

The Role of Cognitive Debriefing and Linguistic Validation in Instrument Development and Modification

Bonnie Teschendorf, PhD
PRO Scientific Advisor

Cognitive Debriefing in Instrument Development

There are two broad uses of cognitive debriefing in the development, modification and utilization of patient-reported outcome instruments in international clinical trials. One use is required as a step in the development of items for a new or modified questionnaire. The second instance is to ascertain evidence of equivalence during translation, linguistic validation, and cultural adaptation of an instrument. Each of these processes use similar methods to interact with study participants, but, they are completed at different intervals in the development and use of a questionnaire.

When a new instrument is being developed for use in a future clinical trial it is necessary for the developer to interview patients with the same condition that will be studied in the trial. This target population is viewed as an expert patient group capable of describing symptoms, health status changes, and functional change for the disease or condition in question. When the interviews are conducted and the analysis completed a draft copy of the items, or questions, can be tested with yet another patient group to see if the ideas resonate with them. The second group of patients, who must meet the same criteria for disease severity as the original population, are asked to review the draft instrument and respond to a pre-set series of questions. These interviews are referred to as cognitive debriefing because of the type of questions being used; structured questions rather than open ended questions. The interviewer asks the participant to provide thoughts about the meaning of each item and to comment on comprehension of the item. The goal is to determine if the participant understands the question in the same way that the original group of patients understood and described it. If the item could be interpreted in yet another way, then it may require modification. It is often quite difficult to word an item in a way that is interpreted as originally intended. Modifications sometimes need to be re-tested with additional cognitive debriefing exercises.

The draft instrument and the cognitive debriefing plan must address several important components. First, the cognitive debriefing sample should include the right patient population; the cognitive debriefing session is best conducted in person, often in groups, and should use a standard interview guide. The interview guide is a well thought out series of questions allowing the subject to ‘think aloud’ about the topics being discussed so that the transcript reflects the way that each interviewee considers the question and responds to it.

Instrument development requires a lengthy and complex process, and is usually a costly endeavor. An important step following the finalization of items and response options is the psychometric validation of the instrument, a step intended to assess the measurement properties of the proposed instrument. Measurement properties include reliability, sensitivity or ability to detect change, and forms of validity. Psychometric assessment is covered in a

separate paper. Content validity, an important part of validation, is unique because it is established based on the qualitative research and cognitive debriefing.

Later Steps in Development Usually Include Linguistic Validation

Linguistic validation uses the original instrument as the source document, yet considers the target language in the resulting reconciliation of words, phrases, and sentences in the items and instruction statements for the language, tone, translation, and meaning of items or semantics in the final version. Several examples of such projects exist in the literature, a particularly interesting one by Quittner et al (2000) offers insight with a project designed to assess the linguistic validation of a new instrument.

Linguistic validation has a long and rich, and well documented history. Some of the original writings on linguistic validation are chapters in *Quality of Life and Pharmacoeconomics in Clinical Trials*, Lippincott-Raven (1996). At the time of these writings there was an attempt to agree on a process and create harmonization among researchers who had begun to use their instruments in international studies. The chapter by Anderson et al (1996) notes that in order to pool data across multinational studies it is essential to be assured that the items in questionnaires mean the same thing to all subjects. Thus, an effort to create a standard for harmonization began in earnest.

Equivalence in a cross-cultural context has been described as having four components (Anderson, 1996). Several instruments have experienced all of these approaches in the development process. These steps include: demonstrating conceptual equivalence; operational equivalence; scale equivalence; and finally, metric equivalence. Conceptual equivalence reflects the item wording and the similarity in meaning from the source version to the target version. It is not simply the literal translation of words but rather the understanding of native speakers that phrases are equivalent. Often this level of equivalence requires that translators involved in a project speak both languages fluently.

Operational equivalence refers to the capacity of the instrument to result in similar outcomes despite being administered in more than one format. This could be the issue when converting a paper-based instrument to electronic or in changing from self-administered to proxy-administered. The versions must be evaluated to determine if they result in the same outcome with similar reliability and validity. These issues continue to be discussed in contemporary scientific circles and have more recently been made explicit in an ISPOR Task Force paper (Coons, et al, 2009).

Construct or scale equivalence reflects the similarity in the cultural adaptation and translated versions of the instrument and how they perform psychometrically. Are both versions equally responsive to change, for example? Metric equivalence is the ability to find the same level of intensity or severity within the scale among subjects with the same disease state.

Some instrument developers use a sequential model for their work and develop the instrument in one language, later developing equivalence for a second or third language. Other developers use simultaneous development methods, creating more than one version from the outset and conducting multiple language versions with comparability, examples using the second method include SF-36 and EUROQoL instruments.

Equivalence also depends on content validity which may be based on semantics (meaning) and was called out in the FDA PRO Guidance document as being critical to the development of a new patient-reported outcome

instrument. The content validity is also important for translation and cultural equivalence, as efforts are made to retain the content validity when translations are made. Native speakers are able to affirm that the meaning is the same for each item. Testing the item during cognitive debriefing with native speakers can affirm that participants agree with the original meaning of the item.

Learning from these early examples of international harmonization efforts helps to understand the continuing refinement of these efforts, a practice that continues to this day. The work that has recently been completed by the ISPOR Task Forces reflects the ongoing efforts to refine the processes and to design harmonized approaches to accomplish the task.

References:

Anderson RT, Aaronson NK, Lepelge AP, Wilkin D. International Use and Application of Generic Health-Related Quality of Life Instruments (Chap 67) in *Quality of Life and Pharmacoeconomics in Clinical Trials*, Second Edition, B Spilker, Ed. Lippincott-Raven Publisher, Philadelphia. 1996.

Coons SJ, Gwaltney CJ, Hays RD, Lundy J, Sloan JA, Revicki DA, Lenderking WR, Cella D, Basch E. Recommendations on Evidence Needed to Support Measurement Equivalence between Electronic and Paper-Based Patient-Reported Outcome (PRO) Measures: ISPOR ePRO Good Research Practices Task Force Report. *Value Health* 2009, 12:4:419.

Quittner AL, Sweeny S, Watrous M et al. Translation and Linguistic Validation for a Disease-Specific Quality of Life Measure for Cystic Fibrosis. *J Pediatric Psych* 2000; 25:6:403-414.

Wild D, Grove A, Martin M, et al. Principles of good practice for the translation and cultural adaptation process for patient-reported outcomes (PRO) measures: report of the ISPOR task force for translating adaptation. *Value Health* 2005; 2:94-104.

Wild D, Eremenco S, Mear I, et al. Multinational trials – Recommendations on the Translations Required, Approaches to Using the Same Language in Different Countries, and the Approaches to Support Pooling the Data: the ISPOR Patient Reported Outcomes Translation & Linguistic Validation Good Research Practices Task Force Report. *Value Health* 2009; 12:430-440.

Willis GB, *Cognitive Interviewing: A Tool for Improving Questionnaire Design*. Sage. 2005. Thousand Oaks, CA.

Dr. Teschendorf is a medical sociologist with experience working in health care settings, academia, and industry. She has expertise in qualitative research and a long standing interest in patient-reported outcomes and electronic data collection.