

April 4, 2006



Division of Dockets Management (HFA-305)  
Food and Drug Administration  
5630 Fishers Lane, Rm. 1061  
Rockville, MD 20852

**RE: Docket No. 2006D-0044 – Draft Guidance for Industry on Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims**

Merck & Co., Inc. is a leading worldwide human health products company. Through a combination of the best science and state-of-the-art medicine, Merck's Research and Development (R&D) pipeline has produced many important pharmaceutical products available today. These products have saved the lives of or improved the quality of life for millions of people globally.

Merck Research Laboratories (MRL), Merck's research division, is one of the leading biomedical research organizations. MRL tests many compounds as potential drug candidates through comprehensive, state-of-the-art R & D programs. Merck supports regulatory oversight of product development that is based on sound scientific principles and good medical judgment.

In the course of bringing Merck drug product candidates through developmental testing and clinical trials, Merck scientists address issues affected by this proposed Guidance. We have extensive experience in the clinical development of drug candidates and in the development and validation of Patient-Reported Outcomes (PROs) and have utilized that experience to author the comments below.

**General Comments**

We commend the Food and Drug Administration (the Agency or FDA) for their commitment to providing guidance to industry concerning PRO Measures in the context of product development and incorporation into approved product labeling. We recognize this is a nascent field for guidance development and appreciate the efforts the Agency has put forward in the generation of the document. In addition, the Agency's participation in workshops to introduce the PRO guidance and continued efforts to communicate this new information to sponsors is appreciated.

*Good Review Management*

Overall, the guidance provides for a very detailed evaluation of the PRO instrument throughout its life-cycle from development through to modification to ultimately, data analysis. The draft guidance recommends that detailed data on the development, adaptation, and validation be provided to the Agency in order for evaluation and acceptance of the instrument. A template designed to help format the submission of information to the FDA would be valuable and ensure that appropriate topics are covered (either by submission of data or an evaluation of why the parameter is not justified for the particular instrument).

The FDA recommends that sponsors discuss with the appropriate review division how best to plan for the interpretation of study findings. We welcome this guidance and are requesting that specific details concerning good review management of PRO evaluation (when to engage the Agency, the appropriate process to follow) be provided in the final guidance document. We suggest that PRO discussions occur earlier than the end-of-Phase II meeting and that ample time be allowed for discussion of PROs as part of the meeting process.

*Overall Development of PRO Instruments*

The draft guidance appears to be focused on the technical aspects of the life-cycle of a PRO rather than on identifying the potential biases that could prevent a sponsor from developing a fair and balanced claim. We strongly suggest that the PRO Guidance focus on any aspects of PRO modification that could result in a differential bias in the context of a randomized controlled clinical trial. This focus should be clearly articulated in the Guidance as the document is specifically intended to address the use of PROs in a clinical trial environment.

*Historically Accepted PRO Measures*

The PRO Guidance does not address the use of previously developed PRO measures which have been "validated" through use and have been recognized through their use as a standard for measurement in a disease area. The PRO Guidance should address this issue and indicate that there may be exceptions for the use of measures where there has been an extensive history of use in clinical trials and demonstrated response to established treatments. The Guidance can cite types of measures where this may be the case such as the 4-point migraine severity measure, functional disability scale and associated symptoms which are recommended by the International Society of Headache and the WOMAC measure for osteoarthritis trials based on the core set of measures recommended by the Outcome Measures in Rheumatology Clinical Trials III (OMERACT) group.

*Lack of Flexibility*

Often, as reflected in our specific comments that follow, the guidance becomes too prescriptive and the appropriate flexibility for developing, modifying, validating or interpreting the PRO instruments is lost. We have attempted to identify the sections that include the most limiting language. For example, the need for revalidation following

changes to an instrument should be based on the degree of the change, the nature of the instrument and the robustness of the original validation results. Not every change to a validated instrument necessitates re-validation; this point is not clearly stated in Section D: Modification of an Existing Instrument. Additionally, the Guidance does not make provision for the use of “tried and true” measures that may not have been developed or validated in the past based on best practices but which have a long history of use in clinical trials and have been incorporated into labels as a result of this experience. There must be provision for these measures to be used since replicating this level of “validation” through use in trials is invaluable to the interpretation of the results by researchers and clinicians.

#### *Health Care Resource Units*

In general, we believe that PROs should also include self-reported, unscheduled utilization of “health care resource units” not required or mandated by the protocol (such as hospitalization, emergency room or doctor office visits, and/or rescue medication use) as valid patient reported measurements. The draft guidance reflects only patient questionnaires or diaries; we request that “health care resource units” that can be accurately reported by patients be included as acceptable PRO measurements.

#### *International Harmonization*

While we appreciate the leadership role the FDA has taken with the development of this guidance document, we encourage the FDA to work with international regulatory agencies, possibly through the ICH process, to harmonize requirements for the development, validation and use of PRO instruments. It will be important to utilize similar approaches worldwide in order to streamline clinical development through participation in global clinical trials. The guidance touches on incorporating changes in language and culture as a first step toward facilitating development of global instruments.

### **Specific Comments**

We have tabulated our specific comments as follows: identification of the line in the draft guidance (in *italics*) followed by suggested edits (underlined) and our rationale supportive of the proposed changes. Our specific comments may be found in Attachment 1.

### **Recommendation**

Regulatory guidance on the development, use and approval of PRO-based labeling claims is valuable and provides the opportunity for assessing the patient’s perspective in the drug development process. The current draft guidance is often too prescriptive in its approach and, as a result, eliminates the needed flexibility to design and maintain PRO instruments that are appropriate for the myriad unique drug development programs. In addition, we strongly suggest that the guidance document focuses on those aspects of PRO modification that could result in a differential bias in the context of a randomized

controlled clinical trial. The guidance should recognize those PRO instruments that historically have been accepted as standards for measurement in certain disease areas.

We recommend that the Agency includes a section on good review management principles describing the opportunities (and mechanisms) for interaction between the sponsor and the Agency during the PRO development process. Finally, we encourage efforts toward international harmonization concerning the development, use and incorporation of PRO measures.

We appreciate the opportunity to share our comments with respect to the FDA Draft Guidance for Industry on Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims. Please do not hesitate to contact me, should you have any questions.

Sincerely,

A handwritten signature in black ink, appearing to read "Taryn Rogalski-Salter".

Taryn Rogalski-Salter, PhD  
Director  
Regulatory Policy

Attachment enclosed

## ATTACHMENT 1

### Specific Recommendations and Comments

Section	Paragraph Figure/ Table Line No.	Proposed Change	Comment/ Rationale
I.	22-24	<i>It also describes our current thinking on how sponsors can develop and use study results measured by PRO instruments to support claims in approved product labeling<sup>2</sup>. Footnote 2 Labeling, as used in this guidance, refers to <u>all information contained in the approved product labelling including the medical product description and summary of use, safety, and effectiveness that must be approved by the FDA.</u></i>	We are requesting a clarification to the term “ <i>approved product labeling</i> ” and associated footnote 2. It is unclear why the footnote refers only to parts of the label as historically FDA considered any information in any section of the label a claim, regardless of whether it is an indication. As such, the footnote should be expanded as indicated.
I	46-47	<i>...contained in a questionnaire or interview <del>schedule</del> along with ....</i>	The word <i>schedule</i> should be eliminated as it is not clear that it adds to the meaning of the sentence.
III. B.	164/Table 1 First Line	<i>Attribute: Intended use of the measure</i> <ul style="list-style-type: none"> <li>• <i>To evaluate <del>adverse events</del> tolerability</i></li> </ul>	A PRO measure should not be used to ‘ <i>evaluate adverse events</i> ’ but may be used to measure tolerability.
IV.	173-174	<i>... before <del>studies</del> Phase II/III clinical trials are initiated</i>	Suggested edit clarifies which studies are being referenced.
IV.	178-179	<i>“When considering an instrument that has been modified from the original, the FDA <del>generally plans to evaluate</del> may request an evaluation of the modified instrument <del>just as it would a new one</del> with patient cognitive interviews to assess the appropriateness of the change(s)”</i>	Reasonable modifications for clarity or modifications which do not change the intent of the questions or response options should not be considered as new measures. Considering a modified instrument as a ‘new one’ is not always appropriate and may not be warranted based on the types of modifications made.
IV. A. 1.	214-216	<i>“If the concept of interest is general (e.g. physical function), a single-item PRO instrument is usually unable to provide a complete understanding of the treatment’s effect because a <u>stand alone</u> single item cannot capture all the domains of the general concept”</i>	The suggested edit clarifies the reference to single-item PRO instruments being those PROs that are not supplemental to other measures, but stand on their own, especially for more general concepts. The Guidance should allow for appropriate ways in which a single, more specific supplemental item can capture important information, e.g. specific symptoms or a summation of related highly correlated symptoms capturing the same concept.
IV. A. 3.	275-279	<i>If substantive differences are thought to be present with respect to age, sex, ethnic identity, and cognitive ability, the FDA <del>plans to compare</del> may request a <u>comparison (e.g. through cognitive interviews)</u> of the patient population used in the PRO instrument development process to the study populations enrolled in clinical trials. <del>to determine whether the instrument is appropriate to that population</del></i>	Less restrictive language is needed concerning whether the instrument is appropriate to the population with respect to patient age, sex, ethnic identify, cognitive ability. As written, it appears that slight differences in demographics could cause FDA to question the instrument and its measurement characteristics for the new population. If there are substantive differences, consideration should be given to ways in which one could demonstrate appropriateness in the new population (e.g., cognitive interviews).

**[Docket No. 2006-0044] – Draft Guidance for Industry on Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims**

Section	Paragraph Figure/ Table Line No.	Proposed Change	Comment/ Rationale
IV. B. 3	334-337	<p><i>If a patient diary or some other form of unsupervised data entry is used, the FDA plans to review the protocol to determine what measures are taken to ensure that patients make entries according to the study design and not, for example, just before a clinic visit when their reports will be collected. <u>FDA recognizes that even with proper controls, last-minute entries may occur and the percentage of this occurrence should not differ between randomized treatment groups. In addition, FDA recognizes that not all studies can be conducted using electronic diaries to assure automatic date and time stamping of entries.</u></i></p>	<p>Despite standardization and training on diary completion, FDA should recognize that with paper diary it may not be possible to ensure that a certain proportion of diary entries do not occur just before a clinic visit. The percent of this occurrence should not differ by randomized treatment group. The FDA should recognize that not all studies can be conducted using electronic diary methods to assure date and time stamping of entries which would facilitate tracking.</p>
IV. B. 3	339-343	<p><del><i>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.</i></del></p> <p><u>PRO instrument recall periods should be appropriate to the disease and events that are being recalled. A longer recall period may be appropriate for infrequent events. The ideal recall period for each PRO measure should be considered as a part of the conceptual model, including considerations such as the disease, population, duration of the study, frequency of events, ability to remember event (e.g. hospitalization), method of collection (electronic or paper), etc.</u></p>	<p>As written, this section is too prescriptive. The guidance should recognize that each PRO measure and its use in a population may differ which can impact the recall period. As examples, it might be most appropriate to capture the number of migraine headaches in a migraine population over the past 1 week period or 1 month period or 3 month period depending on the population studied. It may be most appropriate to capture acute pain symptoms hourly or every 30 minutes but more appropriate to capture chronic pain symptoms daily or perhaps weekly if the pain is not episodic or variable.</p>

**[Docket No. 2006-0044] – Draft Guidance for Industry on Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims**

Section	Paragraph Figure/ Table Line No.	Proposed Change	Comment/ Rationale
IV. B. 4.	Table 2	<p><i>Types of Response Options. Anchored or categorized VAS. A VAS that has the addition of one or more intermediate marks positioned along the line with reference terms assigned to each mark to help patients identify the locations (e.g., half-way) between the ends of the scale. VAS scales have not been shown to be any more precise than other types of response option scales.</i></p>	<p>Edit suggested for clarity and to inform the reader that research demonstrates no differences in precision</p>
IV. B. 4.	363	<ul style="list-style-type: none"> <li>▪ <i>Response options are appropriately ordered and appear to represent equal intervals.</i></li> </ul>	<p>Please clarify what is meant by “appear to represent equal intervals”.</p>
IV. B. 6.	380-394	<p><i>Development of Format, Instructions, and Training. PRO study results can vary according to the instructions to patients or the training given to the interviewer or persons supervising PRO data collection. Sponsors should consider all PRO instrument instructions and procedures contained in publications and user manuals provided by developers, including procedures for reviewing completed questionnaires and re-administration to avoid missing data or clarify responses. Other important considerations include the format of the questionnaire, the final wording of PRO instruments as implemented in clinical trials, and any potentially important changes in presentation or format. Examples of changes that can alter the way that patients respond to the same set of questions include:</i></p> <ul style="list-style-type: none"> <li>▪ <del><i>Changing an instrument from paper to electronic format</i></del></li> <li>▪ <del><i>Changing the timing of or procedures for PRO instrument administration within the clinic visit</i></del></li> <li>▪ <del><i>Changing the order of items or deleting portions of a questionnaire</i></del></li> <li>▪ <del><i>Changing the instructions or the placement of instructions within the PRO instrument</i></del></li> </ul>	<p>Clarification is needed within this section. We agree that changes listed in bulleted form could potentially alter patient responses and consideration needs to be given to the impact of the changes on how patients respond to questions; justification for the change may be necessary. In addition, this section is too prescriptive. The list of changes that may alter the way patients respond is very broad and lumps more drastic changes (e.g. <i>deleting portions of a questionnaire</i>) with less drastic and probably inconsequential changes (e.g. <i>changing the placement of instructions</i>). Therefore, we suggest deleting the list of examples.</p>

**[Docket No. 2006-0044] – Draft Guidance for Industry on Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims**

Section	Paragraph Figure/ Table Line No.	Proposed Change	Comment/ Rationale
IV. B. 6.	396-407	<p><i>It is important that the PRO instrument format used in the clinical trial be <del>consistent with</del> <u>similar to</u> the format that is used in the instrument validation process. Format refers to the <del>exact</del> appearance of the instrument. Instrument format is specific to the mode of administration, including paper and pencil, interviewer-administered or supervised, or electronic data collection. <u>Minor modifications to the format of a PRO instrument may be necessary and will not compromise the integrity or validity of the measure. If more substantive modifications are made there should be evidence that these were tested with patients in cognitive interviews and found to be acceptable.</u> The FDA plans to review the PRO instrument in the format used in the clinical trial case report forms, including the order and numbering of items, the presentation of response options in single response or grid formats, the grouping of items, patterns for skipping questions that are not applicable, and all instructions to patients in the interview schedule or on the questionnaire.</i></p> <p><i>The FDA recommends that the PRO instrument development process includes the generation of a user manual that specifies how to incorporate the instrument into a clinical trial in a way that minimizes administrator burden, patient burden, missing data, and poor data quality.</i></p>	<p>Requiring that the format of a PRO instrument used in the clinical trial be consistent with the format used in the instrument validation process is too prescriptive. Minor modifications to the format may be appropriate and necessary for use in a clinical trial (e.g. changing from a line on which to place an “x” versus a box or having the responses indented) and will not impact the validity of the measure or invalidate the results as used in a clinical trial.</p>

[Docket No. 2006-0044] – Draft Guidance for Industry on Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims

Section	Paragraph Figure/ Table Line No.	Proposed Change	Comment/ Rationale
IV. B. 7.	416-422	<p><del>A scoring algorithm creates a single score from multiple items. 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 response options varies by item. The same weighting concerns apply with added complexity when combining domain scores into a single overall score.</del></p> <p><u>A scoring algorithm creates a single score from multiple items. Highly correlated items in a scale (e.g. &gt; 0.95) may not add additional information and may result in a given concept being given greater weight than other items within a scale score. This may be appropriate if it is important to give greater weight to a specific concept. However, in most cases to avoid respondent burden it is best to reduce the number of redundant items within a scale. In some cases, one question or item may adequately capture a specific concept which can be established with statistical testing. The same concerns apply when combining domain scores into a single overall score. Each domain should provide a separate concept that is important to the overall objective of the measure. Unequal weighting of items or domains should be justified based on importance to patients.</u></p>	<p>We recommend that this paragraph be deleted or extensively modified in order not to cause misinterpretation. The FDA should recommend that researchers should strive for equal weighting of questions or domains within a measure. Unequal weighting of questions or domains may require justification of weights in each separate population which may be impractical and limit the ability to utilize a measure more broadly. We agree with the FDA that researchers should consider whether some items in a scale are highly correlated, and if they are, that they may be weighting certain items (especially if highly correlated or redundant) more than others within a scale or domain. This may or may not be appropriate, depending on the importance of the item to patients.</p>
IV. C	478-480	<p><del>The sociodemographic and medical characteristics of any sample used to develop or validate a PRO instrument <u>determine</u> may affect its appropriateness for future clinical study settings.</del></p>	<p><del>Determine</del> is too prescriptive for this statement.</p>

[Docket No. 2006-0044] – Draft Guidance for Industry on Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims

Section	Paragraph Figure/ Table Line No.	Proposed Change	Comment/ Rationale
IV. C	481/Table 4	<i>Reliability, What is assessed: Whether the items in a domain are intercorrelated, as evidenced by an internal consistency statistic (e.g., coefficient alpha)</i>	This statement in Table 4 contradicts line 416 – 417: <i>Equally weighted scores for each item are appropriate only when the responses to the items are relatively uncorrelated.</i> Please clarify within the guidance.
IV. C	481/Table 4	<i>Validity, Test: Content-related <u>or face validity</u></i>	For clarity, addition of the term “face validity” is suggested as it is commonly being used to refer to <i>content-related</i> .
IV. C	481/Table 4	<i>Validity, FDA Review Considerations: <del>Have patients similar to those participating in the clinical trial confirmed the completeness and relevance of all items?</del></i>  <u>Inclusion of subjects/patients with the disease being studied and of a similar age range to those who will participate in the clinical trial to confirm completeness and relevance of items.</u>	It is not feasible to request that <i>patients similar to those participating in the clinical trial</i> confirm completeness and relevance of all items in the PRO. Instead we suggest that the FDA reviewer considerations be modified as indicated. The intent of the validity measurement or construct-related validity can be assessed by subjects/patients with the disease being studied and of a similar age range.
IV. C	481-2 Table 4	<i>Validity, Test: <del>Ability to predict future outcomes (also known as predictive validity)</del></i>	We recommend deletion of this as a test of the validity of a PRO.
IV. C	483/Table 4	<i>Interpretability, Test: Smallest difference that is considered clinically important; this can be a specified difference (the minimum important difference (MID)) or, in some cases, any detectable difference. The MID is used as a benchmark to interpret mean score differences between treatment arms in a clinical trial.</i>	We recommend that the guidance be edited to allow for flexibility in the method(s) used for establishing a MID as no single approach is universally accepted or appropriate (in addition, see lines 566-567 where it is recommended to use <i>a variety of methods</i> which is inconsistent with this section suggesting that a single approach or a combination of approaches may be acceptable).  This section should also recognize the difference between Minimal Important Change (MIC) within groups and Minimal Important Difference (MID) between groups. The methodology for determining MIC is more developed than MID. Furthermore, the application of MIC to MID may not be valid based on the study design used to establish MIC.
IV. C. 1	495-496	<i>Internal consistency reliability, in the absence of test-retest reliability, <del>does</del> <u>may</u> not generally constitute sufficient evidence of reliability for clinical trial purposes.</i>	The word “ <i>does</i> ” should be replaced as indicated.
IV. C. 1.	496-497	<i>When PRO instruments are interviewer-administered, <del>inter-interviewer reproducibility is critical</del> <u>it is important to standardize administration of the measure and train interviewers on this standard in order to increase inter-interviewer reproducibility.</u></i>	Edits suggested for clarity.

**[Docket No. 2006-0044] – Draft Guidance for Industry on Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims**

Section	Paragraph Figure/ Table Line No.	Proposed Change	Comment/ Rationale
IV. C. 2	510-512	<i>If developers expected the instrument to discriminate between patient groups (e.g., between patients with different levels of severity), the FDA is interested in evidence that shows the instrument <del>meaningfully</del> discriminates.</i>	It is unclear what is meant by the term “ <i>meaningfully</i> ”.
IV. C. 3.	527-529	<p><del><i>The extent to which the PRO instrument’s ability to detect change varies by important patient subgroups (e.g., sex, race, age, or ethnicity) can affect clinical trial results.</i></del></p> <p><u>The extent to which the PRO instrument’s ability to detect change can vary by important patient subgroups. The statistical analysis plan for a clinical trial should test the differences between important subgroups (sex, race, age).</u></p>	Difference in response by subgroup is generally examined as a part of the statistical analysis plan in a clinical trial. It may not be possible to <i>a priori</i> determine if there are important differences in response by subgroup. Additionally, the differences in response may not be due to the measure but rather due to the subgroups response to treatment. Separate validation studies in each subgroup may not be possible, but differences can be examined within the trial and acknowledged.
IV. C. 4.a	545-547	<del><i>If PRO instruments are to be considered more sensitive than past measures, It can be useful to specify a minimum important difference (MID) as a benchmark for interpreting mean differences.</i></del>	This statement should be revised as we are unclear why PRO instruments are automatically considered more sensitive than past measures such as change in blood pressure or lung function.
IV. D.	583-585	<p><del><i>For example, small nonrandomized studies may be adequate to assess the results of changing a response scale from vertical to horizontal.</i></del></p> <p><u>For example, cognitive debriefing may be adequate to assess the ability of patients to complete a modified response scale from vertical to horizontal.</u></p>	Edits suggested for clarity.
IV. D	591-593	<i>The FDA recommends additional validation to support the development of a modified PRO instrument when one or more of the following modifications occur.</i>	For clarity the FDA should specify what ‘ <i>additional validation</i> ’ includes.
IV. D.	581-670	<i>D. Modifications of an Existing Instrument</i> (entire section)	This section combines modest changes and more drastic examples of changes to a PRO instrument. The Guidance should clarify the extent of changes illustrated (major changes, minor changes) and not make all changes subject to the same validation requirements. For example, many of the modifications should be assessed in cognitive interviews or focus groups. However, the examples in Section I, Revised Measurement Concept, may require re-validation.

**[Docket No. 2006-0044] – Draft Guidance for Industry on Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims**

Section	Paragraph Figure/ Table Line No.	Proposed Change	Comment/ Rationale
IV. D. 3	621	<ul style="list-style-type: none"> <li>Wording or placement of instructions</li> </ul>	Change in 'Placement of instructions' should not require re-validation
IV. D. 3	622	<ul style="list-style-type: none"> <li>Wording or order of the items</li> </ul>	Change in 'Order of items' should not require re-validation
IV. D. 3	624	<ul style="list-style-type: none"> <li>Recall period associated with an item</li> </ul>	Change in 'Recall period' should not require re-validation
IV. D. 3	627	<ul style="list-style-type: none"> <li>Scoring (including creation of summary scores, subdomain scores, or cut-points)</li> </ul>	Change in 'Scoring' should not require re-validation
IV. D. 4	639-641	<ul style="list-style-type: none"> <li>Instructions or procedures for administration within a trial differ from those used in validation studies (can alter the meaning of the responses from that of the original version)</li> </ul>	Change in 'Instructions' should not require re-validation
IV. D. 5	659	<ul style="list-style-type: none"> <li>The harmonization of different versions <u>which can be accomplished either through a formal comparison or by a determination of conceptual equivalence of the versions.</u></li> </ul>	It should be recognized that an acceptable definition of the term <i>harmonization</i> , as used in this section, can encompass both formal and informal aspects. Harmonization can be a formal meeting to discuss the translations and harmonize versions or can be a less formal, but still acceptable practice, of harmonizing by cross-comparison of the translations for conceptual equivalence. Additionally, having all translations of a single PRO performed by a single vendor enhances the ability to maintain harmonization.
IV. D. 5.	660	<ul style="list-style-type: none"> <li><del>The evidence that measurement properties for translated versions are comparable.</del></li> <li>Cognitive testing of translations establishes that the translated PRO is conceptually equivalent with the original PRO. Within clinical trials tests of interaction by country can be performed to test consistency of the translated PROs across countries.</li> </ul>	Determining the measurement properties of translated PROs for each country may not be feasible due to sample size. Tests of interaction within the clinical trial should be sufficient to examine differences by country which may be due to the translated PRO. In a randomized clinical trial, the sponsor has the burden if a measure is not psychometrically equivalent in that the increased variability in response on the PRO may decrease the ability to show a difference in response to treatment.
IV. D. 6	666	<ul style="list-style-type: none"> <li><del>The PRO instrument was not developed and validated for use in a clinical trial.</del></li> </ul>	PRO are almost always developed and tested initially in observational studies; therefore, this sentence should be deleted. There are examples of physiologic tests such as spirometry, blood pressure, and cholesterol which were initially developed for use in surveys to identify high risk individuals. However, these measures have been used very successfully in clinical trials without modification other than standardization of measurement which is something that always should be done with PROs or any measure to assure that the measure is being administered in a standard fashion across patients and sites.
IV. D. 6	667-668	<ul style="list-style-type: none"> <li><del>A PRO instrument developed and previously used as a stand alone assessment is included as a part of a battery of measures.</del></li> </ul>	It is not unusual to include a PRO as part of a larger battery of measures within a clinical trial. A validated PRO is still validated whether or not it is used alone or as part of a battery or measures. Therefore, this sentence should be deleted.

**[Docket No. 2006-0044] – Draft Guidance for Industry on Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims**

Section	Paragraph Figure/ Table Line No.	Proposed Change	Comment/ Rationale
IV. D. 6	664-670	<p><del>Other Changes: Other changes to the PRO instrument or the way in which it is assessed that may necessitate additional validation include:</del></p> <ul style="list-style-type: none"> <li><del>▪ The PRO instrument was not developed and validated for use in a clinical trial</del></li> <li><del>▪ A PRO instrument developed and previously used as a stand-alone assessment is included as a part of a battery of measures</del></li> <li><del>▪ A PRO developed to measure a treatment benefit is subsequently used to measure a decrement as interpreted by a score change in the opposite direction</del></li> </ul>	Entire section should be deleted (as per specific comments above)
V. A.	709-710	<p><del>If the goal of PRO measurement is to support claims, we recommend that measurement of the PRO concept be clearly stated as a specific study objective or hypothesis.</del></p>	Wording changed for appropriateness.
V. E.	802-803	<p><del>The FDA recommends that sponsors discuss with the appropriate review division how best to plan for the interpretation of study findings.</del></p>	The FDA should clarify the process of engaging the review divisions, specifically, at what stage of development, how engagement is facilitated, who should be included in the discussions, etc.
V. F.	830-835	<p><del>The use of electronic PRO instruments, however, may pose a problem if direct control over source data is maintained by the sponsor or the contract research organization and not by the clinical investigator. The use of electronic PRO instruments, however, may pose a problem if the electronic source records are not maintained. Technology must ensure that only the investigator can maintain electronic source records. The FDA considers the investigator to have met his or her responsibility when the investigator retains the ability to control maintain and provide access to the records that serve as the electronic source documentation for the purpose of an FDA inspection.</del></p>	<p>We suggest elimination of the requirement for investigator to have 'direct control' over the electronic source record and replace with a requirement that the "technology must ensure that only the investigator can maintain the electronic source records." 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.</p> <p>A third 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.</p>

**[Docket No. 2006-0044] – Draft Guidance for Industry on Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims**

Section	Paragraph Figure/ Table Line No.	Proposed Change	Comment/ Rationale
V. F.	840-843	<i>Because electronic PRO data (including data gathered by personal digital assistants or phone-based interactive voice recording systems) are part of the case history, the FDA expects electronic PRO data to be consistent with the data standards described in that guidance.</i>	Clarify what ‘data standards’ are being referenced. After referencing 21 CFR Part 11, this sentence indicates that electronic PRO data should be ‘consistent with the data standards described in that guidance’. 21 CFR Part 11 contains requirements for electronic records/signatures but does not reference any ‘data standards’.
V. F.	848-849	<p><i>Sponsors should also plan to avoid the following:</i></p> <ul style="list-style-type: none"> <li>▪ <i>Direct PRO data transmission from the PRO data collection device to the sponsor <u>without verifiable controls to ensure that the investigator can appropriately maintain source records (i.e., the sponsor should not have exclusive control of the source document)</u></i></li> </ul>	We are requesting clarification on the acceptability of direct data transmission from device to technology vendor under contract from the sponsor. This is the current model for most studies involving electronic PRO. Direct PRO data transmission should be acceptable if verifiable controls are in place to ensure that the investigator can appropriately maintain source records.
V. F.	852	<p><i>Sponsors should also plan to avoid the following:</i></p> <ul style="list-style-type: none"> <li>▪ <i>Removal of investigator accountability for confirming the <del>accuracy</del> <u>completeness</u> of the data</i></li> </ul>	<p>Since PROs are supplied directly by the patient, it is not possible for an investigator to confirm the “accuracy” of such data. Accuracy is patient-dependent; only the patient him/herself can attest to the accuracy of the data they have reported.</p> <p>To the extent possible, and depending on the PRO administrative procedures, the investigator (or delegate) could confirm the completeness (e.g. all questions/ checkboxes completed) and authenticity (e.g. attest it was indeed the patient who supplied the info) of the data, but cannot confirm the accuracy of the patient reported outcomes.</p>
VI. E	1028-1031	<i>When clinical trials show small <u>but significant mean</u> effect sizes, rather than considering results in terms of an MID, it may be more informative to examine the distribution of responses between treatment groups to more fully characterize the treatment effect and examine the possibility that the mean improvement reflects very different responses in subsets of patients.</i>	Small effect sizes should be significant to be meaningful.