**SUPPLEMENT TO Clinician and Patient Reported Endpoints in CNS Orphan Drug Clinical Trials: ISCTM Position Paper on Best Practices for Endpoint Selection, Validation, Training, and Standardization**

***A Position Paper of the ISCTM Rare Disease/Orphan Drug Development Working Group***

*(Note to editor: Supplement references are marked with S then cited in the order they appear in the supplement; supplement reference list follows supplement text)*

I. Validity and Reliability

Validity

Patient-reported outcome (PRO) measures, to be useful as primary or secondary endpoints for clinical trials in orphan diseases, need to be well conceptualized and have sound psychometric qualities. In addition, regulatory agencies require evidence supporting the content and measurement properties of PRO measures (see S1 for FDA guidance). Although providing evidence supporting reliability (i.e., internal consistency, test-retest reliability, etc.) is important, instrument developers and users need to establish content validity and construct validity. [S2] [S3] [S4]

*Content Validity*

Content validity is normally evaluated and confirmed based on qualitative research, including individual interviews and focus groups. This qualitative research usually proceeds in a series of stages starting with literature reviews and interviews with experienced clinicians, followed by individual interviews and focus groups to identify relevant concepts associated with the target disease experience (e.g., symptoms; physical, emotional and social functioning and well-being, etc.). This qualitative research may either be narrow (i.e., key symptoms) or broad in scope (i.e., health-related quality of life) depending on the PRO under development. Once saturation of concepts criteria are met, the instrument developer constructs a draft measure, including instructions, item content and response scales. This draft instrument is then evaluated for content and respondent understanding of instructions, item content, and response scales based on cognitive interviews. This cognitive debriefing ensures that the PRO measure covers the relevant concepts and that the instrument is well-understood and comprehensive. This totality of qualitative evidence from concept elicitation to confirmation provides the supportive evidence for content validity.

*Construct Validity*

Validity represents the extent to which a PRO measure reflects what it is intended to measure, rather than some other concept. [S2] The process of evaluating the construct validity of PRO measures involve the accumulation of many kinds of evidence which indicate the extent to which the PRO measures the concept(s) it is intended to measure (i.e., physical function, pain interference, depression, anxiety, etc.). The main approaches for evaluating construct validity include concurrent validity, known groups validity, and responsiveness. The evaluation of construct validity is based on confirming various hypotheses concerning the relationship between the targeted PRO and convergent and divergent measures. For example, an assessment of pain interference may examine the correlation between this measure and pain numeric rating scales, disability days, physical function and mobility measures, which are hypothesized to be moderately to strongly correlated. Divergent validity is evaluated by examining the correlations between the target PRO instrument and other measures thought to be unrelated to this measure. Known groups validity evaluates mean differences on the target PRO measure by groups defined by other clinical or patient-rated scales. For example, mean pain interference scores by groups defined by patient global assessments of pain interference, disability days, and pain intensity scores.

*Responsiveness to Change*

Responsiveness refers to the ability of a PRO measure to assess underlying change in the target PRO concept. [S2] [S5] Responsiveness is an aspect of construct validity. PRO measure changes may be evaluated based on changes in clinical status, health event, treatments of know efficacy, and global assessments of change derived from patients or clinicians. For example, patient-rated global change in clinical status may be assessed with a response scale ranging from very much better to very much worse. Responsiveness to change is most frequently measured based on the effect size, standardized response means, and responsiveness statistics. [S2] Analysis of variance methods may be used to compare mean PRO score changes in groups that rate themselves as improved compared with no change and worsening groups. Demonstrating responsiveness to change in clinical status is critical for applying a PRO measure in clinical trials.

Content and construct validity is supported by accumulating evidence supporting validity across multiple studies. The greater the evidence supporting validity across studies, the more confident researchers and clinicians are concerning the construct validity of the PRO measures. Responsiveness is also demonstrated with evidence from multiple studies using multiple methods for identifying groups that have changed in their clinical status.

II. Adaptive Instruments

DEVELOPING ADAPTIVE INSTRUMENTS

Adaptive measures may provide an alternative approach to developing patient or caregiver reported outcome measures for rare disorders. Existing item banks provide pretested and psychometrically sound measures of symptoms and physical, social, and psychological well-being (see, for example, the Patient Reported Outcomes Measurement Information System: PROMIS). [S5] The number of these item banks are increasing, providing pools of items that can be used to populate measures of dimensions targeted for rare disorders. PROMIS methodology to develop instruments is based on decreasing the redundancy of and understanding the breadth of the item content within item banks, using innovative methods, such as Item Response Theory, which allows the identification of the most representative items on the targeted domain. Items from the PROMIS bank were used, for example, to develop a scale for the rare pediatric and adult musculoskeletal disorder fibrodysplasia ossificans progressive. Items from the item bank were chosen with the input of the patient advocacy group for the disorder, which allowed qualitative interviews with patients to help define impact and item meaningfulness.

Some advantages of measures based on existing item banks are (1) availability of a pool of previously evaluated items; (2) usually broad coverage of severity of the domain (i.e., physical function, anxiety, social participation, etc.); (3) ability to target items for instrument development; (4) availability of equivalent language translations for international studies; and (5) availability of short forms for some instruments. Key challenges are common to all instrument development for rare disorders, including (1) limited availability of patients for qualitative and psychometric assessment studies; (2) heterogeneous presentation and course; (3) often need for both pediatric and adult measures; and (4) limited availability of effective treatments.

item pools can be used to create a single, standard disease-specific rating scale that is used uniformly with every subject. A more experimental approach is to develop a truly adaptive questionnaire that would present to subjects only those questions relevant to them, rather than a fixed set of questions, or in a fixed order, as would be expected in a classical assessment. Limitations of truly adaptive questionnaires, where each subject utilizes endpoints tailor-made to their condition, are that they require a completely different analytic approach and that they may introduce other confounds.

Supplement references

S1. FDA. Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims Guidance for Industry. 2009. https://www.fda.gov/regulatory-information/search-fda-guidance-documents/patient-reported-outcome-measures-use-medical-product-development-support-labeling-claims.

S2. Hays R, Revicki D. Reliability and validity (including responsiveness). In Fayers P, Hays R (eds). *Assessing Quality of Life in Clinical Trials, Second Edition*. New York: Oxford; 2005:25–39.

S3. De Vet H, Terween C, Mokkink L, Knol D. *Measurement in Medicine: A Practical Guide*. Cambridge: Cambridge University Press; 2011.

S4. Cappelleri J, Zou K, Bushmakin A, Alvir J, et al. *Patient-Reported Outcomes: Measurement, Implementation, and Interpretation*. New York: Chapman & Hall; 2014.

S5. PROMIS patient reported outcomes measurement information system. www.healthmeasures.net.