



Bristol-Myers Squibb Company

0664 '06 APR -7 A9:25

Joseph F. Lamendola, Ph.D.
Vice President
Global Regulatory Strategy
Pharmaceutical Research Institute

P. O. Box 4000 Princeton, NJ 08543-4000
Tel 609-252-3781 Fax 609-252-7781
Joseph.Lamendola@bms.com

April 6, 2006

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

Re: Docket No. 2006D-0044; Draft Guidance, Patient-Reported Outcomes Measures: Use in Medical Product Development to Support Labeling Claims (PRO Guidance), *Federal Register Volume 71, pages 5862-5863, (February 2, 2006).*

Dear Sir or Madam:

Bristol-Myers Squibb (BMS), a diversified global health care company, is pleased to have the opportunity to offer comments on the PRO Guidance. Our company's mission is to extend and enhance human life by providing the highest-quality pharmaceutical and related health care products. For this reason, we are interested in commenting on the PRO Guidance. Our comments are set forth below.

Summary of BMS Comments on Proposal

This document represents years of active collaborative exploration with the Agency, the pharmaceutical industry, academics and interested parties, and we would like to congratulate you on its release. PRO measures represent the very important contribution that patients can make to understanding the value of new medicines. The draft is well written and represents a compendium of information concerning development and execution of PROs sufficient to support labeling. This information is new for the Agency and represents a step forward. While the importance of the guidance for sponsors planning to obtain PRO labeling is obvious, we encourage the FDA to adopt reasonable pragmatism in the application of the PRO Guidance to the drug development process and to recognize research precedents in the PRO field are a means of bridging PRO measures and research deemed acceptable by the Agency.

BMS has prepared both general and specific comments. Our general comments concern:

- The notion of reasonable pragmatism;
- Aligning the standards for PRO development and use with standards for other clinical endpoints (The Level Playing Field);
- Processes for communication with the Agency, and especially with reviewing divisions, concerning PRO measures;
- General comments concerning Section IV

2006D-0044

213

Below, we discuss each of these concerns in more detail.

1. Reasonable pragmatism in the PRO guidance

In designing a measurement strategy for assessing PROs, it is frequently the case that the "perfect" instrument/measure to evaluate a PRO outcome of interest does not exist. Rather than simply 'selecting an instrument' researchers 'design a measurement strategy' consistent with the conceptual framework and the desired labeling. This measurement strategy may involve using previously developed and validated instruments in ways they have been used before, modifying or adapting already existing instruments for a new purpose, or even developing new questions or instruments.

Scientific precedent plays a key role in the process of developing a measurement strategy. Scientific precedent, properly considered, enhances the viability of a measurement strategy and, importantly, lends a foundation with documentation to the research path. The draft guidance appears to suggest that PRO measures that are modified in any way must involve a comprehensive psychometric development program de novo (see lines 178-181), thus the program would benefit only indirectly from scientific precedents that bear on the measurement strategy.

BMS believes that the standard for PRO development and use leading to labeling should rely on reasonable pragmatism for the inclusion of scientific precedents in PRO labeling submissions and documentation. The standard for validation and documentation should be flexible and should recognize the role of scientific precedent. We believe this aspect is consistent with the Critical Path Initiative.

2. Standards for PRO and Other Endpoint Measures

The draft guidance focuses on the methodological standards a PRO measure must meet to be acceptable to the Agency. It is recognized that PRO measures are frequently secondary endpoints in clinical trials. In general, the standards in the PRO Guidance appear higher or more stringent than standards used to assess clinical endpoints in trials. Since PROs are typically a secondary measure and not the primary basis for approval, the Agency should apply even-handedly standards consistent with clinical endpoints in similar circumstances.

Presently the PRO Guidance appears to suggest a more stringent requirement, especially concerning the requirements to validate PRO instruments/measures. We recommend a more practical or economically feasible standard. BMS is concerned that the current standard is onerous and overly concerned with relatively minor 'psychometric issues' and may inhibit the generation of information that will be welcomed by patients and physicians.

Finally, although the guidance focuses on 'PATIENT reported outcomes' do the same rules apply for caregiver/proxy-based outcomes? Is it possible to have a claim based on caregiver data? This line of research could be eminently reasonable for populations such as young children, the mentally or physically disabled, and Alzheimer's patients.

3. Processes for communication with the Agency and Reviewing Divisions

During the recent 'FDA Guidance on PROs' organized by the Mayo Clinic (February 23-25), there were many questions directed to the FDA concerning specific applications of the Guidance.

The ubiquitous response from the FDA was 'it depends'. In fact, during the FDA summation on the last day of the meeting the following was stressed in a slide:

"It Depends: On the concept underlying the claim, On the disease, On the target population for the product (e.g., severity, age, etc.), On the study design, On the expected treatment effect, and On the development stage."

It is well recognized in the PRO field that interaction with the Agency during various phases of drug development is advisable to consult on various aspects of the PRO work being performed. It would be very helpful if the guidance could address optimal interaction practices as they relate to PROs, with particular emphasis on the practices of the reviewing divisions. This is important in light of recent public comments by FDA staff noting that the application of these guidelines would be done on a "case-by-case" or "it depends" basis. It would be most useful to have a section in the document relating to the process of communication with the FDA. Will any discussion with the FDA have to be tied to a drug development program? Would FDA be willing to provide guidance during the development stage of instruments? Additionally, clarity regarding meeting conduct (e.g. separate EOP2 meeting, requested FDA meeting participants) relative to the incorporation of the PRO as a primary versus secondary endpoint would be valuable.

General comments on section IV

A. Validation of modifications to instruments/measures

Section IV of the guidance document provides recommendations regarding the validation of *all* modifications and translations that appear too burdensome. Modifications of existing measures that are minor should not require extensive re-validation. They should not be held to the same development and testing standard of newly developed instruments. Scientific experience engenders developmental value that should be acknowledged. In addition, a separate full validation for translations by language should not be required. Standard procedures short of psychometric testing should be adequate.

B. Recall periods

Longer recall periods (e.g., past week, past month...) are supported historically and extensively in the literature. The recall period should be guided by scientific precedent as well as by the conceptual framework for the measure. The concern over accuracy of recall is a matter of test-retest reliability over the recall period of interest – the error component is the burden of the sponsor to bear in study design and standardization of the assessment. Measurement error at random only serves to diminish the treatment effect. It is unnecessary for the guidance document to be prescriptive regarding the recall period. It should recommend that the recall period be reasonable for the condition and PRO under study. Assessing outcomes over an extended time period is often necessary since asking patients to report only their current experiences may not provide a representative sampling. It should also be acknowledged that the strength of memories is likely to vary for different populations, settings and therapeutic objectives. The statement that 'response over a period of time may threaten the accuracy of the PRO data' is too broad and not based on scientific evidence.

Specific Comments

Line 74: Although the guideline specifies that the *amount and kind of evidence that the FDA expects to support a labeling claim measured by a PRO instrument is the same as that required by any other labeling claim* we feel that the standard set for measurement tools for clinician reported outcomes such as HAQ (Health Assessment Questionnaire) are much lower. Further in Line 462 *It will be important to determine from empirical data submitted whether the conceptual framework has been demonstrated.* – this (as well as many other comments relating to validation, scoring algorithm, etc - line 475) would not be possible for most of the older instruments such as Karnofsky Scale, McGill Pain Questionnaire, Hamilton Anxiety Scale etc. Sensitivity to historical and scientific precedent is strongly in the final guidance.

Line 178-179 (& Line 581): *When considering an instrument that has been modified from the original, the FDA generally plans to evaluate the modified instrument just as it would a new one.* This position is too stringent. Minor modifications should be associated with less than a full psychometric validation requirement. There are many situations where knowledge and methods can be extended. In a clinical measure such as blood pressure, ambulatory monitoring (ABPM) has not entailed a comprehensive revalidation effort. In a similar context ABPM in Greece should not be fundamentally different from ABPM in another geographic setting. The scientific foundation for a measurement should influence the breadth and scope of the validation effort, with more experience translating to a lesser revalidation burden. Finally, we feel that measurement error at random can only serve to diminish treatment effects and the framework of a well controlled trial design with blinding and other features, this is a sponsor burden but does not fundamentally compromise validity.

Line 154: *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.* Whereas this may be true for symptom checklists, in general, PRO measures do provide an assessment of effectiveness of treatment taking both the efficacy and the adverse affect into consideration. This task should be dictated by the conceptual framework and the proposed claim structure.

Table 1 (bullet 3): Standard adverse event reporting should not be supplanted or replaced by patient reports. These aspects should remain separate and distinct.

Line 341: *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.* We feel that this is too stringent in some disease areas (e.g. such as those designed to assess migraine or epilepsy) where patients are expected to be symptom free at the time of evaluation and hence are expected to provide the average number of episodes over a time period.

Line 380: Although the points relating to *Development of Format, Instructions, and Training* are reasonable, two points are worth considering. (i) not all instruments are accompanied with instructions and procedures relating to administration and (ii) minor violations have little or no significance within the RCT framework.

Lines 400 to 401: While we welcome the Agency's review of ePRO instrument to provide sponsors with guidance on selecting the most qualified instrument, submission of the instrument in electronic format or in the format provided to trial subjects may provide unique issues. This guidance should clarify the Agency's expectations regarding the submission of the ePRO instrument and whether or not a description of the instrument with associated system diagrams would be acceptable.

Line 418: *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.* This is largely true for questionnaires developed to assess for symptom burden. In order to avoid overemphasis on similar items we suggest summary scores to be based on averages rather than on totals.

Line 547: *An MID is usually specific to the population under study.* We disagree. The MID is usually specific to study constraints (i.e. population / comparator / time period).

Line 654: *Sponsors should consider whether generally accepted standards for translation and cultural adaptation have been used to support the validity of data from a translated/adapted PRO instrument . . .* Whereas this is a reasonable we suggest that the agency encourage sponsors to carry out pre-specified analysis of summary scores to study differential item functioning between countries / cultures to identify items that may not have worked well in different settings in the RCT.

Line 710: *It is important that the protocol include the exact format and version of the specific PRO instrument to be administered.* This is reasonable if this means that the protocol should include the English version of the intended questionnaire (including the instructions to the patient and the study coordinators). However, given the timelines involved, it would be unrealistic to expect the actual CRF pages or the screen-shots of the electronic data capture device at the time of finalizing the protocol.

Line 717: *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.* We feel that this statement is too restrictive for Oncology. Open label studies are often the design of Oncology clinical trials, for registrational and research purposes. Oncology is a therapeutic area in which PROs are an essential tool in determining treatment choices. Given the guidance regarding a conceptual framework, accomplishing this step along with the rationale should dictate the design.

Lines 725-733: Blinding and Randomization. It is stated that access to prior responses can bias results when unblinding is a possibility. However, this question is not in agreement with the current literature and several authors have advocated that the access to prior response can improve the evaluation of the current response. Also, a PRO instrument with many questions is suggested as a way of limiting bias in the absence of blinding. However, using an instrument with too many questions can result in missing data that could also induce bias in the treatment comparison.

Line 774: *The frequency of PRO assessment should correspond with the demonstrated measurement properties of the instrument and with the planned data analysis.* We suggest the inclusion of 'expected time profile of treatment affect' as one of the attributes necessary to assess the frequency of PRO assessments.

Lines 824 to 825: While we agree with the principles stated regarding the responsibility of the investigator to prepare and maintain adequate and accurate case histories, we would like clarification of the terms "prepare" and "maintain" as they relate to ePRO, or PRO in general. This guidance should clearly explain the agencies expectations on how the investigator will fulfill his or her obligations.

Lines 827 to 835: We agree with the guidance recommendations regarding paper PRO instruments; however, the guidance should also reflect the Agency's view on an acceptable model for electronic source data that will meet the regulatory expectations of control, accessing and maintaining ePRO data. A recent DSI position (April 2004) states that a "Trusted Third Party" may be acceptable provided that the sponsor does not maintain the electronic source data. This system provides investigator site access, control of content (with audit trail, as necessary) and maintenance for the electronic source record. This stance appears contrary to Lines 830 to 832 of the draft Guidance which notes that data controlled by the sponsor or a contract research organization and not by the investigator "may pose a problem." This guidance should clarify how the agencies approach ePRO models that are web-based or use a server for electronic source records that are not physically located at the investigator site.

Lines 838 to 839: We commend the Agency for continuing to apply the recommendations from recent guidance documents on 21 CFR Part 11. We would like some further guidance on the topic contained within the "Computerized Systems Used in Clinical Trials" Guidance document regarding the statement "When original observations are entered directly into a computerized system, the electronic record is the source document." This guidance should clarify which record (or certified copy thereof) should the investigator control and maintain. This issue is of particular concern with eDiary systems which often involve "off-line" storage with subsequent uploading to a centralized server.

Line 852: *Removal of investigator accountability for confirming the accuracy of the data.* This is confusing – for example when data are collected using IVRS technology – investigators are completely left out of the data collection process and therefore will not be able to confirm the accuracy of the data. Line 852 also somewhat contradicts the statement in line 127 '*Self-completed questionnaires that are given directly to patients without the intervention of clinicians are often preferable to the clinician-administered interview and rating*' as well as the issue of patient privacy raised in Line 444.

Line 1003: *Some other approaches involve imputation of missing data on a per-patient basis.* We suggest this is reworded to reflect that methods to handle missing data may be based on imputation methods as well as those not based on imputation (such as mixed model, inverse probability weighting and pattern mixture) to study the robustness of the assumptions.

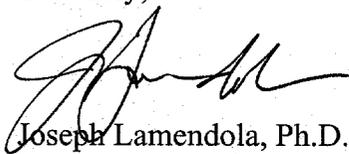
Line 1023: *If the MID is truly to be the smallest effect considered meaningful, however, it would be logical to establish the null hypothesis to rule out a difference less than or equal to the MID. This is rarely done, and would have major implications for sample size. We feel that the MID might be a means to quantify the size of the clinically meaningful difference and the sentiment reflected in Line 1023 is in contradiction with the statement in Line 886 (The statistical analysis considerations for PRO endpoints are not unlike statistical considerations for any other endpoint used in drug development).*

Line 1033: *When defining a meaningful change on an individual patient basis (i.e., a responder), that definition is generally larger than the minimum important difference for application to group mean comparisons. Currently, population MIDs are typically used to define responders. Does the agency suggest a separate larger MID should be used in the future?*

An additional question on the MID. At the Mayo Clinical Workshop on Patient Reported Outcomes, the FDA clarified that the MID should be applied to between group differences rather than within group differences. Our question is: since all the methods that derive the MIDs (distribution-based, anchor-based, or the Cohen's effect sizes) are all based on within group changes, can the MIDs that generated from these methods be used for between group comparison? For example, the MID for the Health Assessment Questionnaire (HAQ) for evaluating physical function is 0.22, can this number be applied to judge whether the difference between the active treatment group and the placebo group is clinically meaningful or not?

BMS appreciates the opportunity to provide comment and respectfully requests that FDA give consideration to our recommendations. We would be pleased to provide additional pertinent information as may be requested.

Sincerely,



Joseph Lamendola, Ph.D.
Vice President, Global Regulatory Strategy