

Alan Goldhammer, PhD

Associate Vice President,
US Regulatory Affairs



April 4, 2006

Dockets Management Branch (HFA-305)
Food and Drug Administration
5630 Fishers Lane, Room 1061
Rockville, MD 20852

Re: Docket Number 2006D-0044; Draft Guidance for Industry on Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims; 71 Federal Register 5862; February 3, 2006

Dear Sir or Madam:

The Pharmaceutical Research and Manufacturers of America (PhRMA) is pleased to submit these comments on the Draft Guidance published by the Food and Drug Administration (FDA) on February 2, 2006 titled "Patient-reported Outcomes Measures: Use in Medical Product Development to Support Labeling Claims" (Draft Guidance). PhRMA represents the country's leading research-based pharmaceutical and biotechnology companies which are devoted to inventing medicines that allow patients to lead longer and more productive lives. PhRMA and its member companies have a significant interest in the provisions of the Draft Guidance and its potential significance on the design and interpretation of clinical trials. PhRMA's detailed comments are set forth below.

General Comments

PhRMA congratulates the FDA on its efforts to address the important topic of patient-reported outcomes (PRO) measures in medical product development to support labeling claims.

PhRMA appreciates the FDA's willingness to discuss the development and use of PRO measures in the context of product development. As this is an area that has received little attention in past dialogs between FDA and the sponsors, we suggest that clarification of expectations regarding the timing, content, and participants in such discussions would be helpful for planning purposes.

The PRO Draft Guidance proposes that the sponsor provide extensive information regarding the development and performance of each PRO instrument to the FDA. It would be very useful to have a template to guide such submissions to ensure that appropriate topics are covered in a format that is efficient and effective. PhRMA would be pleased to work with FDA to produce such a template.

While PhRMA recognizes the importance of providing evidence of reliability and validity of PRO measures used in labeling, we believe it is important that there be a balance between what may be considered ideal evidence and practical limitations of what can be accomplished in the context of the clinical trial setting. The Draft Guidance covers a number of topics that may be more or less relevant to a specific stage of product development. Evaluation of validity should not be a checklist, but rather evolves over time as new instruments are being developed or

Pharmaceutical Research and Manufacturers of America

950 F Street, NW, Washington, DC 20004 • Tel: 202-835-3533 • FAX: 202-835-3597 • E-Mail: agoldhammer@phrma.org

modified. Presentations given by the FDA on this Draft Guidance suggest that there is more flexibility than the Draft Guidance would indicate. It would be helpful to clarify that such flexibility underlies the application of the Draft Guidance.

The Draft Guidance suggests that PRO instruments will be held to a much higher standard than other endpoints. It would be helpful to know what efforts the FDA will make in revising the Draft Guidance to ensure that the playing field is more equal between PRO measures and other endpoints used in clinical trials, e.g., surrogate endpoints, hybrid clinical disease measures. We urge the FDA to consider the cost as well as the benefits of higher standards – for any measures – in a way that fulfills the goal of the Critical Path Initiative to optimize the provision of new drugs and drug-related information to patients.

Specific Comments

Section I. Introduction

A PRO measurement is defined as “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 any else).” This definition is quite limited, since it excludes those who cannot speak for themselves (e.g., cognitively impaired persons, infants and young children). We suggest that the definition be modified to read, “A PRO is a measurement of any aspect of a patient’s health status that comes directly from a patient, or when patients are unable to speak for themselves (e.g. cognitively impaired), is provided by immediate caregivers.” (lines 31-32) This is also more consistent with the comment on lines 694-699 that indicate an assessment by a proxy may be appropriate. In addition, there are circumstances when an interview, by telephone or in person, is the best way to collect these data, and the interviewer may be a clinician. We recommend adding “The responses to a PRO instrument may be directly recorded by the patient or a response provided to an interviewer who reads the item to the respondent and records, but does not interpret the response”.

Section II. Patient-Reported Outcomes – Regulatory Perspective

This section provides important information but does not appear to provide a specific regulatory perspective. Thus, the section heading is somewhat confusing and PhRMA suggests that renaming this section would provide clarity.

Throughout the document the words effectiveness and efficacy are used interchangeably. The meaning of these terms may vary across professional groups. Therefore, PhRMA suggests the terms be defined in the Draft Guidance.

The Draft Guidance indicates that PROs used in a clinical trial to support efficacy should measure adverse consequences of treatment separately from the effectiveness of treatment. PhRMA recommends that FDA consider modifying this statement because in some situations patients may have difficulty distinguishing between the two, for example, in the case of fatigue in cancer patients. Furthermore, in some cases it could be valuable to quantify the net health impact of a treatment and to include such information in product labeling. There are well-validated PRO measures that assess effectiveness and adverse events together. In addition, possibly at this point in the Draft Guidance, it would be most useful to clarify that PRO instruments are not adverse event collection instruments as generally used for safety labeling

purposes and should not be reconciled with the standard spontaneous adverse event collection that is part of all clinical trial protocols (with the potential exception of serious adverse events).

Section IV. Evaluating PRO Instruments

The Draft Guidance (lines 178-181) states that FDA generally plans to “evaluate the modified instrument just as it would a new one. Therefore, in such instances we encourage sponsors to document the original development processes, all modifications made and updated assessments of its measurement properties.” PhRMA believes that this approach is too stringent and not pragmatic for minor changes to an instrument. In many cases minor modifications to an existing instrument would not require substantial revalidation such as re-establishment of measurement properties. This same idea is covered under Section IV B-6, where PhRMA notes that not all changes listed may be important enough to warrant a re-validation study. In many cases a cognitive debriefing study may yield sufficient evidence to show that the changes do not impact responses. We suggest modifying the sentence above to read: “When considering an instrument that has been significantly modified from the original, the FDA generally plans to evaluate the modified instrument as it would a new one. Minor modifications may not require revalidation or the extent of validation required for minor modifications may be substantially less.”

Section IV A. Development of the Conceptual framework and identification of the Intended Application

Figure 1 of the Draft Guidance shows the PRO instrument development and modification process. While PhRMA agrees that this figure represents the ideal process, it is worth noting that existing instruments, often developed by academic researchers, may not necessarily have followed all the development steps outlined in Figure 1, yet their appropriateness for use in a given patient population may have been demonstrated over time in numerous clinical studies. In such cases, it would be inappropriate to require re-validation, or additional development work to fulfill the steps outlined in Figure 1. PhRMA recommends that the agency address how it will evaluate the use of existing PRO measures in clinical trials for use in labeling and promotion since this is the most frequent use of PRO measures in drug development.

The Draft Guidance notes (lines 231-234) that “if improvements in a score for a general concept (e.g., physical function) is driven by a single responsive domain (e.g., symptom improvement) while other important domains (e.g., physical abilities and activities of daily living) did not show a response, a general claim about improvements in physical function would not be supported.” The Draft Guidance further notes (lines 951-954) that “if analysis of scores for the individual component endpoints of a composite shows the improvement is driven primarily by a single domain (e.g., performance of a specific activity), the findings for the composite score would not support a general claim (e.g., psychological or emotional benefit, or even general physical state if all that is shown is symptom improvement).” Does this imply that a narrower claim would be allowed?

The Draft Guidance describes what appears to be a measurement model (lines 249-261, including Figure 2), yet it is referred to as a conceptual framework. It is not clear why new terminology is being introduced, and PhRMA believes that it would be preferable to use the

traditional “measurement model” terminology in order to avoid confusion. If the FDA intends Figure 2 to be different from a measurement model, this difference needs to be clarified.

The Draft Guidance also states (lines 275-279) that FDA plans to compare the patient population used in the PRO instrument development process to the study populations enrolled in clinical trials to determine whether the instrument is appropriate to that population with respect to patient age, sex, ethnic identity, and cognitive ability. PhRMA recognizes that the need for comparability between the PRO instrument development population and the target population for the clinical trial. However, the degree of correspondence implied is unclear and there is concern that the list is too specific. Furthermore, the clinical trial population itself is a subset of the population who will receive marketed products and it is unrealistic to expect that the population studied during instrument development will be completely comprehensive to either the clinical trial or post-marketed population. In addition, the potential need to re-validate and/or adapt instruments to account for minor differences in populations would lead to a proliferation of instruments and instrument versions. This could move the field away from standards and normative data that is highly desired for industry and FDA endpoints. The extent of comparability may vary depending on the condition being studied. PhRMA believes that it would be preferable to state more generally that FDA plans to compare the patient populations in the PRO development process and the target populations to ensure that they are similar and that the instrument is applicable to the clinical study population.

The Draft Guidance (lines 298-300) states: “The FDA plans to review instrument development (e.g., results from patient interviews or focus groups) to determine whether adequate numbers of patients have supported the opinion that the specific items in the instrument are adequate and appropriate to measure the concept.” It remains unclear what is adequate from the FDA’s perspective. If it is a “rule of thumb” value, please elucidate. If it is not a “rule of thumb” value, what factors will be used to define this value?

The Draft Guidance states (lines 307-308) “Of course, it would be critical to know that each item refers to something that patients actually do.” The meaning of this statement is unclear. What if some (or many) patients cannot do the activity in question? This should not necessarily invalidate an item. For example, the item may be geared towards achieving a higher level of functioning than is attained by most patients with the disease. This item might be useful in expanding the bandwidth of the measure (e.g. distinguishing between good functioning and excellent functioning). PhRMA suggests rewording the sentence as follows: “It may be useful to know whether each item refers to something that patients actually do.”

The Draft Guidance addresses the issue of patient recall (lines 339-343) stating “PRO instruments that require patients to rely on memory, especially if they must recall over a period of time, or to average their response over a period of time may threaten the accuracy of the PRO data. It is usually better to construct items that ask patients to describe their current state than to ask them to compare their current state with an earlier period or to attempt to average their experiences over a period of time.” This is an extremely complex topic with many unresolved issues. Furthermore, there is insufficient scientific evidence to support this statement. PhRMA suggests that the recall period should be informed by the disease and the question being asked. If the averaging of items has adequate psychometric properties, and averaging is justified conceptually, then the use of items that average across time should be considered acceptable

Table 2 describes various response options used in PROs. The description of Visual Analogue Scale (VAS) includes a sentence that states, "These scales often produce a false sense of precision." PhRMA suggests deletion of this sentence as there is no scientific evidence to support it.

Section IV B. Development of Format, Instructions and Training

Under Section IV B-6 (lines 380 to 395), the Draft Guidance implies that any format, instruction or training changes to a PRO would require re-validation. In many cases instructions are improved from the original to clarify directions for the respondent. Other changes can include minor formatting changes. PhRMA believes that many of these changes may not warrant revalidation of the instrument, but rather a simple cognitive debriefing to ensure that the changes would have little impact on patient responses. PhRMA believes that it would be helpful for the FDA to clarify this in the Draft Guidance.

Section IV B-7 "Preliminary Scoring of items and domains" states, "Equally weighted scores for each item are appropriate only when the responses to the items are relatively uncorrelated. Otherwise, the assignment of equal weights will overweight correlated items and underweight independent items. Even when items are uncorrelated, assigning equal weights to each item may overweight certain items if the number of response options or the values associated with the response options varies by item. The same weighting concerns apply with added complexity when combining domain scores into a single overall score." (lines 416-422) Since the available evidence suggests that weighting of items does not influence scoring, PhRMA suggests that lines 416-422 be deleted from the Draft Guidance.

Section IV C: Assessment of Measurement Properties.

In Table 4: Measurement Properties Reviewed for PRO Instruments used in Clinical Trials, the ability to predict future outcomes (also known as predictive validity) is listed. PhRMA believes that a requirement for predictive validity to be established may be unrealistic for many PROs that do not have sufficient longitudinal data. To satisfy this criterion would unnecessarily prolong clinical trial programs. PhRMA recommends that predictive validity be listed in Table 4 as a PRO measurement property that is desirable but not necessary to have in order for a PRO instrument to be considered sufficiently validated to support a claim.

The Draft Guidance states (table 4 line 483 for the ability to detect change) "Ability to detect change is always specific to a time interval." It further asks: "Has ability to detect change been assessed for the time interval appropriate to study?" These statements imply that responsiveness must be evaluated over the same time interval as that of the clinical trial. It may not be practical to assess the ability to detect change over the same time interval, particularly when that interval is long. In such circumstances the change in the PRO instrument seen in the trial should provide sufficient evidence that the instrument is responsive.

Also in Table 4, under "Minimally important difference" (line 483) the Draft Guidance states that the FDA is specifically requesting comment on the appropriate review of derivation and application of responder definitions when used in clinical trials. In response to this request, PhRMA presents the following comments.

Defining responders may depend on a specific treatment, disease or patient population and methods to establish responders have not been well established. Further research is needed to establish best practices and standards to determine 'responders'. Until then, PhRMA recommends that a pragmatic approach be taken to implementing any research strategy to determine this outcome in clinical trials.

PhRMA recommends that the FDA remain flexible with respect to the methodology used for establishing a minimally important difference (MID) since no single approach is universally accepted or appropriate. Table 4 (line 483) suggests that a single approach or combination of approaches may be acceptable, yet in lines 566-567, the Draft Guidance recommends that multiple methods be used to discover whether there is concordance among methods confirming the MID. PhRMA recommends that the FDA revise language in the Draft Guidance to reflect flexibility in evaluating approaches to establishing MID.

The Draft Guidance states (lines 523-524) "If there is evidence that PRO scores are affected by changes that are not specific to the concept of interest, the validity of the PRO instrument should be questioned." Scores on PRO measures may change due to events that are not directly related to the measure. This does not necessarily invalidate the scale. For example, a patient may suffer from a significant physical injury (e.g., a broken leg), yet in addition to reductions in scores measuring physical functioning, one would not be surprised to see reductions in scores on measures of psychosocial and/or mental health. This should not invalidate those psychosocial and/or mental health scores. PhRMA suggests this sentence be deleted.

Section IV D. Modifying an Existing Instrument

The Draft Guidance Section IV D includes language that is very prescriptive and recommends re-validation of PROs for even minor modifications. PhRMA believes that it is unnecessary for sponsors to undertake small randomized studies to revalidate an instrument where there are only minor modifications. In cases where there are minor modifications to an existing instrument, cognitive testing should be sufficient. PhRMA recommends that sponsors justify the scope of re-validation required for a modified instrument early in a development program

Section V. Study Design

The Draft Guidance suggests that open label studies are rarely credible (line 718). There are a number of instances when blinding is not possible such as clinical trials in oncology or medical devices. However, this does not mean that the study loses credibility. PhRMA recognizes that open label studies are not ideal, but they should not be ruled out entirely, particularly when there is a high demand for active treatment comparative trials that are often unable to be blinded. We support the statement that blinding and identifying a control group is not always feasible (lines 735-738); therefore, a sponsor should consult with the FDA on such issues when needed.

Section V F. Specific Concerns When Using Electronic PRO Instruments

PhRMA recommends that the requirement for investigators to have 'direct control' over the electronic source record be eliminated and replaced with a requirement that the "technology

must ensure that only the investigator can maintain the electronic source records. " (Lines 830-834)

A reasonable interpretation of 'maintain' is that the technology must ensure that only the investigator can create, modify, or delete the records and has continuous access. The technology must also ensure that the sponsor cannot create, modify, or delete source records. A part 11 compliant audit trail is essential to confirm this.

The rationale for this request is that a party vendor or technology provider can ensure that appropriate controls are established to prevent and detect unauthorized access and/or changes to the source data. Proof (validation and audit trail) of these controls can be made available to FDA in an inspection. Although the investigator does not directly manage these controls, the end result is that it creates a verifiable environment where only the investigator can maintain electronic source records.

Section VI. Data Analysis

A. General Statistical Considerations

The Draft Guidance states on lines 871-872 that "unplanned or post hoc statistical analyses are usually viewed as exploratory and, therefore, unable to serve as the basis of a claim of effectiveness". We agree that patient-reported outcome measures that will serve as the basis of a claim should be pre-specified in study protocols and statistical analysis plans. However, there may be situations where post hoc statistical analyses are appropriate to clarify results or more fully understand the benefits to the patient. We propose that the FDA further expand this statement to allow for the possibility that, as is the case with clinical endpoints, unplanned analyses may be necessary and appropriate to support label claims.

C. Statistical Considerations for composite measures

The Draft Guidance states on lines 940-942 that when using composite endpoints, "it is critical to ensure that patients enrolled in a clinical study are impaired in all domains". However, it is not necessarily a requirement, when using composite endpoints based on a combination of clinical criteria, that all patients be impaired for all criteria at baseline. PhRMA believes that the Draft Guidance should provide greater flexibility and allow for discussions between the sponsor and the reviewing division on the appropriate entry criteria and sample size requirements for studying treatment effects on composite measures. Otherwise, the FDA would appear to be proposing a higher standard for PROs than for other clinical endpoints

D. Statistical Considerations for Patient-Level Missing Data

The Draft Guidance recommends that procedures for handling missing data be pre-specified in study protocols and statistical analysis plans. Some procedures may be pre-specified appropriately (e.g., handling missing items within domains). However, in other instances, the pre-specification of rules for handling missing data may be inappropriate without an understanding of why the data are missing. PhRMA believes that a more appropriate approach may be to emphasize the need for sponsors to conduct a thorough evaluation of such missing

data prior to the analysis of PRO measures between treatment groups. The quantity of missing data and associated reasons (e.g., missing completely at random, missing at random, missing not at random) should be described. The sponsor should then identify the most appropriate approach, or approaches, for handling these missing data once the patterns are known.

On line 1012, the Draft Guidance appears to recognize the need for flexibility as it recommends that when “a higher proportion of patients have missing data, the FDA recommends the use of several different imputation methods (including a worst-case scenario in which missing data are assumed to be unfavorable for those on the investigational treatment and favorable for those in the control group).” While we agree with the FDA on the need for flexibility, we suggest that alternate methods be based on scenarios that are clinically and medically reasonable, rather than hypothetically possible, but highly unlikely. We, therefore, disagree with the requirement of a worst-case scenario in which missing data are assumed to be unfavorable for those on the investigational treatment and favorable for those in the control group.

E. Interpretation of Study Results

PhRMA recommends clarification of the section describing the need for minimally important difference (MID) information. On line 1023, the FDA suggests that the MID may serve as a benchmark for interpreting the clinical importance or relevance of study results. This appears to contradict an earlier statement in Line 886 in which the FDA states, “The statistical analysis considerations for PRO endpoints are not unlike statistical considerations for any other endpoint used in drug development”. Since statistically significant differences in clinical endpoints are sufficient for a claim, we suggest clarification of whether or not there is an additional hurdle for PROs versus other clinical endpoints.

In summary, PhRMA supports many aspects of the Draft PRO Guidance. We feel that it provides a step forward in recognizing the importance of assessing the patient’s perspective in drug development. However, we believe that due to a variety of factors such as the diversity of disease conditions and treatments as well as the lack of consensus on some topics, e.g., methods for identifying the MID, assessment processes need to remain flexible rather than fixed. We welcome the opportunity to discuss this important topic with the agency at each stage of product development and support this process becoming an established part of the development process.

Sincerely,

A handwritten signature in cursive script, appearing to read "Alan Goldhammer".