Clinical Outcome Assessments: Establishing and Interpreting Meaningful Within-Patient Change
The Center for Strategic and International Studies • Washington, DC
April 4, 2017

Discussion Guide

In recent years, patients, advocates, researchers, regulators, and other stakeholders have highlighted the importance of engaging patients more directly in medical product development and the regulatory review process. A key focus of these efforts has been advancing the science of patient input: identifying rigorous, systematic approaches to incorporating patient perspectives in medical product development, developing patient centered outcomes, and applying innovative drug development tools to capture data on those outcomes. A critical component of this advancement has been the development and implementation of well-defined and reliable clinical outcome assessments (COAs), which measure how a patient feels, functions, or survives, and determine whether or not a drug provides clinical benefit.¹

Most registration trials utilize COAs as primary endpoints, as pre-specified secondary endpoints, or as other endpoints that provide supportive information on the drug’s effects.² However, there are ongoing challenges with the use of COA endpoints in the medical product development process. One of these challenges relates to the interpretation of meaningful within-patient change in the outcome assessment over time – specifically, how to establish the threshold of change in the measure that can be interpreted as being clinically meaningful, rather than just statistically significant. The US Food and Drug Administration (FDA) published guidance for one type of COA, patient-reported outcomes, in 2009, which includes some considerations and recommendations for interpreting meaningful within-patient change. However, challenges remain and further conversation is needed to address outstanding issues, including how best to choose and apply the appropriate method for deriving thresholds to interpret change, as well as whether and how emerging methods might be applied in certain contexts.

To support further progress in this area, and under a cooperative agreement with FDA, the Duke-Margolis Center for Health Policy is convening this expert workshop in order to 1) explore and discuss methodologies and best practices surrounding meaningful within-patient change, and 2) identify specific recommendations on methodologies used to derive and interpret meaningful within-patient change with use of COA endpoints in medical product development.

Clinical Outcome Assessments in Medical Product Development
COAs are generally divided into four broad categories, depending on how the assessment is conducted and reported: patient-reported outcomes (PROs), clinician-reported outcomes (ClinROs), observer-reported outcomes (ObsROs), and performance outcomes (PerfOs). A description of each of the four types is provided in Table 1 below.

Table 1: Clinical outcome assessment definitions³

<table>
<thead>
<tr>
<th>COA type</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Patient-Reported Outcomes (PROs)</td>
<td>A measurement based on a report that comes directly from the patient (i.e., study subject) about the status of a patient’s health condition without amendment or interpretation of the patient’s response by a clinician or anyone else.</td>
</tr>
</tbody>
</table>
### Clinician-Reported Outcomes (ClinROs)
A measurement based on a report that comes from a trained healthcare professional after observation of a patient’s health condition. Most ClinRO measures involve a clinical judgment or interpretation of the observable signs, behaviors, or other manifestations related to a disease or condition. ClinRO measures cannot directly assess symptoms that are known only to the patient.

### Observer-Reported Outcomes (ObsROs)
A measurement based on a report of observable signs, events, or behaviors related to a patient’s health condition by someone other than the patient or a healthcare professional. Generally, ObsROs are reported by a parent, caregiver, or someone who observes the patient in daily life and are particularly useful for patients who cannot report for themselves (e.g., infants or individuals who are cognitively impaired). An ObsRO measure does not include medical judgment or interpretation.

### Performance Outcomes (PerfOs)
A measurement based on a task(s) performed by a patient according to instructions that are administered by a health care professional. Performance outcomes require patient cooperation and motivation.

COAs may be further subcategorized in a variety of ways, such as by the concepts being measured, the method and mode of administration, data collection, or analysis, or whether the measure is specific to a particular disease, population or is ‘generic’ (i.e. applies across diseases or populations).

### Interpreting Meaningful Within-Patient Change in Medical Product Development
The usefulness and relevance of a COA endpoint relies on its ability to accurately capture the impact of an intervention and communicate those results to patients and other stakeholders. Of particular importance is the degree to which an observed change in a measure can be interpreted as clinically meaningful and indicates a treatment benefit. For example, when measuring a concept such as pain intensity using a COA, the meaning of a 1-point change on a 0- to 10-point pain scale is not readily apparent. In other words, it is not self-evident whether these numerically small changes translate to a clinical benefit for patients.

Methods for deriving meaningful thresholds for interpreting change have been subject to much debate, and there is currently no established consensus on best practices, particularly in medical product development. Furthermore, most of the work to date on methods for interpreting meaningful within-patient change has focused almost exclusively on PROs. While some principles can be applied broadly across all COA types, there are context-specific issues that still need to be addressed for ClinROs, ObsROs, and PerfOs.

FDA released final guidance on the use of PROs to support labeling claims in 2009. In addition to outlining the process for developing and evaluating a PRO, the guidance clarified FDA’s thinking on a number of areas regarding the interpretation of meaningful within-patient change. Although group-level or individual-level responses to treatment may be used to support an efficacy claim, the guidance emphasizes the importance of providing individual-level responses to treatments to help with the interpretation and communication of treatment benefit and to guide treatment decision-making. These individual level responses should utilize an a priori responder definition, which represents the threshold for within-patient change – i.e., the change in a measure over time that should be interpreted as a meaningful level of clinical benefit. This responder definition should be determined empirically, in the context of intended treatment indication. The guidance highlighted one primary approach, anchor-based methods, to derive meaningful within-patient change.
Anchor-Based Methods for Establishing Meaningful Within-Patient Change

Anchor-based methods examine the associations between the concept of interest targeted by a COA and the same or closely related concept measured by independent measures, often other COAs. For the anchoring measure, external criteria are used to identify patients who have experienced a clinically important change in their condition in order to derive the threshold(s) for interpreting change. These thresholds are then used to establish the meaning of a change on the measure of interest. To be useful, the selected anchor(s) should be plainly understood in context, easier to interpret than the COA itself, and sufficiently correlated to the targeted COA. For example, in one trial evaluating a treatment for severe itching caused by plaque psoriasis, the Itch Severity Score was used to evaluate treatment impact. To help interpret the change in itch severity scores, a PRO known as the Patient Global Assessment of psoriasis disease severity was employed as the anchor.

Although anchor-based methods have been the most widely used, there are some limitations to this approach. For anchor-based methods to be applicable, the anchor should be associated with the COA, and the patients must have experienced a measurable meaningful change related to the condition or disease being evaluated. However, if there are no adequate anchors, if there is little detectable change in the anchor or in the COA of interest, or if there is not sufficient clinical trial experience to draw from, it may not be possible to derive interpretation thresholds. This can be particularly challenging in disease contexts where patients experience a wide variety of symptoms and levels of disease severity, making it difficult to not only develop COAs that reflect the heterogeneity of their disease, but also define what constitutes a meaningful change.

Furthermore, there are challenges with evaluating the direction of the meaningful change, specifically how to derive thresholds that interpret stability or worsening on the COA, not just improvement. In the development of COAs, the goal of a treatment is usually to demonstrate an improvement for patients. However, for some diseases, particularly in oncology, an effective treatment may be one that slows or halts deterioration in patient functioning. In these situations, thresholds that correspond to worsening must be established, but difficulties remain in quantifying these thresholds. (For additional background information on this approach, please see Attachment I.)

Emerging Novel Methods: Opportunities and Challenges

Given the challenges and limitations associated with anchor-based methods, there is substantial interest in identifying alternative approaches that may be useful or supportive in interpreting clinically meaningful change. Several emerging methods may have the potential to serve as alternatives or supplemental approaches to anchor-based methods. Examples of these novel approaches include bookmarking/standard-setting, scale-judgment, and exit-interview methods. (Note: this list is not exhaustive and includes only some examples of novel approaches; other emerging methods should be considered in the broader context.)

In the bookmarking/standard-setting method, patients and experts are presented with clinical vignettes of a disease in order to reach a consensus on thresholds for severity levels to interpret change. (See Attachment II.) In the scale-judgment method, panels of judges evaluate pairs of completed tests to determine whether the amount of change specified by the responses before and after a treatment indicate an important difference. (See Attachment III.) The exit-interview method is a more qualitative approach, whereby patients who recently completed a clinical trial are interviewed to provide insight into how observed changes in the measure of interest compare to their own perception of treatment benefit. (See Attachment IV.)
Each of these methods has advantages and limitations that may make them appropriate for use, depending on the COA, therapeutic area, or type of intervention. However, these methods are still relatively new and will require further research and stakeholder discussion regarding their suitability for use in clinical trials.

**Regulatory Efforts to Support COA Development and Implementation**

Since publishing its 2009 guidance on the use of PROs in medical product development, FDA has taken a number of important steps to facilitate the development and uptake of more patient-centered COAs and to better incorporate the patient voice into regulatory decision-making. Many of these efforts have been driven in part by the FDA’s obligations under the 2012 re-authorization of the Prescription Drug User Fee Act (PDUFA V), which specifically committed FDA to advancing the use of PROs in medical product development. PDUFA V also directed the FDA to develop a new program known as the Patient-Focused Drug Development (PFDD) Initiative. Through this initiative, FDA aims to systematically gather patient input and perspectives on a number of specific disease areas and the available treatment options for those conditions. Since that time, FDA has held 20 PFDD meetings, and will complete four more by the end of fiscal year 2017.

The FDA has signaled its ongoing dedication to this effort in its recently released set of draft commitments for the upcoming reauthorization of PDUFA in 2017. Under these commitments, the FDA will develop a series of guidance documents that will concentrate on approaches for translating these initial PFDD meetings into COA tools that can be used to collect meaningful patient and caregiver input. A better understanding of the best practices and emerging methods for deriving and interpreting meaningful within-patient change will play an important role in informing next steps around this process.

**Meeting Objectives**

The purpose of this workshop is to advance the discussion on meaningful within-patient change by identifying best practices and developing specific recommendations for methodologies that can be used to derive meaningful within-patient change. Discussion will encompass special considerations for establishing meaningful change in small and heterogeneous study populations, as well as how threshold determinations may differ across the four types of COA (patient-reported outcome, observer-reported outcomes, clinician-reported outcomes, and performance outcomes).

**Session I: Exploring the Use of Emerging Methods to Derive and Interpret Meaningful Within-Patient Change: Bookmarking/Standard-Setting**

**Objective:** While there has historically been an emphasis on anchor-based and distribution-based methods for interpreting change in COA scores, emerging methods may have the potential to serve as alternatives or supplemental approaches. This session will examine the applicability and validity of the bookmarking/standard-setting methods and highlight any remaining questions in assessing its appropriateness of use.

**Questions to address:**

- When is this approach useful in clinical trial settings (e.g., special populations such as pediatric, rare disease, etc.)? How do we operationalize this approach? How feasible is this approach?
• How would this approach for deriving and interpreting meaningful change differ across the four types of COA (patient-reported outcome, observer-reported outcomes, clinician-reported outcomes, and performance outcomes)?
• What are the advantages and disadvantages of the bookmarking/standard-setting methods for use in drug development?

Session II: Exploring the Use of Emerging Methods to Derive and Interpret Meaningful Within-Patient Change: Scale-Judgment

Objective: This session will examine the applicability and validity of another emerging approach - the scale-judgment method - and highlight any remaining questions in assessing its appropriateness of use.

Questions to address:
• When is this approach useful in clinical trial settings (e.g., special populations such as pediatric, rare disease, etc.)? How do we operationalize this approach? How feasible is this approach?
• How would this approach for deriving and interpreting meaningful change differ across the four types of COA (patient-reported outcomes, observer-reported outcomes, clinician-reported outcomes, and performance outcomes)?
• What are the advantages and disadvantages of the scale-judgment method for use in drug development?

Session III: Exploring the Use of Emerging Methods to Derive and Interpret Meaningful Within-Patient Change: Exit-Interviews

Objective: This session will examine the applicability and validity of another emerging approach - the exit-interview method - and highlight any remaining questions in assessing its appropriateness of use.

Questions to address:
• When is this approach useful in clinical trial settings (e.g., special populations such as pediatric, rare disease, etc.)? How do we operationalize this approach? How feasible is this approach?
• How would this approach for deriving and interpreting meaningful change differ across the four types of COA (patient-reported outcomes, observer-reported outcomes, clinician-reported outcomes, and performance outcomes)?
• What are the advantages and disadvantages of the exit-interview method for use in drug development?

Session IV: Exploring the Use of Anchor-Based Methods to Derive and Interpret Meaningful Within-Patient Change

Objective: The 2009 PRO Guidance highlighted the importance of traditional anchor-based methods for interpreting meaningful change of COA scores. This session will tackle outstanding questions related to the implementation of anchor-based methods, including techniques for choosing the appropriate anchors.

Questions to address:
• When is this approach useful in clinical trial settings (e.g., special populations such as pediatric, rare disease, etc.)? How do we operationalize this approach? How feasible is this approach?
• What are best practices to selecting an appropriate anchor?
  o What are the qualities of an anchor that should be considered?
  o Does the anchor need to be of the same type of COA as the target COA? (e.g., can a PerfO or ClinRO serve as an anchor for a PRO? Or vice versa?)
  o Are there situations where anchors may not be necessary (e.g., single global item with verbal rating scale)?
• How would this approach for deriving and interpreting meaningful change differ across the four types of COA (patient-reported outcomes, observer-reported outcomes, clinician-reported outcomes, and performance outcomes)?
• What are the advantages and disadvantages of the anchor-based methods for use in drug development?

Session V: Reflecting on the Methods for Deriving Thresholds for Interpreting Meaningful With-Patient Change

Questions to address:
• What factors should be considered when selecting among methods for documenting meaningful within patient change?
• What are the considerations for meaningful within-patient change when patients cannot report for themselves (e.g., pediatric, cognitively impaired)?
• What are additional advantages and disadvantages that have not been discussed for any of these methods for use in drug development?
• What are some of the other emerging methods that have not yet been discussed?