taken into account in assessing results" (p. 18-19)



Level of Evidence Needed

- Multiple pieces of supporting evidence increases confidence in psychometric properties
 - Two or more focus groups and saturation for content validity
 - Multiple experts to judge content validity
 - Cross-validation or replication of empirical associations in two or more samples of sufficient sample size
 - Cannot assume that measure will perform as well in every conceivable sample but a measure that works well in multiple applications is likely to perform well in many circumstances.

Degree of Additional Psychometric Evidence Depends on Intended New Application

- Adults -> children
- Educated -> Less educated
- Self-administered -> phone
- ✤ White -> Asian
- Men -> women

Less -> more severity of targeted condition
Age 18-29 -> 30-39 vs. 75+



Evidence in Phase III Trial

Validating an instrument within a "Phase II clinical trial ... obviously entails some risk because you could make the argument that it's not well-defined and reliable yet" (Powers, Medical Officer, Pink Sheet, April 17, 2006)



Trial Period Recommended

- FDA can better understand how the guidelines are applied in practice and assess need for guideline revisions
- Issues that generate problems for FDA or for sponsors



Summary

FDA has done a good job in drafting the guidelines
Flexibility in evaluating sufficient psychometric properties is important



Discussion

Margaret Rothman, Ph.D. Executive Director, HE&P, PGSM RW Johnson Pharmaceutical Research Institute



Premature Ejaculation: An Observational Study of Men and Their Partners

Donald L. Patrick, PhD, MSPH,* Stanley E. Althof, PhD,[†] Jon L. Pryor, MD,[§] Raymond Rosen, PhD,¹ David L. Rowland, PhD,** Kai Fai Ho, PhD,^{††} Pauline McNulty, PhD,^{‡‡} Margaret Rothman, PhD,[#] and Carol Jamieson, BSc^{§§}

*Department of Health Services and Epidemiology, University of Washington, Seattle, WA; [†]Case Western Reserve University School of Medicine, Beachwood, OH; [‡]Albany Medical College. Albany, NY; [§]University of Minnesota, Minneapolis, MN; [¶]Robert Wood Johnson Medical School, Piscataway, NJ; **Valparaiso University, Valparaiso, IN; ^{††}ALZA Corporation, Mountain View, CA; ^{‡†}Johnson and Johnson Pharmaceutical Services, LLC, Raritan, NJ; ^{§§}Johnson and Johnson Pharmaceutical Services, LLC, Mountain View, CA, USA

Corresponding Author: Donald L. Patrick, PhD, MSPH, Department of Health Services and Epidemiology, University of Washington, 1959 NE Pacific St., Box 357660, Seattle, WA 98195-7660, USA. Tel: (206) 616-2981; (206) 543-3964; E-mail: Donald@u.washington.edu

ABSTRACT

Introduction. Premature ejaculation (PE) is the most common male sexual dysfunction affecting men and their partners. Lack of community-based data describing this condition limits understanding of PE and its outcomes.

Aim. To characterize PE in a large population of men with and without PE using patient-reported outcome (PRO) measures elicited from men and their partners.

Methods. 4-week, multicenter, observational study of males (\geq 18 years) and their female partners in monogamous relationships (\geq 6 months). Screening, baseline, and follow-up visits scheduled at 2-week intervals. Clinicians diagnosed PE utilizing DSM-IV-TR criteria. Intravaginal ejaculatory latency time (IELT), measured by a stopwatch held by the partner, was recorded for each sexual intercourse experience. Subject and partner independently assessed PROs: *control over ejaculation* and *satisfaction with sexual intercourse* (0 = very poor to 4 = very good), *personal distress* and *interpercond difficulty* (0 = pot et all to 4 = outroppely) and escentive of PE (0 = pope to 3 = course).

🔄 🛛 1 of 10 🛛 🔊 🚺 🙆 🙆

HH

Appendix: Session Abstract

This session proposes standards for evaluating and documenting the psychometric qualities of PRO measure of use in medical product development and to support labeling claims. We will summarize methods for assessing reliability and validity (including responsiveness) of measures and provide guidance for evaluating these psychometric properties. The presentation will cover the kinds of evidence needed to indicate that a PRO measure has a sufficient level of reliability and validity, evaluation approaches that can be used when a measure is revised, and the types of reliability and validity evaluation that are appropriate during different phases of clinical trials.