

VIA ELECTRONIC DELIVERY

March 31, 2006

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

**Re: Docket Number 2006D-0044**

**Request for Comments on: Draft Guidance for Industry on Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims**

Dear Sir or Madam:

Eli Lilly and Company (Lilly) respectfully submits the following written comments regarding the February 3, 2006 Federal Register notice announcing availability of a draft guidance for industry entitled “Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims.”

Lilly is a leading, innovation-driven corporation committed to developing a growing portfolio of best-in-class and first-in-class pharmaceutical products that help people live longer, healthier, and more active lives. We are committed to providing *Answers that Matter* – through medicines and information – for some of the world’s most urgent medical needs.

Lilly congratulates the FDA on developing this draft guidance through a process that has included close collaboration with leading experts in the area of patient reported outcomes. This draft guidance, when finalized, can provide much-needed information and general principles on evidence that can be used to support product labeling statements that are based on patient reported outcomes.

Following are suggestions that Lilly believes will enhance the clarity of the Draft Guidance. We first present some general comments for consideration. The remaining suggestions are organized by topic, and reference line numbers in the Draft Guidance for ease of review.

## **Draft Guidance for Industry on Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims**

### **General Comments**

- The Draft Guidance is extremely well-written and thorough. The diagrams, tables, and examples help to explain some complex and multi-dimensional issues.
- There needs to be more clarity in the Draft Guidance regarding where in labeling PRO-based claims would be located. The Draft Guidance is not clear on whether “label claims” can appear in any section of labeling (e.g., clinical trials), or are restricted to certain sections of labeling (e.g., indications).
- The guidance document is quite prescriptive, especially given the subject matter. Some of the strict declarations may not be feasible or applicable to all PROs. There needs to be some flexibility for things such as precedent (e.g., validation by years of use). Some examples are included in the line-referenced comments.
- There should be more explanation in the guidance on how PRO results can be used in promotion practices. Will requirements for including a PRO claim in promotion be the same as those listed in the guidance for including a PRO claim in labeling?
- There needs to be greater allowance for PROs that affect Quality of Life or other “holistic” concepts. As written, the Draft Guidance concentrates the PRO concept around how a drug is affecting individual domains and appears to overlook the more general concepts.
- Based on public presentations made by FDA on the Target Product Profile (TPP), it is clear that the TPP would be an ideal tool to facilitate the discussions between the

Review Division, SEALD, and DDMAC. Please coordinate the Draft Guidance on PROs with any Guidance on TPP and vice versa.

### **Specific Comments**

#### Adverse Events and PROs (Lines 154-156):

The Draft Guidance states, "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."

Safety, including "adverse consequences of treatment," can be (and usually is) assessed within a clinical trial without explicitly being part of the PRO instrument. Not all PRO instruments should be expected to include an assessment of adverse effects. Suggest that guidance not require that PRO instruments explicitly include measures of adverse consequences if such consequences are captured elsewhere within the clinical trial.

#### Interaction with SEALD (Lines 173 and 190):

Although the draft guidance provides framework for scientific development of PROs, and clarifies semantics, the actual process for FDA-industry interactions to achieve labeling is left unclear. **In line 173**, FDA asks sponsors to start 'early in development' to understand goals. We suggest committing an additional paragraph at this point to provide the following clarifications, for example:

1. Specifically state that SEALD may only be approached through the therapeutic divisions.
2. **Line 190**; specify how the wheel and spokes diagram corresponds to expected FDA interactions for Sponsors developing new PRO instruments: for instance, "i. Conceptual framework" to be discussed at IND submission or EOP1 meeting. "ii. Create Instrument to be discussed at EOP1" and "iii. Assess measurement properties and their impact on Phase 3 protocols to be discussed preliminarily at EOP2 meeting".

Claims from Multidomain Instruments (Lines 227 and 912):

**In line 227**, the discussion of multidomain instruments begins. Although there is substantial text, there is little actual recommendation as to how labeling decisions will be made. We suggest consideration of harmonization with the EMEA's brief but clear "CHMP reflections paper on the regulatory guidance....". The CHMP reflection paper states that a HRQOL claim could be made when it "...demonstrates robust improvements in all or most of these domains". **In line 912** of the FDA Draft Guidance, the language suggests that for a HRQOL claim all domains that are 'important' must be affected by treatment, but significant separation on most domains and the index score would satisfy EMEA standards. In addition, we suggest that more detail be made available to help understand how much information would appear in the clinical trials (or other) section of labeling; e.g., all domains vs those identified *a priori*.

Associations Among Conceptual Framework Elements (Lines 249 to 260):

The Draft Guidance states, "For measures of general concepts, the FDA intends to review how individual items are associated with each other, how items are associated with each domain, and how domains are associated with each other and with the general concept of interest..."

Please clarify what kind of specific evidence the agency wants to consider that demonstrates association between the elements of the conceptual framework. The simplicity of **Figure 2** beginning on line **258** would suggest that a traditional analytic approach of item to total correlations, item to remainder correlations, exploratory factor analysis, confirmatory factor analysis, and the like would be appropriate for determining these relationships.

However, more complex conceptual models can lend themselves to more complex analytic approaches, such as path analysis, that aim to understand the direct and intervening effects of a variety of independent variables to patient perception of the concept of interest. Such effects can be derived from the domains that compose the concept but also from other independent influences such as demographics and disease state characteristics. These latter analyses are not trivial undertakings and may not be immediately thought of as something in which the Agency has a specific interest. Clarification of the precise information that will be of interest, similar to the specificity that was given in Table 4 dealing with measurement properties, will greatly

help in the preparation of both the types of variables that will need to be collected and in the analytic plan in the early phases of instrument development.

Types of Response Options (Line 351, Table 2)

Under “Visual Analog Scale” the Draft Guidance comments, “These scales often produce a false sense of precision.” In the absence of scientific evidence for this statement, this sentence should be removed. If FDA wishes to discourage this response option, this should be clearly stated, with justification.

Submission of Documentation of Item Development (Line 462 to 464):

The Draft Guidance states, “The FDA intends to examine the final version of an instrument in light of its development history, including documentation of the complete list of items generated and the reasons for deleting or modifying items, as illustrated in Table 3.”

Please clarify what specific items of information from the item generation process the agency will want in a submission to examine the development history of an instrument. While the concept of an item history provides a convenient and brief account of the changes that a particular item may have undergone in the development process, it does not provide the substance or specifics of the information that created the history. It is not the same as viewing the cognitive debriefing transcripts, exploratory factor analysis, or analysis of floor and ceiling effects that may have produced a modification or elimination of an item.

Submitting the substance of the history of the items in an instrument can be a quite voluminous task. Does the Agency desire submission of the item histories alone, the item histories and all of the supporting documentation, or something in between such as the item histories and summary reports of the supporting documentation? Clarification of the desired level of documentation will give the Agency the precise information in the submission format that it needs for initial review.

Predictive Validity (Table 4 and Line 504):

The goal in most cases with PROs is to assess the current condition of a patient or the condition of the patient in the recent past. While the current condition of a patient often is correlated with future condition, this is highly dependent upon the disease / condition being assessed and the time-frame involved. Some conditions are reversible or cyclic in nature – and others are long-term. Two scales which are equally reliable and valid at assessing a patient’s current condition could have dramatically different ‘predictive validity’ – due simply to the topic being measured and not the scale. For these reasons, predictive validity could be mentioned as one tool for supporting validity – and very important in some cases, but the guidelines should not indicate that predictive validity be a core requirement for psychometric validation in all cases.

Minimal Important Difference (Lines 539 to 569):

In **line 539**, the Draft Guidance states the importance of defining a Minimal Important Difference (MID). The examples given [**lines 551-565**] include both anchor based and distribution based methods and then the Guidance suggests [**line 567**] that concordance confirms the choice. Given that protocols and Statistical Analysis Plans should *a priori* define MIDs for Phase 3, it may be difficult to perform *a priori* MID estimations with anchor based approaches given that efficacy in the study population has not yet been determined. It is likely that MID determinations in small Phase 2 trials prior to Phase 3 will most commonly be distribution based with confirmatory MIDs calculated in Phase 3 based upon relevant clinical anchors. The Guidance should allow for these considerations.

In **line 541**, the MID discussion delineates ‘mean effect’ from that minimal important difference to identify a responder. Because labeling will be derived based upon adequate and well-controlled studies of patients, the patient level discussion of ‘responders’ is not appropriate here as it is pertinent to ‘Section 4: Choice of methods for Interpretation’, but not to MID. We suggest that the patient level discussion of ‘responders’ be removed from this section; in **line 569**, there is the separate discussion of identifying responders outside the context of MID and this section could be retained.

Application to a New Population or Condition (Lines 610-615)

The Draft Guidance mentions disease severity level as one example of a new population or condition. The guidance should specify how a difference in disease severity is going to be defined. Unless a disease already has well-accepted severity levels that are clinically distinct, then severity may not constitute a new population or condition. Please allow some flexibility on what constitutes a new population or condition.

Similarly, the Draft Guidance suggests that re-validation is needed in a variety of cases. An example included in the Draft Guidance is when “Patients in the proposed trial differ in age, gender, race, or developmental or life stage from those for instrument development and validation” (lines 614-615). Please indicate the amount, type, and robustness of revalidation needed when a population or condition is slightly different from what was used to validate the instrument. Also, please indicate more clearly when a population from a clinical trial might be considered “new”. (For example, if an instrument was validated on a population that included people through age 70, presumably it may be used without revalidation on a population that includes some people between 70 and 75 years of age.

Multiplicity (Lines 876 to 922):

(Lines 876-880) – There needs to be clarification that the issue of multiplicity regarding the primary analysis should be addressed in the protocol – not simply the Statistical Analysis Plan. This is a core feature of the design of a study – with impact on the sample size and primary inferences.

(Lines 885-890) – When two or more variables are all required to be significant before any claim can be made – it is always the case that one does not need to adjust the significance levels downward (not ‘generally’). In fact, research suggests that usually a modest upwards adjustment to the significant levels could be performed. This is an important downside of requiring multiple measures to be significant, and the warning of this in this paragraph is helpful. It would be beneficial, however, to also have an expanded discussion of more

practical methods such as the gatekeeper approach and to contrast with the multiple significance approach.

**Missing Data General (Lines 973 to 1017):**

Overall, this section on missing data would be greatly strengthened by including a discussion of the mechanisms giving rise to the missing data: (e.g., Missing Completely at Random (MCAR), Missing at Random (MAR), and Missing not at Random (MNAR)). This is the standard approach in the literature for discussing missing data and is especially valuable when comparing the strengths and weaknesses of the various methods. It is difficult to compare and suggest methods without an appropriate framework and focus on assumptions utilized by each approach. For instance, likelihood based methods are valid under different missing data mechanisms than are other methods such as generalized estimating equations when assessing repeated measures.

(Lines **979-981**) – The Draft Guidance states, “FDA encourages sponsors to obtain data on each patient at the time of withdrawal to determine the reason for withdrawal. When available, this information can be taken into account in the analysis.” It is not yet clear in the literature how to best take such information into account. For instance, methods such as pattern mixture models are a possibility, but such methods are new and in their early stages of development. Thus, it seems inappropriate to make such a general comment without a specific recommendation.

(Lines **987-988**) – This paragraph discusses the completers analysis and states that this strategy is “generally inadvisable.” The point made in this and similar paragraphs is valid (no single gold standard, completers analyses may not be appropriate). As mentioned earlier, however, this discussion would benefit from the use of the standard missing data mechanisms. For instance, one could say that completers analysis is not recommended because the MCAR assumption necessary for its validity is unlikely to be true.

(Line **992**) – This paragraph discusses last observation carried forward (LOCF). Again – utilizing the standard mechanisms would be helpful. One could clearly state that LOCF requires the assumption of MCAR in addition to the time trend assumption. The mechanism discussion would be helpful in lines **1003-1007** as well.

(Lines **1012-1017**) – The concept of using a variety of imputation methods is appropriate. However, rather than mentioning a method such as ‘worst case imputation’ that generally is too conservative in the very situations of most interest (where there is a moderate to substantial amount of missing data), the guidance could provide more helpful suggestions. For instance, one could state that a variety of methods relying on different assumptions should be used in order to assess how departures from specific assumptions might have influence on the results. Again, this emphasizes focusing on the assumptions of each method and tying them back to the missing data mechanism.

### **Conclusions**

Again, Lilly believes that this Draft Guidance, when finalized, will provide much-needed information and general principles on evidence that can be used to support product labeling statements that are based on patient reported outcomes. We thank FDA for this initiative and for this opportunity to comment and look forward to continuing to work with FDA on this important topic.

Respectfully submitted,

ELI LILLY AND COMPANY

Catherine A. Melfi, Ph.D.  
Scientific Director,  
US Regulatory Affairs

Don Buesching, Ph.D.  
Research Advisor,  
US Health Outcomes Research