



White Paper

Answering Questions about FDA's Draft Guidance on Patient Reported Outcomes (PROs)

date:
4 December 2008

Summary

In February, 2006 The U.S. Food and Drug Administration (FDA) published a draft of its “Patient-Reported Outcome (PRO) Measures: Use in Medical Product Development to Support Labeling Claims”¹ (hereafter referred to as the Draft Guidance). This document describes how the FDA evaluates PRO instruments used as efficacy endpoints in clinical trials. It also describes the FDA’s current thinking on how sponsors can develop, utilize, and document PRO instruments in a manner that will increase the likelihood of supporting claims in approved product labeling.

This FDA document implicitly gives credibility to PROs as the basis for evaluating drugs and biologics and provides structure for a more efficient, effective and appropriate use of these tools. Challenges exist, however, in interpreting various aspects of the Draft Guidance, which (as of Nov 2008) remains in draft form.

The goal of this paper is to provide answers to questions frequently asked by researchers as they develop their PRO strategies, based on information gathered from the Draft Guidance and FDA presentations and publications that have emerged following its issue.

Background

The Draft Guidance defines Patient Reported Outcomes as the measurement of any aspect of a patient’s health status that comes directly from the patient (i.e., without the interpretation of the patient’s responses by a physician or anyone else). This includes an entire spectrum of patient measurements, ranging from the very specific and direct, such as symptom reports, to broader and more complex constructs, like health related quality of life measures.

The Draft Guidance has had significant impact on how the FDA reviews clinical programs. There are numerous examples of situations where the agency’s Study Endpoints and Labeling Division (SEALD) has reviewed a clinical program based on PROs, and requested documentation of the sponsors’ adherence to the Draft Guidance. Naturally, this impacts the manner in which researchers implement and interpret PROs in support of their global development goals.

The first question to ask may be, “*Why use PROs at all?*” The answer lies in the fact that patient reports are the best available method for getting information on both a disorder and the treatment for the disorder. Even when a disorder-related or treatment-related event is observable, the patient is in the best position to report on it. This may explain why PROs are being used more frequently and often are replacing reports that were traditionally gathered via clinician measurement. For example, in a paper published in Controlled Clinical

Trials (2004)², Wilke et. al reviewed the use of PROs in labeling from 1997 to 2002, and found that of the 215 labels approved during that time period, 30% included PROs; 11% of were approved based on PROs alone.

General Questions about the Draft Guidance

1 Does the Draft Guidance apply to instruments developed and used prior to its release?

Yes. One example of a program that implemented PROs during its clinical development prior to Feb., 2006 which was later viewed through the lenses of the Draft Guidance is Varenicline (Pfizer), a medication approved for smoking cessation in May, 2006, which included PRO-based label claims for:

- Reduction of urge to smoke
- Reduction of nicotine withdrawal
- Reduction of reinforcing effect of smoking

These concepts were measured with PRO instruments that had been used to support claims for previous smoking cessation medications. The Summary Basis for Approval (SBA) explicitly included a review by the SEALD team. While the urge to smoke claim was approved, the claims for withdrawal, craving, and reinforcement were denied. The reasons for denial included that the content validity of the measures was not established, that bias may have been introduced when patients were asked to recall their experiences over a period of time, and that there was an inconsistency of effect (i.e., the overall withdrawal claim was not supported by data that demonstrated all of the individual symptoms of nicotine withdrawal were affected in the same way). All of these concepts are addressed in the Draft Guidance.

2 How can a sponsor receive feedback from FDA on its strategy for administering PROs? What types of questions should be asked of the agency?

There are a number of means by which a sponsor can solicit FDA feedback, including:

- Requesting a Type C meeting
- Requesting written feedback on a briefing book
- Submitting the PRO strategy as part of a scheduled meeting

The PRO measurement issues that a sponsor may raise with the FDA will vary from program to program. The following are examples of issues that could be important to address with FDA.

Concepts/Conceptual Model

In order to identify concepts that may be important to target with PRO instruments, it can be useful to create a conceptual model that outlines the relationship between the compound and various impact areas. Sponsors may choose to request feedback from FDA about the relevance of the concepts in the target population and the appropriateness of targeting these concepts with PRO instruments.

Measurement Strategy/Content Validity

It may be possible to seek FDA feedback on a measurement strategy, including the details of how a PRO instrument would be selected, developed, and/or administered, (i.e., how, when, where and how often). Depending on the stage of a particular program, it may be appropriate to present a 'dossier' to the agency that outlines the measurement characteristics of a particular instrument.

FDA has emphasized the importance of content validity in developing and evaluating PROs. In this context, content validity is established when:

- Concepts that are addressed by PRO measures are important to that patient population
- Concepts comprehensively address the domain of interest
- Patients understand the meaning of questionnaire items and responses

Qualitative patient input collected through interviews with patients is critical in addressing these dimensions of content validity.

Endpoint Development

A sponsor may also request feedback on its overall plan for establishing and documenting endpoints in a program. This development process is captured in the FDA's "wheel and spokes diagram"³ that is presented in Figure 1.

In order to facilitate endpoint development, PRO Consulting® has developed a systematic method for developing endpoints that:

- identify the relevant PRO and other endpoints associated with a specified drug and/or clinical program
- inform the appropriate measurement strategies necessary to accurately evaluate those endpoints, and;
- provide documentation that is consistent with recommendations provided in the Draft Guidance.

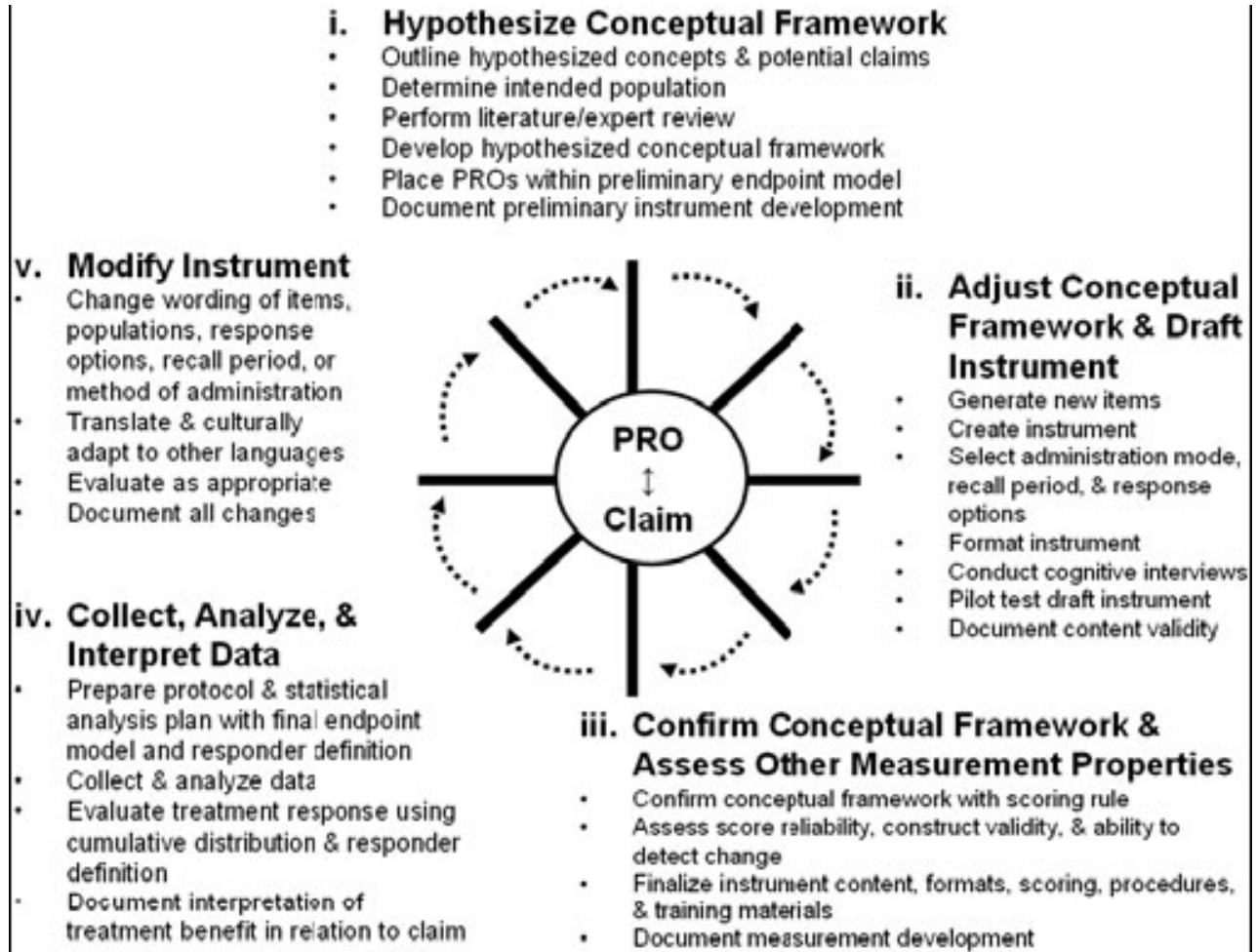


Figure 1. Wheel and Spoke Diagram³

PRO Instrument Modification and Validation

1 Does changing the order of PRO instrument items or deleting portions of the questionnaire alter the way a patient responds to a questionnaire?

It might. If these types of changes are made to a PRO instrument, it is advisable to ensure that patients understand and respond to the items in the same manner as in the original instrument. This may be done through qualitative and/or quantitative methods.

2 *If an instrument is modified, is evidence supporting the psychometric properties of the modified measure required?*

Yes. However, again, there is not a “one size fits all” answer. To quote the Draft Guidance, “The extent of additional validation recommended depends on the type of modification made,” (lines 582-583). To better understand this, it’s important to consider the various scenarios where a sponsor might make changes to a measure, instead of using it in its original form, some of which include:

- Using an established PRO instrument with a brand new population;
- Reducing the length of items or reducing the length of the entire measure;
- Administering only a specific sub-domain within a broader or generic measure;
- Changing a response scale.

There is no easy way to determine what is required when any of these, or other forms of modifications are made. Determining the extent of additional validation requires careful consideration of how the changes may impact patients’ understanding of the items.

3 *Is it a requirement that development, modification and validation of a PRO instrument occur in a linear manner?*

We believe that the answer is no, based on the ‘Wheel and Spokes’ diagram (figure 1) as outlined in the Draft Guidance. The development of PROs is an iterative process, where information from patients or data from a clinical trial may cause sponsors to consider modifying a PRO instrument or strategy.

4 *Should a sponsor conduct validation activities as past of its phase III study?*

We believe that the answer is no. The primary reason is that it is simply good science to validate a measure prior to using it as an endpoint in pivotal trials. It also increases risk to the sponsor, as the data may not confirm the psychometric ‘soundness’ of the instrument. It is possible to leverage Phase II studies to establish the validity or other psychometric characteristics of the instrument, or to run a separate study in order to generate that information. Data from Phase III studies can be used to supplement existing data regarding scale performance.

Migrating to Electronic PRO Measures

1 When migrating a paper and pencil measure to an electronic platform, which strategies should be used to ensure that patients are using the electronic measure appropriately?

A lot of work has been done to determine what level of revalidation is required for paper-electronic migrations, including a 2008 review of over 250 studies that quantitatively addressed the association between paper and electronic versions of the same questionnaire⁴. This meta-analysis found that overwhelmingly, paper and electronic modes of administration are equivalent. That finding set the framework for a hierarchy of validation requirements based on various modification levels (see Table 1). The primary conclusion is that when migrating to electronic, if the modifications to the items or responses are minor, cognitive debriefing and usability testing would be required. However, if more substantial modifications are being made in which the changes may be expected to alter patients’ understanding of the items or responses then it may be advisable to conduct quantitative equivalence tests (see next question).

Table 1. Determining Level of Validation

| Modification | Rationale | Examples | Level of Evidence |
|--------------|--|--|--|
| Minor | The modification can be justified on the basis of logic and/or existing literature. No change in content or meaning. | Non-substantive changes in instructions (e.g., from circling the response to touching the response on a screen). | <ul style="list-style-type: none"> • Cognitive debriefing • Usability testing |
| Moderate | Based on the current empirical literature, the modification cannot be justified as minor. May change content or meaning. | Change in mode of administration involving different cognitive processes (e.g., paper [visual] to IVR [aural]). | <ul style="list-style-type: none"> • Equivalence testing • Usability testing |
| Substantial | There is no existing empirical support for the equivalence of the modification and the modification clearly changes content or meaning | Substantial changes in item response options Substantial changes in item wording | <ul style="list-style-type: none"> • Full psychometric testing • Usability testing |

2 *What is the difference between Cognitive Interviewing, Usability Testing, and Equivalency Testing?*

The goal of Cognitive Interviewing is to ensure that patients understand the items in the electronic platform in the same way they understand the items in the paper and pencil measure. The method would require a relatively small sample of the target patient population. Typically an experimenter sits right next to the patient and asks him/her to think out loud as they are completing the measure. As the patient verbalizes the items and responses the experimenter can document how the patient encodes and responds to the items.

In Usability Testing, the goal is to ensure that patients can navigate the software and electronic device in its intended manner. For example, it would be important to show that patients know how to move from one screen to another and that they can access all response options for every item. The sample size for Usability Testing depends on the complexity of the software and how significant the paper to electronic migration is; small changes require only a small sample size, whereas larger changes may require significantly more.

In formal Equivalence Testing, patients from the target population would complete both the paper version of the measure and the electronic version of the measure and then a quantitative index of that association would be derived, such as the intraclass correlation coefficient.

3 *What needs to be included in a Cognitive Debriefing Report to the FDA?*

The content of a cognitive debriefing report will vary depending on the goals of the program. However, if developing a new instrument or confirming the appropriateness of an existing instrument, it is important to include at least a couple of key elements. First, it is important to include an item tracking matrix, which shows an initial 'long list' of items and how the items were altered or deleted through the process of developing the measure. Second, it is important to demonstrate that saturation was reached during the interviewing process. Saturation means that the inclusion of additional patients would not have resulted in new information. Although not required, FDA may request to review transcripts of the interviews, so it is important to maintain this documentation.

Conclusion

One of the most significant accomplishments of the Draft Guidance is to establish scientific standards for the development and use of PROs in clinical development programs. These standards increase the likelihood that PROs will accurately address important aspects of the patient's experience. Following these standards may increase the likelihood of successfully using PROs to support label claims.

References

- 1 Federal Register, Vol. 71, No. 23; Friday, February 3, 2006
- 2 Wilke, Burke, Erickson. Measuring treatment impact: A review of patient reported outcomes and other efficacy endpoints in approved product labels. *Controlled Clinical Trials*, 2004; 25:535-552.
- 3 Patrick DL, Burke LB, Powers JH, et al. Patient-reported outcomes to support medical product labeling claims: FDA perspective. *Value Health* 2007; 10 Supp. 2: S125-137.
- 4 Gwaltney CJ, Shields AL, Shiffman S. Equivalence of electronic and paper-and-pencil administration of patient-reported outcome measures: A meta-analytic review. *Value Health* 2008; 11:322-33.