



Amgen  
One Amgen Center Drive  
Thousand Oaks, CA 91320-1799  
805.447.1000

**Date:** ► April 4, 2006

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

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*Draft Guidance for Industry on Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims*

Dear Sir/Madam:

Amgen is a global biotechnology and pharmaceuticals products company based in Thousand Oaks, CA, which strives to serve patients by transforming the promise of science and biotechnology into therapies that have the power to dramatically improve people's lives.

We are pleased to provide the following comments on the draft guidance, *Draft Guidance for Industry on Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims*. Below we have identified two general concerns that are of primary importance. Our detailed comments are provided in an attachment to this letter.

#### **Communication with FDA on PRO measures**

It would be useful to have greater clarity on when to meet with FDA for consultation regarding potential PRO claims. Given the information provided in the draft guidance, it seems clear that, if there is uncertainty regarding the FDA's acceptance of a PRO instrument, then having an initial discussion at the End of Phase 2 Meeting is too late. Also, it would be helpful if the guidance clarified how FDA's Study Endpoint and Label Development Office will be involved in such meetings and the participation, as well as role, of the various reviewing offices.

#### **Status of existing instruments**

The draft guidance provides great detail on the development of new PRO instruments; however, for the most part, sponsors are not developing new instruments. Most sponsors are using existing instruments that have been developed by others, eg, academic researchers.

Therefore, the document could be made more useful to sponsors if guidance were provided on how the Agency will evaluate the use of existing PRO measures in trials for labeling claims. What types of documentation will be considered sufficient to demonstrate to the FDA that an existing, widely used tool has been adequately validated? Will approved language in labeling

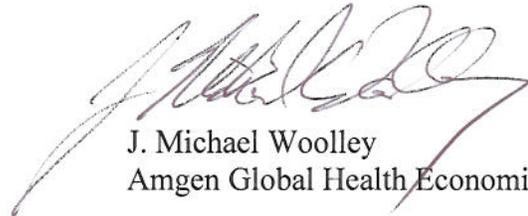
imply that a PRO instrument used in support of such language is valid and reliable for that indication and that population? Alternatively, might a given PRO instrument be only considered acceptable if it was reviewed and approved for labeling within a specific timeframe (e.g., the last five years)? If none of these are the case, then it is suggested that the FDA maintain a public document that describes which PRO measures are acceptable for which indications and populations. This last suggestion is designed to improve transparency since finding this out in private consultation with the FDA may induce a lag and/or unnecessary development costs.

If you have any questions regarding our comments, or how we may assist with further development of this guidance, please contact Jenny Peters at (805)-447-8840.

Sincerely,



Jenny Peters  
Amgen Global Regulatory Affairs & Safety



J. Michael Woolley  
Amgen Global Health Economics

**ATTACHMENT**

<b>(Line numbers) “Original Text”</b>	<b>Comment</b>	<b>Suggested New Text</b>
<p>(31 - 32) “A PRO is a 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).”</p>	<p>There may be times when patients are unable to provide responses (e.g., too ill, too young, etc.). Under these circumstances it is possible that a proxy report, provided by a close family member for example, may be acceptable.</p>	<p>“A PRO is a measurement of any aspect of a patient’s health status that <b>typically</b> comes directly from the patient (i.e., without the interpretation of the patient’s responses by a physician or anyone else).”</p>
<p>(153 - 156) “Some PRO instruments (e.g., health-related <i>quality of life</i> instruments) attempt to measure both the effectiveness and the side effects of treatment. PRO instruments that are used in clinical trials to support effectiveness claims should measure the adverse consequences of treatment separately from the effectiveness of treatment.”</p>	<p>This statement is true if the positive effects of treatments manifest in areas mutually exclusively from those that reflect adverse consequences. However, in the case where the positive and negative effects overlap, PRO instruments are designed to elicit patients’ evaluations of the concept of interest, taking into account the combined effect. Thus it may not be desirable, nor feasible, to separate the different effects.</p>	<p>Delete these two sentences.</p>
<p>(166) “At regular intervals throughout a study”</p>	<p>This sentence seems to indicate that PROs should be assessed at regular intervals in studies. In some situations it may be appropriate to measure PROs at irregular intervals during the study.</p>	<p>“At <b>appropriate</b> intervals throughout a study”</p>

**ATTACHMENT**

<b>(Line numbers) “Original Text”</b>	<b>Comment</b>	<b>Suggested New Text</b>
<p>(231 - 234) “For example, 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.” See also (951 - 954) “In general, 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).”</p>	<p>Please confirm in the guidance that a narrower claim would be allowed.</p>	
<p>(298 - 300) “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.”</p>	<p>It remains unclear what is an adequate number of patients from the FDA’s perspective. Please identify the factors that will be used to define this number.</p>	

**ATTACHMENT**

(Line numbers) “Original Text”	Comment	Suggested New Text
<p>(307 - 308) “Of course, it would be critical to know that each item refers to something that patients actually do.”</p>	<p>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).</p>	<p>“It <b>may</b> be <b>useful</b> to know whether each item refers to something that patients actually do.”</p>
<p>(339 - 343) “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.”</p>	<p>This is an extremely complex topic with many unresolved issues. Furthermore, there is insufficient scientific evidence to support this statement. 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 should be considered acceptable. For example, it may be the best way to determine the effect of an acute event on patient outcomes after the event has passed. Responses assessing outcomes over an interval may strike the appropriate balance between patient burden and accuracy. While the text is somewhat softened by the words “may” and “usually”, it remains too prescriptive.</p>	<p>Delete paragraph (339 – 343).</p>

ATTACHMENT

(Line numbers) "Original Text"	Comment	Suggested New Text
<p>(416 - 420) "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."</p>	<p>Item weights are not necessarily related to item correlations. An important concept, perhaps covered by more than one item (hence likely to be highly correlated) may still be more highly weighted (or at least equally weighted) relative to items that assess less critical outcomes within a domain. For example, in the SF-36 PF scale the item "walking more than a mile" is correlated with "walking several blocks" (which is also correlated with "walking one block"). These items are less correlated with "lifting or carrying groceries". However, equal weighting is likely reasonable since mobility is a very important component of physical functioning.</p>	<p>Delete sentences.</p>
<p>(481) Table 4 (for validity) "Ability to predict future outcomes (also known as predictive validity)"</p>	<p>The ability to predict future outcomes provides additional support of validity for a PRO instrument or domain (or scale). However this is not a necessary property of PRO instruments and its absence should not reduce or put into question the validity of a PRO instrument.</p>	<p>"Ability to predict future outcomes (also known as predictive validity). <b>This property is not necessary to demonstrate the validity of a PRO instrument.</b>"</p>

**ATTACHMENT**

(Line numbers) “Original Text”	Comment	Suggested New Text
<p>(483) Table 4 (for the ability to detect change) “Ability to detect change is always specific to a time interval.” and “Has ability to detect change been assessed for the time interval appropriate to study?”</p>	<p>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.</p>	<p>Delete both sentences.</p>
<p>(566 - 567) “If an MID is to be applied to clinical study results, it is generally helpful to use a variety of methods to discover whether concordance among methods confirms the choice of an MID.”</p>	<p>Sometimes a specific method for estimating the MID may be appropriate. In such cases, this appropriate method may lead to the best estimate of the MID, even if it is not consistent with results using other methods.</p>	<p>“If an MID is to be applied to clinical study results, it <b>may be</b> helpful to use a variety of methods to discover whether concordance among methods confirms the choice of an MID.”</p>
<p>(585 - 589) “On the other hand, if the PRO instrument is to be used in an entirely new population of patients, a small randomized study to ascertain the measurement properties in the new population may minimize the risk that the instrument will not perform adequately in a phase 3 study.”</p>	<p>What is meant by a randomized study here?</p> <p>Conducting a study may not be all that can be done to minimize risk, so suggest changing “minimize” to “reduce”</p>	<p>“On the other hand, if the PRO instrument is to be used in an entirely new population of patients, a small study to ascertain the measurement properties in the new population may <b>reduce</b> the risk that the instrument will not perform adequately in a phase 3 study.”</p>

**ATTACHMENT**

<b>(Line numbers) “Original Text”</b>	<b>Comment</b>	<b>Suggested New Text</b>
(590 - 593) “The FDA intends to consider a modified instrument as a different instrument from the original and will consider measurement properties to be version-specific. The FDA recommends additional validation to support the development of a modified PRO instrument when one or more of the following modifications occur.”	The language here is too stringent. Very small changes to a PRO instrument (e.g., cosmetic changes) may not necessitate these analyses. For example, a small change in wording of instructions or moving it to the top of a page may not warrant the revalidation of the PRO instrument.	<b>“Depending upon the size and scope of the modification, the FDA may consider a modified instrument as a different instrument from the original and may consider measurement properties to be version-specific. The FDA may recommend additional validation to support the development of a modified PRO instrument when one or more of the following modifications occur.”</b>
(666) “The PRO instrument was not developed and validated for use in a clinical trial.”	There are PRO instruments that were not developed for use in clinical trials that still perform well in them. This should not invalidate their use.	“The PRO instrument was not validated for use in a clinical trial.”
(681 - 683) “It is important that PRO instruments developed for adults are not used in pediatric populations unless the measurement properties are similar in all age groups tested.”	It should not be necessary to show that the measurement properties are similar, rather that the measurement properties support its validity and reliability. Furthermore, this need not be the case for all age groups, rather only for those age groups completing the PRO instrument in the clinical trial.	Delete sentence.
(717 - 718) “Because responses to PRO measures are subjective, representing a patient’s impression, open-label studies, where patients and investigators are aware of assigned therapy, are rarely credible.”	There are enough cases where open label studies do provide some useful data that this language should be modified.	<b>“Because responses to PRO measures are subjective, representing a patient’s impression, open-label studies, where patients and investigators are aware of assigned therapy, are open to question.”</b>

**ATTACHMENT**

<b>(Line numbers) “Original Text”</b>	<b>Comment</b>	<b>Suggested New Text</b>
<p>(939 - 942) “In any such composite, it is critical to ensure that patients enrolled in a clinical study are impaired in all domains (e.g., psychological or emotional well-being) because they cannot improve in domains if they are not impaired in whatever concept the domain measures.”</p>	<p>It may be infeasible to restrict enrollment in a trial to patients who are impaired in all domains of the PRO instrument. However, to the extent that they are less impaired, this reduces the potential to demonstrate efficacy and increases the type II error. Furthermore, if improvement in the composite score is driven largely by only one item or domain (or scale), this point is covered by other language in the guidance (e.g. lines 231 and 951).</p>	<p>Delete sentence.</p>