Developing Personalized Clinical Outcome Assessments
April 5, 2017

Meeting Summary

Background

Over the last several decades, there has been growing recognition of the importance of incorporating the patient perspective into medical product development. A key component of this effort has been the development and implementation of fit-for-purpose clinical outcome assessments (COAs) that accurately and reliably measure meaningful treatment outcomes and can be used as study endpoints. Ensuring that COAs are patient centered – that they identify and assess outcomes that are relevant and important to patients – is critical to their usefulness. One challenge in the development and interpretation of COAs is the heterogeneity of signs, symptoms, and functional impairment that patients exhibit in numerous disease areas. Patients may experience some, but not all, of the symptoms associated with their disease or condition (as with many rare diseases), or may experience one common symptom but exhibit significant variation in other symptoms (such as with migraines).

In such cases, it is challenging to develop a single outcome assessment that is relevant and applicable to the entire target patient population. A standardized COA applied uniformly to all patients may overlook symptoms that are important to some patients or include symptoms that are irrelevant to others, potentially rendering the assessment insensitive to meaningful treatment benefit. A possible solution to this problem includes employing a “personalized COA” approach, in which the signs, symptoms, and functions assessed may vary across patients in an effort to measure the most important and relevant outcomes in each individual. While this approach shows promise, there are a number of challenges relating to their development, implementation, and analysis.

The Potential for Personalized COAs

The concept of personalized COAs is still emerging, and there is currently a paucity of information on best practices surrounding the methodologies and analytical approaches to take, or how best to apply them in different disease contexts. The U.S. Food and Drug Administration (FDA) has released a draft guidance for industry that outlines how a more personalized approach may be applied in the area of acute treatment for migraines. The draft guidance notes that because migraine is characterized by a complex array of symptoms that may be unique to each patient, a demonstrated effect on headache pain alone is not sufficient for approval. Instead, one should also examine associated symptoms of migraine that are most bothersome to the individual patients, such as nausea, photophobia, and phonophobia. Though the release of this draft guidance indicates the growing interest in such approaches across other therapeutic areas, there are still a number of unanswered questions about their use in medical product development.

There are several existing methods for assessing patients’ symptoms and functions in clinical trials that may represent a personalized COA approach. These include the ‘most bothersome symptom(s)’ (MBS) approach (such as what is described in the FDA draft guidance on migraine), goal attainment scaling (GAS), and computerized adaptive testing (CAT). Each of these approaches warrants further discussion, as there are outstanding issues that need to be addressed to ensure that they are fit-for-purpose (i.e.,
appropriate for the regulatory purpose and context of use for which they are intended). These include the uncertainty over how to adequately reconcile the personalization of COAs so that the results could be pooled and standardized for the purpose of comparisons across treatments and studies. It is also unclear whether personalized COAs can be sufficiently evaluated for use in trials to support approval and labeling, and if so, how to determine when the use of personalized COAs is most appropriate. At base is the question of whether personalized COAs can be treated as a single study endpoint for purposes of analyses.

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

FDA has taken a number of important steps in recent years to facilitate the development and uptake of more patient-centered COAs. 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 directed the FDA to implement 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. More recently, the 21st Century Cures Act and the reauthorization of PDUFA have committed the agency to developing a series of guidelines on how patient experience data can be more systematically incorporated into the drug development and review process. Exploring potential methodological and analytical approaches to personalized COAs will be an important contribution to the agency’s broader efforts to advance the science of patient input more generally, and promote the development of fit-for-purpose COA measures more specifically, and will also help to clarify and prioritize for action the outstanding questions that will need to be addressed through future research.

**Meeting Objectives**

Under a cooperative agreement with the U.S. Food and Drug Administration (FDA), the Duke-Margolis Center for Health Policy convened an expert workshop on April 5, 2017 entitled “Developing Personalized Clinical Outcome Assessments” to advance the discussion on personalizing COAs in medical product development. The objectives of the workshop were to: 1) explore and discuss methodologies and best practices surrounding personalized COAs, and 2) identify specific recommendations on methodologies to explore and advance the use of personalized COAs in medical product development. The day’s discussion was broken out into three sessions, each one dedicated to a potential approach for personalizing COAs: MBS, GAS, and CAT.

This workshop provided an opportunity for representatives from across academia, industry, government, and patient advocacy to engage and discuss the challenges associated with personalizing COAs, with discussion encompassing considerations for small, heterogeneous, and rare disease study populations, as well as how approaches may differ across the four types of COA: patient-reported, observer-reported, clinician-reported, and performance outcome assessments.

**Exploring the Applicability of the ‘Most Bothersome Symptoms’ Approach for Personalizing COAs**

The ‘most bothersome symptoms’ approach is a flexible method for evaluating diseases with variable manifestations, and—as its name implies—is designed to measure only the symptoms that are bothering each individual patient the most. The MBS approach allows patients themselves to select the
outcomes (e.g., signs, symptoms or functional impairment) that are most bothersome to them, in recognition of the fact that patients may be differentially impacted by outcomes specific to the disease or condition. It may be particularly useful in disease areas where patients do not experience all symptoms of the disease, or may manifest different symptoms of the disease (as is often the case for rare disorders, genetic disorders, or other disorders with heterogeneous presentations). The rationales and principles of this approach are also applicable for assessing other concepts, such as the most-impacted function, the worst symptoms, or the most-meaningful outcomes.

Participants generally agreed that MBS may be most feasible for a disease where there is an array of symptoms that might impact a patient, but it is not required for all symptoms to improve. However, for a disease that has variable symptoms and no overlap, the MBS approach to personalized COAs might be more challenging to develop and evaluate.

Statistical Challenges for the Use of a ‘Most Bothersome Symptoms’ Approach

A major focus of the discussion was how to address intra-patient variability in symptom progression or severity over time. The most relevant or bothersome symptoms for a given patient may change over time, making it difficult to accurately measure and evaluate those symptoms that matter most to patients. However, the point was made that if a researcher begins a trial examining a specific symptom which then improves over the course of the study, then this may actually mean that the treatment is working for the originally specified ‘most bothersome symptom’ and the trial is successful, regardless of whether another symptom becomes more bothersome. Some participants still expressed concerns over the issue of intra-patient variability over time, and noted that it would be difficult to assess treatment benefit if patients saw an improvement in the ‘most bothersome symptom’, but also experienced a worsening of other symptoms over the course of the trial. One compromise solution is to assess whatever are the most bothersome symptoms at the time of assessment and treat all most bothersome symptoms equally. The treatment is said to be effective as long as the severity of the most bothersome symptoms (regardless of which) is meaningfully reduced from early time points. By only measuring the most bothersome symptoms, it is not possible to measure a lack of meaningful worsening in those symptoms that had not been identified as most bothersome. It was suggested that MBS may therefore be more appropriate in cases where symptoms of the disease are unidimensional and tend to progress in the same direction. This will make the MBS approach more feasible, but could limit its application in diseases where the personalized COA is most needed, i.e., in cases where symptoms manifest in ways that are more difficult to predict.

Appropriateness of the Term ‘Bothersome’

In addition to statistical challenges, participants also discussed the appropriateness of the term ‘bothersome’ and whether it was correct for this context. Instead of focusing on what is most bothersome to patients, some suggested focusing on the symptom on which patients score the worst. Others pointed out that patients may accommodate to symptoms over time, and another symptom might worsen or emerge as more bothersome. This again, circles back to the question of intra-patient variability. When identifying the most bothersome symptom, an open question was whether the goal is to assess the symptom that impacts patients the most in their daily lives. The uncertainty over how to define ‘bothersome’ was a consistent theme throughout the discussion, and more work is needed to reach a consensus on its meaning and use in this context.
Exploring the Applicability of Goal Attainment Scaling for Personalizing COAs

Goal Attainment Scaling enables patients with the same disease to set their own treatment goals under the supervision of a medical professional. Although patients may differ in the type and number of goals they set, the attainment of these goals is measured in a standardized way, allowing for a standardized evaluation of an intervention even when patients are in different stages of their disease. It may be particularly useful for small, rare, or heterogeneous patient groups where generic outcome measures may not be appropriate or applicable.

Statistical Challenges for the Use of Goal Attainment Scaling in Clinical Trials

Participants noted a number of statistical challenges with the use of GAS in a clinical trial setting. One particular consideration was related to the weighting of goals in various contexts, as it was unclear whether all goals on a scale should be weighted the same. Though not all goals that might be set by a patient may be of equal value, differential weighting might add a layer of difficulty to the analysis. Participants also questioned whether the goals themselves need to be measured in standardized units. If patient goals are measured in different units, there was some uncertainty over how to weight and combine these goals together to analyze the results.

Another issue related to how to control for situations in which participants in one arm of the trial select goals that are easier to attain than the goals set by participants in the other arm, as such differences could skew the analysis. Participants did note that this issue could be mitigated through randomization of patients to treatment assignment, but some participants were skeptical as to whether this would adequately resolve the issue. One potential solution is to randomize patients to study arms stratified by treatment goals.

Participants also expressed concerns about the difficulty in interpreting the results, and highlighted issues with how to determine the meaningful benefit of a treatment. For example, it was unclear whether meaningful treatment benefit should be defined as “achieving all goals” and, if not, how many of the pre-established goals should be achieved in order to meet that threshold.

Operational and Logistical Challenges for the Use of Goal Attainment Scaling in Clinical Trials

Participants raised several logistical and operational challenges to implementing GAS into clinical trials. In particular, they noted challenges related to implementing this process across study sites so as to ensure consistency, as well as adequately training medical professionals to lead patients through the process. The time-intensive nature of the method would make it challenging to do on a large scale across multiple sites or countries.

There was also uncertainty over how to manage patient expectations and what rules should be applied to the goal-setting process. Some participants questioned whether goals should be limited to only those that could be impacted by the treatment, while others noted that limiting patient goals may mean something truly important to the patient is overlooked (i.e., the assessment is not completely personalized). Constraining goals to a specific subset of these symptoms may not provide a full picture of the patient experience. It is also unclear how best to apply GAS when a key goal might be to slow disease progression instead of improving symptoms or cure the disease.
Finally, participants pointed out that GAS has been typically applied in clinical practice, where the clinician tries to work with the patient to help them achieve that goal. However, in a traditional clinical trial, the aim is often to minimize care variability to isolate treatment effect. This may contradict the goal of GAS, which is to identify personalized approaches to treating each patient.

**Exploring the Applicability of Computerized Adaptive Testing for Personalized COAs**

Under a CAT approach, the computerized item selection algorithm selects questions or “items” from an item bank based on a patient’s responses to previously answered questions. The objective is to increase the score precision and to reduce patient burden by selecting items that are closest to the patient’s level on the construct being assessed (e.g., symptom severity). For example, a patient with mild symptoms only has to answer items associated with mild symptoms without being asked about items associated with severe symptoms, and vice versa. Although patients receive different questions based on their individualized responses, scores are standardized and can be compared using a common scale.

CAT is different from the other two personalized approaches as it is a ‘domain-based’ approach, wherein item banks that contain items assessing the same construct are developed. The item bank enables administering items that are tailored to the severity level of individual patients with respect to the particular domain (or concept) of interest. This has the advantage of increasing efficiency, as it limits questions to those that most appropriately reflect each patient’s experience with a disease, hence decreasing the number of questions needed and reducing the burden on patients.

**Statistical Challenges in Comparability and Interpretation for Computerized Adaptive Testing**

Participants voiced a number of concerns with utilizing CAT to personalize COAs. First, the item selection algorithm is usually based on quantitative rules aimed at increasing measurement precision, which may miss items that are relevant and important to assess. To mitigate the possibility of dropping items with important content, one suggestion is to administer a hybrid CAT with a set of common items that are administered at the end of the session to ensure content coverage. Relatedly, because patients may receive a different set of questions based on their responses to previously answered questions, it can be challenging to compare and interpret the results across and within patients in a trial that uses CAT. While this does have the advantage of reducing item burden compared to static forms, static forms better allow for comparing groups at the level of item content. Some participants raised the possibility of incorporating a static short form into the analysis in order to calibrate the CAT. However, the addition of a short form could negate the benefits of using CAT in the first place.

In addition to challenges for comparing across patients, it may also be difficult to make comparisons for individual patients over time. Depending on the course of the disease or timeline of the trial, there could be scenarios where patients are presented with an entirely different set of items at the end of the trial than they received at baseline. It may be difficult to understand item progression or to interpret the results if it is not possible to compare the change in individual items. Participants felt that addressing these issues would be important if CAT is applied in clinical trials.

**Additional Considerations for Using a Computerized Adaptive Testing Approach**

Participants stressed the importance of having a clearly designed study before employing a CAT approach. Researchers must evaluate the contents of the relevant item bank upfront and ensure it is appropriate for their trial population, as some key symptoms may not be reflected in the item banks and
it may take considerable time to develop new items. Translation of the item bank to other languages and cultures may also prove challenging.

It is also necessary for researchers to decide how they intend to establish a meaningful change for the trial population. Conventional approaches, such as anchor-based approaches can be used, but there is also the possibility of developing model-based clinical vignettes. These determinations, among others, are necessary to identify before applying CAT.

**Applicability of Personalized Approaches to other COAs Aside from PROs**

The personalized approaches described above have been primarily used in the context of PROs, but there was substantial interest among meeting participants regarding their potential application to other types of COAs. Especially in rare disease settings, for example, a personalized COA approach could have utility for PerfOs. Patient with rare disease may have significant functional impairment preventing them from performing a large number of tasks as part of a PerFO. Only completing tasks that are directly relevant to the patient would significantly reduce burden on these patients. A similar approach could also have value for ClinROs, where clinicians may focus only on evaluating those signs, symptoms, or functional concepts that are relevant to each patient.

Given the relative lack of experience in adapting these approaches to other COAs, however, participants noted that more work is needed to understand how best to apply the personalized COA approach to other COAs (including how to adapt the technological platform used to administer the assessments) and interpret the resulting data.

**Major Takeaways and Areas for Additional Work**

Overall, participants agreed there was potential in each of these approaches, and while each could be useful in certain contexts, the framework for that use needs to be well-defined. It is essential to have clearly defined and interpretable endpoints in order to advance the use of these methods. Participants also reinforced throughout the day’s discussion that patient engagement is crucial. Patients need to be included during all phases of a trial – before the trial starts, throughout the trial, and after the trial has ended. Where possible, patients should also be engaged in the pre-competitive space, in order to minimize the burden on patients and patient groups moving forward.

One of the most important areas for future work is addressing the statistical challenges associated with each of these three approaches, particularly how to combine data for analysis. Personalized COAs imply that different items or assessments have been administered, but the goal is to integrate them together to draw meaningful conclusions from the data. Some participants questioned whether simple averaging could be used or whether more specialized techniques are required. There were also open questions regarding the personalizing of outcome measures to include only those that are relevant to individual patients. There was concern that important information could be eliminated by tailoring the data collection too narrowly.

Another area that will require further examination is how to include this information in a label. There may be lessons that can be learned from labels that include composite endpoints, but more work in this area is needed to address the diversity of personalized COAs.
Lastly, some participants raised the possibility of incorporating these personalized approaches into clinical care, so as to support the generation of real world evidence (RWE) on these products in the postmarket setting. This, however, is a longer-term goal, as current efforts to define the potential applications and methods for real world evidence development are in relatively early stages.