Advancing Development and Use of Patient-Reported Outcomes in Drug Development: Near-Term Opportunities

October 6, 2014

Meeting Summary

Background

In clinical research, the most direct method for assessing a patient's symptoms, functional status, treatment preferences, or quality of life is through the collection of patient-reported outcomes (PROs). Drug developers, clinicians, patients, payers, and regulatory agencies are increasingly recognizing the importance of capturing the patient experience across the drug development spectrum, from early stage research through to post-marketing surveillance. There is particular interest among these stakeholders in using PROs to support labelling claims, which are a key source of information for both providers and patients.

The U.S. Food and Drug Administration (FDA) has made considerable efforts to facilitate the use of PRO instruments in clinical drug development. In 2009, the agency released final guidance to industry on the use of PRO instruments to support labeling claims. This final guidance outlined FDA's current thinking on the instrument development process, the evaluation criteria the agency will use to assess these instruments, and the major considerations related to clinical trial design and data analysis. FDA also developed a formal, voluntary process for qualifying drug development tools (such as PRO instruments) for use across multiple drug development programs. The goal of this process was ultimately to minimize redundant data collection and streamline the drug development process. Draft guidance on the qualification process was released in January 2014.

While these guidances represented an important step in formalizing and communicating the agency's approach to evaluating PRO instruments, progress has been slow. Only one PRO instrument has been fully qualified by the FDA, and PRO label claim approvals have declined slightly since the publication of the 2009 final guidance. Furthermore, most of the label claims granted between 2000 and 2012 were in cases where the PRO was specified as the primary endpoint, rather than as a secondary endpoint which can provide valuable supportive data on the effects of a drug.[1]

The barriers to accelerating the development and use of PRO instruments stem from a range of logistical, methodological, and communication challenges, which exist at both the industry and agency level.[*] As part of its work to facilitate the use of PROs in drug development, FDA is partnering with the Engelberg Center for Health Care Reform at the Brookings Institution to discuss the main barriers to their use, and identify promising strategies that can help to address them. This series of workshops will also help to inform a public meeting that FDA will hold in spring 2015, which will focus on the development, use, and qualification of clinical outcome assessments (COAs) more broadly, among other topics.

[*] A fuller description of these barriers is available here: http://www.brookings.edu/~/media/events/2014/07/16%20pro%20outcomes/brookings_pro_expert_workshop_discussion_guide_final.pdf

On July 16, 2014, the Engelberg Center convened the first in this series of expert workshops, "Enhancing the Development and Use of Patient-Reported Outcomes in Drug Development". Subject matter experts and stakeholders from FDA, patient advocacy groups, academia, and industry offered their perspectives on the implementation of the PRO guidance, with a particular focus on the evidentiary standards, strategies for streamlining the communication between FDA and sponsors, and new approaches for capturing the patient voice across the drug development continuum. This meeting generated a number of promising ideas, several of which were prioritized for near-term follow up.[†] In particular, FDA and Brookings identified several topics that would benefit from additional stakeholder input at a subsequent meeting. In particular:

Identification of potentially acceptable endpoints (PAEs)

The FDA's drug development tool qualification process provides confidence for both the agency and a sponsor that the instrument being used to support a labeling claim is fit for purpose (that is, a well-defined and reliable assessment of a particular concept). However, the instrument qualification process requires substantially more evidence than would be necessary within the context of an individual drug development program, and thus can require considerable time and resources to complete. For individual drug development programs, FDA could provide more guidance on which instruments are potentially acceptable, despite not having undergone formal qualification. As a part of this effort, FDA's Study Endpoints and Labeling Development (SEALD) team is collaborating with each of the Office of New Drug review divisions within the Center for Drug Evaluation and Research to identify a list of endpoints that are currently considered potentially acceptable for use in various therapeutic areas. Through publication of a potentially acceptable endpoints (PAE) list, FDA hopes to clarify details regarding the context in which certain endpoints are potentially acceptable as primary or secondary endpoints. Though the list would not be exhaustive, it could serve as a communication tool in early discussions between drug developers and FDA, as well as highlight the gaps in existing measures that would need to be addressed before they could be used to support labeling claims. FDA envisions that this list would be an ongoing effort, and would be reevaluated on a regular basis to ensure the information it contained was up-to-date and reflected current FDA thinking.

Adapting legacy instruments to be fit for purpose

Stakeholders have also noted that the current FDA guidances on PRO development and qualification incentivize the development of *de novo* instruments, rather than the adaptation of existing instruments. In many cases, this may unnecessarily slow instrument development timelines or lead to duplication of effort. However, modifying an instrument can affect the adequacy and interpretability of the measure, and potentially decrease confidence in the reliability of its results. In order to streamline the process for adapting existing instruments, FDA could seek input from experts on the feasibility and potential approaches for modifying existing instruments within the framework of 1) qualification and 2) for inclusion on the potentially acceptable endpoint measures list.

Meeting Objectives

In order to build on the momentum of this first meeting, a second expert workshop "Advancing Development and Use of Patient-Reported Outcomes in Drug Development: Near-Term Opportunities" was held on October 6, 2014 at the Brookings Institution. Through this expert workshop, FDA sought

[†] The full meeting summary is available here: http://www.brookings.edu/~/media/events/2014/07/16%20pro%20outcomes/pro%20expert%20workshop%201_meeting%20summary.pdf

specific input on the value and uses of a potentially acceptable endpoints list, using case examples to highlight the benefits and challenges associated with developing such a list. Participants also addressed possible implementation approaches that would help to ensure that the list was both valuable to sponsors and feasible to implement, and were invited to comment on a draft table that might be used to display the relevant information. (see Appendix I for the case examples and table draft provided to meeting attendees).

Participants were also asked to explore key evidentiary considerations for adapting existing PRO instruments to be fit for purpose within the context of qualification or inclusion on the potentially acceptable endpoint measures list. The day concluded with a discussion of next steps, and participants were asked to identify any outstanding topics that present methodological or regulatory uncertainty or have hindered the development of and confidence in the use of PRO instruments in clinical trials. These topics will be consolidated into FDA's research agenda and possibly explored at future meetings.

Potential Value and Drawbacks of a Potentially Acceptable Endpoints (PAE) List

The workshop began with a high-level discussion on the strategic goals and value of developing and publishing a PAE list. Participants were generally supportive of the idea of developing and maintaining a PAE list, and agreed that the current framework developed by FDA (Appendix I) was a solid foundation from which to build. The potential benefits of such a list included increased sponsor confidence about PRO endpoints and instruments with which FDA might have a certain level of comfort, improved clarity on aspects of particular measures that might need modification or further evidence to be considered fit-for-purpose for a given context of use, and identification of therapeutic areas where more work is needed to develop PRO endpoints. Others noted that the PAE list could also help illustrate how various divisions within FDA interpret the PRO guidance, and identify areas where more coordination is needed.

Participants also noted potential drawbacks of a PAE list that would require close attention. In particular, some raised concerns that the list might undermine the qualification process, which requires the collection of more substantial evidence and is therefore more costly and time-consuming. More specifically, the PAE list could signal to drug developers certain measures that FDA considers "good enough" to be used in individual drug development programs (with supporting evidence), therefore decreasing the need for qualifying an instrument. However, the case example from the PRO Consortium's IBS working group (See Appendix I) suggests that, while the initial urgency to develop an IBS endpoint faded after FDA publication of provisional IBS endpoints, there were several other factors at play that likely slowed down their progress. Participants also raised concerns that the PAE list could hinder innovation in endpoint development. Drug developers might be incentivized to select endpoints that have already achieved some level of acceptability, rather than explore new and better ways to better capture treatment benefit. Concerns over feasibility were also raised. For complex domains such as pain—which may have multiple possible instruments that could be used to capture data—it can be very challenging to know how best to extrapolate based on prior approvals. There are numerous considerations that must be taken into account, such as mode of administration, wording, type of measure, type of pain intensity, adequacy of translation, and cultural validity.

Considerations for Developing the PAE List

There was substantial discussion over what information should be included as part of the PAE, and participants noted a lack of clarity over whether the list was primarily focused on endpoints or the endpoint measures that would be used to collect data. Given this confusion, participants stressed the

importance of clear communication about what information is contained in the table, the goals for its development and use, as well as the process that the agency will use for developing and updating the PAE list. In particular, it will be necessary to clearly distinguish between the concepts of interest, the associated endpoints, and the instruments for measuring those endpoints. It was suggested that FDA establish a clear timeline for re-evaluating the list to determine if the endpoints are still acceptable, and that the agency also develop a process for handling situations in which a sponsor has selected an endpoint which is later deemed to be unacceptable.

A systematic method for selecting endpoints for inclusion was emphasized as being foundational. Where possible, this process should utilize systematic reviews and meta-analyses in the literature to support list development, and to rely on existing databases that contain information on PRO instruments where possible (such as PROQOLID or the GEM database). Given FDA resource constraints, participants also suggested that the Agency consider establishing an independent group to supplement the internal work done to build and maintain the list. Such an independent group could ensure that the list reflects current scientific understanding in the field, and could also suggest improvements in its design. Participants stressed that the patient perspective would also be important in identifying which endpoints should be prioritized for inclusion, as the PAE list would ideally align with outcomes that are most meaningful to patients.

Participants also discussed the possibility of capturing additional information in the PAE table, such as the methodological concerns associated with an instrument, or information on what constitutes clinically meaningful change. It was generally agreed that the table would be more useful if it clearly highlighted where gaps exist, which would help sponsors to understand the risks associated with selecting a particular endpoint or measure.

Several questions remain over how best to format the PAE list. Some participants suggested that the table should connect more closely to the Roadmap that FDA developed as part of its guidance to industry on the qualification process (i.e. from left to right, the columns would begin with the concept of interest, then provide information on the context of use, and then provide certain information on the instrument(s)). It was also suggested that two separate (though clearly linked or nested) tables might accomplish the overall goals better; one that identifies potentially acceptable endpoints and their gaps and another that focuses on instruments, what is known about those measures, and also what gaps exist. The amount and type of information contained in the table, the timeline for updating that information, and the process for selecting instruments for inclusion are still being determined. More work is required before it will be ready for use, and FDA was encouraged to continue its process of stakeholder consultation and revision.

Considerations for Implementation and Use

Regardless of the final form, it was noted that a clear communication framework for the use of the list, which describes how the information should be interpreted and deployed by sponsors and other stakeholders, be developed. One suggested framework included the following steps:

- **Step 1**: Identification of the salient outcomes/concepts in a population (e.g., through qualitative work in phase 2 or literature review)
- **Step 2**: Selection of a measure that represents those outcomes

  - Ideally, sponsors could use an existing tool without further testing if it meets particular criteria (i.e., it is qualified, or has been shown to be acceptable in one related or three or more unrelated populations)
- **Step 3**: Identification of endpoints that adequately represent those outcomes and are supported by those measures

This final step would be conducted through collaborative work between sponsors and FDA review divisions. It was noted that the advice given to sponsors through this process should be consistent, which would help to ensure consistency across multiple trials. Several participants agreed that this was a potentially useful framework, though some raised methodological concerns over whether an instrument could be considered "acceptable" based on evidence from just three unrelated populations.

Participants supported this framework and also suggested that it might be helpful to frame the instruments within a risk continuum, whereby qualified instruments are categorized as low-risk (e.g., qualified instruments that do not require supporting evidence of the instrument validity), potentially acceptable instruments are categorized as medium-risk (that is, potentially acceptable in certain contexts but with identified gaps), and instruments with little supporting evidence are higher-risk.

Also emphasized throughout the meeting was that the list be implemented in a way that clearly links to the qualification process. Some suggested that this list be on the pathway toward qualification. For example, if an endpoint measure from the PAE list is used within an individual drug submission (with supporting evidence for its fitness-for-purpose), that the instrument could then be considered as being further along toward qualification. Any remaining evidentiary gaps could then be noted in the PAE list. Stakeholders suggested that such an approach might help to address concerns that the PAE list would undermine the qualification process.

Adapting Legacy Instruments to be Fit-for-Purpose

Developing a *de novo* instrument is a time-consuming and laborious process, and participants generally agreed that there is substantial opportunity to modify existing instruments to support labeling claims. Adapting existing instruments could reduce unnecessary duplication of effort as well as shorten instrument development timelines, therefore incentivizing their uptake in clinical trials. The motivation for adapting an existing instrument may also be related to the instrument itself. For example, the instrument may target the right concept of interest but may lack precision or require adjustment to make it valid within a new context of use. Conversely, the instrument may not target the right concept of interest, but may contain a subset of items that are potentially usable for a particular population. However, there are a range of considerations that must be made in order to ensure that such instruments are fit for purpose under the current FDA guidance.

Participants noted that the process for selecting and adapting an existing instrument should be careful and systematic. Changing a validated instrument too much can reduce confidence in the validity of the results, and can make it difficult to compare those results against past studies or aggregate data based on that instrument. However, it was also noted that further research could be done to compare a modified instrument to the original versions, and that this could help to increase confidence in its use. Ultimately, the scope of the effort required will depend on the starting place of the instrument, how much prior work has been done on the measurement properties, and the context of use.

Participants further suggested that any process for adapting an instrument should begin with identifying and defining the concept (or concepts) of interest that are most meaningful to the patient population in question, and that this initial research be used to inform the decision on whether to adapt an instrument or develop a new one. Whatever instrument is selected to measure that concept, it must clearly connect to the claim that a sponsor wants to make about the drug. Participants requested more clarity from FDA on whether and how the list of potentially acceptable endpoints can inform the selection of an instrument for modification and, ultimately, qualification. Furthermore, copyright issues associated with modifying existing instruments were expressed as well.

Strategies to support instrument adaptation and use

Given the challenges of aligning instrument development with clinical development timelines, it was reiterated that early consultation with the FDA would be helpful for sponsors as they make decisions about 1) whether to adapt an instrument or develop a new one, and 2) optimal approaches to selecting and adapting an instrument. As noted at the previous Brookings PRO meeting, there are structural barriers within industry that disincentivize significant investment in PROs, and many companies are averse to taking on the additional risk of incorporating a PRO endpoint that has not already been used to support a labeling claim. Participants suggested that a meeting could focus on PRO-specific concerns. Such a meeting would ideally take place early in the development process, though several participants noted that this may not be feasible in every case.

It was also suggested that sponsors might consider an enriched or parallel clinical trial strategy, whereby a PRO instrument is used to measure effect size for a subgroup of particularly symptomatic patients. Participants also noted that there are many trials currently underway that use instruments that are well-accepted in the literature, but which are principally used in the European context to support reimbursement decisions. It was suggested that sponsors do more work to refine this data that is already being collected and provide more detail to FDA, which could help increase confidence that these instruments are adequate for use within the contexts they are being applied. Several participants also noted the potential role of the PROMIS Initiative in adapting existing instruments. PROMIS maintains an item bank that developers can potentially use to adapt or build an instrument, perhaps with direct patient input. Recent research (which was cited during the discussion) has also supported the use of PROMIS instruments for certain domains.

Next Steps

At the conclusion of the workshop, participants identified several next steps for FDA, industry, and other stakeholders that could help to advance the discussion around these and other issues related to PRO development and use.

Continue to develop the PAE list

Participants reiterated that the table represents a good first step, though the specific columns will likely require further refinement and consultation with stakeholders. FDA might also consider developing two tables rather than one, as discussed in the section above. Regardless, it will be important for FDA to embed this table within a framework that guides its appropriate use by sponsors, and that sponsors continue to engage regularly with FDA about their specific drug development programs. Participants also acknowledged that FDA resources are limited, and that if the agency should choose to expand this list beyond what was originally proposed, it should consider partnering with other organizations to help support the necessary work. Voluntary advisory boards or public-private partnerships may be useful

mechanisms for obtaining regular feedback and input into the table. Where possible, the agency should also rely on existing information sources on PRO instruments (such as PROQOLID).

Consult with stakeholders to further refine a PRO research agenda

As was highlighted in the afternoon discussion on adapting existing instruments, there are many outstanding scientific issues that hinder the development and use of PRO instruments to support labeling claims (e.g., the use of PROs in open label studies, or the cross-cultural adaptation of instruments in multinational trials). Further research is required to address these questions, and as a next step FDA could consider establishing a process to consult with academic thought leaders and other stakeholders on developing a research agenda that will identify those scientific questions and prioritize them for action. In the FDA background document circulated prior to the meeting, FDA identified several topics that might inform a broader research agenda on PRO development, adaptation, and use:

- Interpretation of PRO data from open-label trials (i.e., can PRO data in unblinded studies be relied upon?)
- Interpretation of meaningful change; methods for determining clinically meaningful change
- Amount and type of additional assessment needed to use an existing instrument in a new population or make other modifications to an existing instrument
- Methodology of using a "pick your most bothersome symptom(s)" approach in diseases or conditions where symptoms vary across individuals
- Relative merits of "worst symptom severity," "overall symptom severity," or "average symptom severity" over past 24 hours
- Use of averages of weekly scores versus all the data points in clinical trials
- Use of mixed methods in instrument development such as concept mapping and content analysis
- Cultural adaptation of PRO measures for use in multinational trials
- Use of social media in instrument development
- Use of computerized adaptive testing in clinical trial outcome assessments

Participants generally agreed that the identified topics were important areas to explore. As a next step, FDA could consider subsequent small workshops that could serve to further develop this research agenda, as well as identify partners who could help implement it.

Appendix I

A. Potentially Acceptable Endpoints List Outline for Draft Guidance Publication

Condition / Symptom	Indication	Concepts of Interest[1]	Instrument or Assessment Name	Context of Use				Guidance Available or DDT in Progress
				Endpoint Definition	Endpoint Model (Endpoint Hierarchy)[2]	Clinical Trial Population Characteristics[3]	Clinical Trial Protocol Key Elements[4]	

[1] In some cases, no instrument currently exists to measure relevant concepts. When concepts are identified as being relevant and important to patients (e.g., through patient-focused workshops), those concepts will be listed here, noting a GAP in availability of instrument or assessment tools for these concepts.

[2] Describes whether the instrument supports primary, co-primary, or secondary endpoint (and, if applicable, other key endpoints that must be considered)

[3] Demographic, cultural, clinical subpopulations

[4] May include trial setting, duration, study design (e.g., non-inferiority trial)

*This list was devised based on previously approved labeling. Concept of interest, instrument, and context of use criteria identified in this list may or may not be applicable for each specific drug development program. Sponsors are encouraged to discuss the specific drug development program needs with the review division.

B. Case Study #1: Irritable Bowel Syndrome (IBS)

In order to explore the possible advantages and disadvantages of the potentially acceptable endpoints initiative, it would be useful to consider the impacts of a similar recently used approach in which FDA attempted to expedite drug development by defining specific provisional endpoints in guidance that could be considered in supporting the approval of drugs for the treatment of the specific subtypes of irritable bowel syndrome (IBS). The applicable details of the IBS provisional endpoints are discussed below. Feedback concerning this approach from IBS drug developers, PRO Consortium working group members, and patients will form the background for the discussion of the potentially acceptable endpoint initiative.

In the past, IBS clinical trials commonly used a single-item patient-reported rating of overall change in condition as the primary efficacy endpoint. Specific IBS signs and symptoms were included as separate secondary endpoints. Examples of single-item patient ratings of change included questions posed to patients about “adequate relief” or “satisfactory relief” or Subject Global Assessment of Relief (SGA) of IBS symptoms. With the publication of FDA’s 2009 Guidance for Industry: “Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims” (the PRO Guidance), FDA’s assessment of PRO measures used to support labeling claims evolved. FDA determined that PRO measures, such as patient ratings of change, were not based on sufficient qualitative research to support conclusions that they captured treatment benefit in a well-defined and reliable way. FDA acknowledged, however, that although a content-valid PRO instrument that measures the clinically important signs and symptoms associated with each IBS subtype would be the ideal primary efficacy

assessment tool in clinical trials used to support labeling claims, such a measure was currently unavailable.

In 2008, Critical Path Institute launched its Patient-Reported Outcome Consortium. The PRO Consortium's IBS Working Group was formed in 2009 to fill an unmet need and develop (and ultimately qualify) measures that could be used as primary endpoints in IBS subtype-specific clinical trials. Shortly after the IBS working group was formed, FDA (also in response to the same unmet need identified by the PRO Consortium) published a draft guidance in 2010 (and revised final guidance in 2012) that summarized provisional standards for IBS subtype trials designs, including primary endpoints.

The development of the suggested provisional endpoints and responder definitions were based on several factors including literature review, FDA's review of qualitative data from failed IBS drug development programs, and expert opinion. Unlike the potentially acceptable endpoints project under development, the IBS provisional endpoints had not been successfully used in a clinical trial to support an approved indication. However, similar to the potentially acceptable initiative, the IBS provisional endpoints lacked complete documentation of evidence that the instruments were well-defined and reliable measures of the concept of interest and were still considered "potential" or "provisional," as opposed to having a more definitive interpretation (i.e., a qualified measure can be relied on to have a specific interpretation and application).

C. Case Study #2: Pain

In order to consider some of the operational challenges of developing a list of potentially acceptable endpoints, it may be illustrative to ground the broader discussion using the example of pain. According the FDA's Draft Guidance for Industry – Analgesic Indications: Developing Drug and Biological Products released in February 2012, "Pain can be categorized according to its duration, acute, or chronic, as well as based on other characteristics, such as breakthrough pain, acute episodes of pain that occur on a background of well-controlled, chronic pain. Pain is subjective in nature and is measured by patient self-reporting of its intensity, and other subjective qualities." Pain has been measured in clinical trials using a variety of instruments such as a single item Numerical Rating Scale (NRS) or Visual Analog Scale (VAS) among others. Different aspects of pain can be measured (e.g., intensity, frequency, duration), and pain measures can ask patients to consider worst pain, overall pain, average pain, typical pain, etc. In the context of the potentially acceptable endpoint project, there are challenges in suggesting, classifying and communicating provisional endpoints measuring pain due to the multiple approaches to pain measurement and varying experience with different approaches across different conditions and therapeutic review divisions.

The potential exists to include on the potentially acceptable endpoints list different assessments of pain for various conditions based on successful use in the context of one NME program or therapeutic area. For example, worst pain intensity on an NRS might have been previously used and is found potentially acceptable for condition X, while worst pain intensity on a VAS might have been previously used and is found potentially acceptable for condition Y. However, evidence to support the use of one measurement approach over another may be lacking.

Considerations exist related to best outlining and communicating such provisional endpoints measuring concepts, like pain, that are caused by a variety of conditions. Might these endpoints be considered as potentially acceptable for the domain of pain across multiple conditions and treatments?

We wish to consider the potential implications of identifying various measurements (e.g., pain assessments) as potentially acceptable when there is still needed research to identify the most appropriate measure of the concept (e.g., response scale, recall, aspect and considerations like worse vs average). We seek to understand the pros and cons of including endpoints on the potentially acceptable endpoints list when evidence is incomplete. In addition, while the documented evidence in line with the principles in the PRO Guidance may not exist for existing measures, might we consider qualification (rather than just including on the potentially acceptable endpoints list) for some assessments for which we have a great deal of experience and are generally thought to work well enough in clinical trials to support approval and claims of efficacy? What criteria should be used to make decisions about qualification verses inclusion on the potentially acceptable endpoints list?

[1] Gnansakthy A, et al. A Review of Patient-Reported Outcome Labels in the US: 2006 to 2010. Value in Health. 2012;15:437-442. Retrieved May 22, 2014 from: http://www.sciencedirect.com/science/article/pii/S1098301511036011