

Wyeth Pharmaceuticals Inc.
P.O. Box 8299
Philadelphia, PA 19101-8299

Wyeth

1797 6 APR -5 19:19

April 04, 2006

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

RE: Docket No. 2006D-0044, February 3, 2006 (71 FR, 5862-5863)

Dear Sir/Madam:

Wyeth Pharmaceuticals is submitting the following comments on the FDA's Draft Guidance for Industry entitled, "Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims."

Wyeth is one of the largest research based pharmaceutical and healthcare products companies and is a leading developer, manufacturer, and marketer of prescription drugs, biopharmaceuticals, vaccines, and over the counter medications. Wyeth appreciates the opportunity to comment on the above-mentioned Draft Guidance; our comments are provided below.

Wyeth supports the development of this FDA Guidance, as we believe there is a need to formally document FDA's current thinking on this topic. Also, we commend the Agency on its efforts in preparing a thorough and well-reasoned Guidance for the industry.

General Comments

The Guidance for Industry "ICH E9 Statistical Principles for Clinical Trials" provides commentary on several topics addressed in this Draft Guidance (e.g., treatment of missing data, creation of composite measures). We recommend that a cross-reference to ICH E9 be included where appropriate.

In addition, the document should allow for a varying degree of evidence required depending on how well established a Patient Reported Outcome (PRO) instrument is or how novel it is. Strict evidentiary requirements should only apply to entirely new, never-before-validated PRO instruments. Older, well established and accepted PROs or modifications of older PROs should not require such extensive re-validation.

2006D-0044

C 11



Specific Comments

I. Request for Clarification in the “Background” Section

The Background (Section II, lines 74-79) states that the use of a PRO instrument to support a labeling claim requires the same amount and kind of evidence as that required for any other labeling claim. However, this Section does not take into account the amount of evidence needed when using an instrument that has been in wide use for a long time versus a novel tool, which may require more data. We believe that this Section should include flexibility for the amount and kind of evidence needed.

We recommend that line 79 be revised to add the following statement “However, the amount and kind of evidence to assess the ability of the PRO instrument to reliably and validly measure the specific outcome at issue may vary depending, for example, on whether the intended use is long-established as valid or is entirely novel for this indication.”

II. Request for References

The Guidance does not contain references for concepts that have originated from the field of psychometric/survey research. Although some of them are widely accepted, it is encouraged that appropriate references be provided where appropriate, particularly for concepts that may not be widely accepted. Specifically, references are requested to support the following:

- Section IV.B.3, lines 339-343. The “Choice of the Recall Period” states “PRO instruments that require patients to rely on memory, ... or to average their response over a period of time may threaten the accuracy of the PRO data.” While this may be accurate for some PRO instruments, there are circumstances where a shorter recall period may not provide an accurate representation of the patients’ experience (e.g., once-a-day versus once-a-month). In general, it should be up to the study sponsor to justify the recall period (duration and frequency) chosen¹.

We recommend that a reference for this statement should be provided. If none are available, we recommend that this statement be deleted.

- Section IV. B. 4, Table 2, line 351. The Description for, “Visual analog scale (VAS)” states that, “These scales often produce a false sense of precision.”

¹ Reference is made to lines 328-329, which states that the sponsor should evaluate the rationale and the appropriateness of the recall period.

Wyeth

We recommend that a reference for this statement should be provided. If none are available, we recommend that this statement be deleted.

- Section V. A. 1, lines 717-718. The “Blinding and Randomization” Section states that PRO measures collected in open-label studies are “rarely credible.”

We recommend that a reference for this statement should be provided. If none are available, we recommend that this statement be deleted.

- Section VI. B. 7, lines 416-419, “Identification of Preliminary Scoring of Items and Domains” states that, “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.” In addition, lines 426-429 states that it is tempting to use the same weights in a clinical trial setting to demonstrate treatment benefit.

We recommend that a reference for these statements be provided. If none are available, we recommend that these statements be revised or deleted.

III. Request for Clarification Regarding Evaluation/Monitoring Adverse Events

Section III. B., line 164 (Table 1), “Taxonomy of PROs Used in Clinical Trials” includes the evaluation of adverse events in the “Intended use of the measure” attribute. Similarly, Section IV. A. 2 (Identification of the Intended Application of the PRO Instrument), lines 269-271 states that the PRO instrument can be developed for a variety of roles including monitoring adverse events. While use of a PRO instrument can provide important information on the patients’ experience, it should not replace the clinician as being the primary reporter of adverse events.

Please confirm that the intent of these Sections is to supplement the clinician’s evaluation of adverse events and to provide another mechanism to characterize a drug’s overall benefit to risk assessment.

IV. Request for Examples Regarding Acceptable Documentation for a Previously Developed Instrument

Section IV. C includes guidance regarding the creation of a new PRO instrument and Section IV. D. includes guidance regarding the modification of an existing instrument. However, the Guidance does not include examples of the type(s) of

Wyeth

documentation (e.g., publication) that would be considered acceptable to demonstrate the proven validity for older, well-validated instruments.

We recommend that examples be included on the type(s) of documentation that would be considered acceptable to demonstrate the proven validity for a previously developed instrument.

V. Request for Consistency of "Evaluating PRO Instruments" Section

The introduction of Section IV, "Evaluating PRO Instruments" states in lines 178-179 that when considering an instrument that has been modified, the FDA generally plans to evaluate the modified instrument just as it would a new one. This language seems to be more strict and inconsistent with Section IV. D., "Modification of an Existing Instrument" (beginning on line 579) which states, for example, "The extent of the additional validation ...depends on the type of modification made." We believe that the level of scrutiny and quality of evidence required should depend on whether the PRO is: (1) widely-accepted and already validated for this use, (2) a modified use of a widely accepted, validated instrument, or (3) entirely novel. The amount of evidence required for validation may vary, depending on the situation, from merely citing the existing validation studies (e.g., in published, peer-reviewed literature) to providing new data from an adequate controlled study that uses the modified or novel PRO.

We Recommend that the following revisions be made:

- *Revise line 178 to the following: "When considering an instrument that is new or has been modified...."*
- *Revise line 179 to add, "The level of evidence required for new or modified PROs may vary depending on the degree of novelty or the modification" after "...as it would a new one."*

VI. Request to Generalize Criteria for Identification of Intended Population

Section IV. A. 3, lines 275-279, "Identification of the Intended Population" states that the FDA plans to compare the patient population used in the PRO instrument development process to the study population with respect to age, sex, ethnic identity and cognitive ability. While we agree that comparability between the PRO instrument development population and the target population for the clinical trial should be assessed to ensure that the populations are similar, we believe that the list of variables that the FDA intends to review (age, sex, ethnic identity and cognitive ability) is too specific. This statement should be revised to reflect that the purpose of the FDA's review is to ensure that the two groups are appropriately similar.

Wyeth

We recommend that line 275 be revised as follows, "FDA will review the PRO instrument development population and the target population for the clinical trial to ensure that the two groups are similar". Alternatively, we recommend that line 277 be revised to add "for example" prior to "age, sex, ethnic identity and cognitive ability."

VII. Request for Clarification of Patient Understanding of Recall Period

Section IV. B. 3., lines 332-334, "Choice of Recall Period" states that FDA will review study protocols to determine what steps will be taken to ensure that patients understand when they should complete the PRO instrument and what steps will be taken to assess their compliance. We believe that assessing compliance with paper-based instruments is impractical and would impose a large new burden on sponsors.

We recommend that line 334 be revised to include an example(s) of how to assess compliance for a paper-based instrument that is not burdensome to the patients.

VIII. Request for Clarification of Evaluation of Patient Understanding

Section IV. B. 5, lines 373-375, "Evaluation of Patient Understanding" states that Sponsors are encouraged to assess patients' understanding of the items in the PRO instruments. The Guidance further states that the FDA's evaluation is likely to include "...a review of a cognitive debriefing report containing a readability test, the script used in the debriefing interviews, the interview transcript, the analysis of the interview results, and the actions taken to remedy any items that created problems." While we believe the intent of this statement relates to the validation of the instrument rather than the use of the instrument, the Guidance is unclear.

Please confirm that this statement is specific to the validation of the instrument rather than the use of the instrument.

IX. Request for Revision of the "Development of Format, Instructions, and Training" Section

Section IV. B. 6, lines 383-394, "Development of Format, Instructions, and Training" seems to include recommendations specific to the modification of a PRO instrument. While we believe these recommendations are important, we also believe they would be more appropriate in Section IV. D., "Modification of an Existing Instrument" (line 662, Other Changes) rather than in the "Creation of a new PRO Instrument." For example, line 394 refers to "Changing the instructions or the placement of instructions within the PRO instrument" as an example of "potentially important changes in presentation or format" that can

Wyeth

alter the way patients respond to the same set of questions. However, this typically would not apply to a new instrument.

We recommend moving the text in line 383 (beginning with "Sponsors....") to 394 to be incorporated into "Other Changes" in Section IV. D. "Modification of an Existing Instrument."

X. Request for Revision of Scoring of Items and Domains

Section IV. B. 7, lines 416-422, "Identification of Preliminary Scoring of Items and Domains" provides recommendations on how to create a single score from multiple items. Specifically, the Guidance indicates in lines 416-419 that, "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." We believe that it should be up to the Sponsor to justify the method(s) used to combine items within a scale as well as methods to combine items across domains to form a composite.

We recommend revising the text on line 416 after the first sentence to the end of the paragraph on line 430 as follow: "It is the responsibility of the Sponsor to justify the method(s) chosen to combine items within a scale (e.g., equally weighted, patient preference weightings) or items across scales to form composite measures." In addition, we suggest that this section contain a reference to Section 2.2.3 of ICH E9 Statistical Principles for Clinical Trials for a discussion on Composite Variables - Considerations for Overall Clinical Development.

XI. Request for Revision of Table 3 to Include "Recall Period"

Section IV. B. 9, line 470 (Table 3), "Common Reasons for Changing PRO Instruments During Initial Development" seems to omit one of the most common reasons; recall period.

We recommend revising this table to include a new Item Property "Recall Period" and to identify the reason for change or deletion as "The population and disease state can affect the appropriateness of the recall period."

XII. Request for Clarification of Assessment of Measurement Properties

Section IV. C., lines 474-475, "Assessment of Measurement Properties" does not include sufficient flexibility for the type of PRO instrument. We believe that the level of scrutiny and quality of evidence required should depend on whether the PRO is: (1) widely accepted and already validated for this use, (2) a modified use of a widely accepted and already validated PRO, or (3) entirely novel. The

Wyeth

amount and quality of evidence required for validation may vary from merely citing the existing validation studies (e.g., in published, peer-reviewed literature) to providing new data from an adequate controlled study that uses the modified or novel PRO.

We recommend that line 475 be revised to add the following after "...e.g., minimum important difference": "The level of scrutiny and quality of evidence required should depend on whether the PRO is: (1) widely accepted and already validated for this use, (2) a modified use of a widely accepted and already validated PRO, or (3) entirely novel."

XIII. Request for Revision of Test-Retest Qualification

Section IV. C. 1, lines 491-492, "Evaluation of Reliability" states, "Test-retest reliability is the most important type of reliability for PRO instruments used in clinical trials." However, test-retest may not be appropriate for all clinical trials (e.g., may depend on disease state).

We recommend that lines 491-492 be revised as follows: "Test-retest reliability is an important type of reliability for PRO instruments used in clinical trials."

XIV. Input on MID in the Clinical Trial Setting

Section IV. C., line 483 (Table 4), "Measurement Properties Reviewed for PRO Instruments Used in Clinical Trials", states in the "Interpretability"/"FDA Review Considerations" that the FDA is specifically requesting comment on appropriate review of derivation and application of an MID in the clinical trial setting. Consistent with the guidance provided in lines 566-567, we believe that it is important to maintain a level of flexibility in evaluating approaches to establish an MID.

We recommend that the guidance for deviation of an MID incorporate flexibility and include multiple methods if there is consistency among methods confirming the MID.

XV. Input on Responder Definitions

Section IV. C., line 483 (Table 4), "Measurement Properties Reviewed for PRO Instruments Used in Clinical Trials", states in the "Interpretability"/"FDA Review Considerations" portion of the table that the FDA is specifically requesting comment on appropriate review of derivation and application of responder definition when used in clinical trials. We believe the context should determine the method of analysis or mode of presentation. Graphical depictions may show the percentage of subjects who responded at every possible response

Wyeth

level in a continuous, graphic fashion without requiring a potentially misleading definition of responder.

We recommend that the Interpretability/FDA Review Considerations portion of Table 4 be revised to include the following: "Graphical depictions showing a continuous plot of the percent change from baseline on the X-axis and the percent of subjects experiencing that change on the Y-axis may be preferable over categorical definitions of responders."

XVI. Request for Revision of "Modification of an Existing Instrument"

Section IV. D., "Modification of an Existing Instrument" states that when a PRO instrument is modified, additional validation may be needed, the extent of which depends upon the modification. FDA lists situations for which they recommend additional validation; these situations include applying the instrument to a population different from the validation population (by demographics, condition or severity levels), alterations to item content or format, or a changed mode of administration. Although there are situations when additional validation should be required, the examples provided seem to be too specific and more strict than what may actually be necessary to adequately demonstrate validation based on type of modification.

We recommend that, for clarity, lines 581-583 be revised as follows: "When a PRO instrument is modified it may be necessary to provide evidence to confirm the adequacy of the instrument's measurement properties. The extent of evidence will depend on the type of modification made." In addition, we recommend that lines 590-591 be revised as follows: "The FDA may consider a modified instrument as a different instrument from the original depending on the extent of the modification."

XVII. Request for Revision of Example in "Missing Items Within Domains"

Section VI. D. 1, lines 969-971, "Missing Items Within Domains" states that "For example, the SAP can specify that a domain will be treated as missing if more than 25 percent of the items are missing; if less than 25 percent of the items are missing, the domain score can be taken to be the average of the nonmissing items." While we agree that the SAP should specify the method of handling missing items, we have concerns about including this type of detailed example. The choice of the method(s) used to handle missing items should be dependent on the measure itself and we believe that providing an example with 25% missing is arbitrary and may mislead Sponsors.

Wyeth

We recommend that Section VI.D, contain a reference to ICH E9 Statistical Principles for Clinical Trial, Section 5.3 "Missing Values and Outliers" for a comprehensive discussion of how missing data should be handled.

We are submitting the above comments in duplicate. Wyeth appreciates the opportunity to comment on the above-mentioned Draft Guidance and trusts that the Agency will take these comments into consideration.

Sincerely,



Roy J. Baranello
Assistant Vice President
Regulatory Policy and Operations
Worldwide Regulatory Affairs