Advantages and Disadvantages of Using Phase III Clinical Trials for Psychometric Evaluation of New PROs

Dennis Revicki

Dennis A. Revicki, PhD
Senior Vice President, Outcomes Research, Evidera

Patient-reported outcomes (PROs) represent one type of clinical outcome assessment that may be specified as primary or key secondary endpoints for clinical development programs of new pharmaceutical or biotech products. The increasing interest in the patients' perspective in understanding treatment benefits and risks requires PRO measures. PROs identified as primary or secondary endpoints for clinical trials need to have adequate evidence supporting content validity and good psychometric properties (i.e., reliability, validity, responsiveness), and have interpretation guidelines. The U.S. Food and Drug Administration (FDA) guidance on PROs for labeling of medical products provides a summary of the evidentiary requirements that the FDA uses to evaluate PROs as endpoints.[1] New PRO measures are developed following a sequence of qualitative research for concept elicitation with patients and clinicians, careful development of item content and response scales, cognitive interviewing studies to ensure respondent understanding and comprehension of the new instruments, and one or more studies evaluating the measurement properties of the new PRO instrument. Ideally, this psychometric evidence is derived from stand-alone observational studies and/or Phase II clinical trials, so that at the initiation of pivotal Phase III clinical trials, information is available on the reliability, validity, responsiveness, and interpretation guidelines for the target PRO measure.

This article briefly summarizes some of the risks and advantages of developing and evaluating the psychometric characteristics and interpretation guidelines within Phase III clinical trials. The summary is based on previous presentations by Johnson, et al.[2] but reflects the perspective of the author and not necessarily the positions of the other presenters.

At times, sponsors and PRO instrument developers need to deviate from the ideal development and psychometric evaluation approach. In the case of accelerated clinical development programs, products for rare medical disorders, and a mismatch between starting the PRO development studies and the clinical development program, the sponsor may be in a situation where the Phase III clinical trial data is needed for the psychometric evaluation. Clearly, there is often a tension between taking the necessary time to systematically develop and evaluate a new PRO measure and interest and progress toward completing the clinical development program as quickly and efficiently as possible. Deviating from the ideal approach for developing and evaluating the measurement characteristics of new PRO measures presents a number of challenges and potential risks for the pharmaceutical industry sponsor.

Basically proceeding with a Phase III clinical trial with a PRO endpoint with unknown psychometric characteristics is very risky. If the PRO is designated as primary or key

secondary endpoint, this approach is riskier than a clinical outcomes assessment (COA) designated as one of several secondary endpoints. Generally, it is not advisable to have a PRO with unknown psychometric qualities specified as primary endpoint.

In some cases, it may be possible to conduct a psychometric sub-study using only part of the overall clinical trial population. However, there may be challenges associated with conducting and maintaining fidelity of a psychometric sub-study, and the sub-study procedures may impact the conduct of the clinical trial. Some of these challenges may be minimized by limiting the psychometric sub-study to well managed and experienced clinical centers. The psychometric sub-study may involve additional clinical and PRO measures, and may require additional clinical center resources. In addition, this approach may result in a reduction in the clinical trial sample that can be used for efficacy analyses (assuming sub-study patients are not included in efficacy analyses). This issue may be minimized by increasing overall sample size to maintain statistical power for efficacy analyses, but also requires an increase in clinical trial expenditures. Regulatory agencies may be concerned about including the psychometric sub-study participants in the clinical trial efficacy analyses. Regulatory agencies may recommend not including the sub-study data in the clinical efficacy analyses because of concern over potential biases. However, it may be possible to include these data in a sensitivity analysis, thus allowing all clinical trial patients to contribute to the efficacy analyses.

There may be increased risk associated with taking the psychometric sub-study approach for determining the reliability, validity, responsiveness, and especially responder definitions for the new PRO. There is always the potential risk that psychometric analyses may demonstrate that the PRO does not have adequate measurement properties (i.e., reliability, validity, responsiveness). This potential risk can be minimized if attention is paid to the concept elicitation and cognitive interviewing stages of PRO instrument development, with the psychometric evaluation confirming that the developers did a good job in constructing the draft PRO measure. In some cases, it may be unknown whether lack of responsiveness is attributable to treatment or the PRO measure. The analyses may find that estimated responder definition criteria is not demonstrated and/or requires larger sample sizes to adequately evaluate responder definitions, often due to inadequate sample sizes for patients improving, remaining stable, and worsening over time.

There are additional specific challenges associated with defining minimal important difference and responder definitions associated with basing these definitions on analyses of Phase III clinical trial data. Ideally, clinical and PRO data from either stand-alone observational studies or Phase II clinical trials are used to evaluate the psychometric characteristics of a new PRO instrument. Basing the clinical responder definitions on data from Phase III clinical trials may result in bias. If the psychometric analyses can be truly masked to treatment status, it may be possible to determine thresholds for clinical responders.

The ideal situation for evaluating responsiveness to clinically meaningful changes in PRO scores and in identifying meaningful responder thresholds for PRO scores is when some subjects are improving, some subjects remain the same, and some subjects are worsening in clinical status over the course of the study. Although basing the responder definitions on clinical trial data is recommended, there may be additional challenges in some cases. For example, in situations where an active treatment is highly effective (e.g., biologic treatments for psoriasis) and there is only a small placebo group, the resultant analyses may inflate estimates of responder definition and responders. In other cases (e.g., congestive heart failure), where the active treatment is not very effective and with small sample sizes, it may be difficult to identify reasonable responder definitions, and these estimates may be attenuated.

"In situations where it is unavoidable to conduct the psychometric analyses based on Phase III clinical trial data, it is essential to mask psychometricians to treatment groups for psychometric analyses of these studies."

In situations where it is unavoidable to conduct the psychometric analyses based on Phase III clinical trial data, it is essential to mask psychometricians to treatment groups for psychometric analyses of these studies. Decisions about item retention and deletion need to be made without reference to treatment group membership. The usual approach is to provide psychometricians with data files without any reference to treatment group status. In addition, no adverse event data is provided, as these data may potentially be used to identify treatment group, especially if there are specific adverse effects associated with the new treatment. The practice is to provide psychometric analysts with only those data files necessary for conducting the planned psychometric analyses.

An innovative approach to handling the masking problem is to set up an independent psychometric evaluation

committee which is tasked with developing and executing the psychometric analysis, much like a data monitoring committee for some clinical trials. The psychometric committee is organized and charged with completing the psychometric analysis masked to treatment group status. The committee can include psychometricians, clinicians and biostatisticians not directly involved with the clinical trial. This committee is masked to treatment group status, reviews psychometric analyses and makes independent decisions about item retention and deletion, domain structure, reliability and validity, responsiveness, and responder definitions for the PRO measures. A report is generated summarizing and documenting these measurement-related decisions for the PRO endpoints.

For regulatory agencies, risks related to reviewing evidence on the psychometric characteristics of new PRO measures intended as primary or secondary endpoints based on pivotal clinical trials are minimal. For example, the FDA will still hold sponsors to standards of evidence summarized in the PRO guidance on PROs[1] regardless of the source of this evidence. However, the FDA may express concern when decisions about final item content, instrument scoring, and especially clinical responder definitions are based on pivotal clinical trial data. There is always the danger associated with unmasking treatment assignments, and in making decisions that may benefit the active treatment under investigation compared with placebo or other comparative active treatments. Regulatory agencies may not be comfortable with the level of evidence for the PRO measure to make confident decisions about the adequacy of the PRO endpoint (i.e., fit for purpose) and the efficacy of the investigated treatment. Regulatory agencies may come under criticism from sponsors and the public for delaying clinical development programs by recommending additional confirmatory PRO development and psychometric evaluation studies. However, unless scientifically sound and adequate evidence on measurement characteristics of the new PRO are available, it is difficult to make informed decisions on efficacy.

There also may be possible risks to patients and the general public associated with PRO endpoints that may not be developed and psychometrically evaluated based on standard approaches. Study participants may be exposed to adverse effects of treatment unnecessarily in clinical trial with inadequate PRO endpoints. For the general public and health care systems, requirements for additional measurement studies to confirm reliability, validity, responsiveness, and responder definitions may delay clinical development programs and providing access to potentially effective treatments. This situation may be particularly troublesome in cases of rare disorders or other medical conditions (e.g., gastroparesis) where there may not be available effective and approved treatments.

In conclusion, deviations from the ideal approach to systematically develop and evaluate the psychometric properties of new PRO endpoints have some risk to the sponsor. These risks can be mitigated somewhat by recognizing these potential risks and developing strategies to minimize the risks. Certainly ensuring that the psychometric analysis and related decisions about the content, scoring, and responsiveness of the new PRO measure is masked to treatment helps to minimize these risks. The organization of an independent psychometric evaluation committee with established standards and methods may provide further assurances that decisions regarding the PRO measure are made separate from bias related to treatment. PRO endpoints represent important and meaningful assessments for understanding the effectiveness of new treatments. For some medical disorders, PROs are the main approach for evaluating treatment effects, and sponsors and researchers need to ensure that these measures are developed and evaluated to most reliably and validly assess health-related outcomes. ■

For more information, please contact Dennis.Revicki@evidera.com.

REFERENCES

[1] U.S. Food and Drug Administration (FDA).Guidance for Industry Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims (2009). Available at: https://www.fda.gov/downloads/drugs/guidances/ucm193282.pdf. Accessed March 21, 2017.

[2] Johnson LL, Coons C, Chen WH, Revicki DA, Kammerman L. Developing PRO Instruments in Clinical Trials: Issues, Considerations and Solutions. Drug Information Association Statistics Community Webinar, March 17, 2017.