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 determines 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 Ensuring that COAs are patient centered – that they identify, measure, and evaluate symptoms or functions that are meaningful to patients – is central to their effectiveness. 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, including many rare diseases. 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 effects. A possible solution to this problem includes employing a “personalized COA” approach, in which the COA may vary across patients in an effort to measure the most important and relevant signs, symptoms, functions, as well as the degree of severity of these concepts in each individual. While this approach shows promise, there are a number of outstanding questions relating to the development, implementation, and evaluation of personalized COAs.

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 personalized COAs, and 2) identify specific recommendations on methodologies to explore and advance the use of personalized COAs 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>
<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</td>
</tr>
</tbody>
</table>
amendment or interpretation of the patient’s response by a clinician or anyone else.

| **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 or population, or is ‘generic’ (i.e. applies across diseases or populations).

**The Promise of Personalized COAs**

The strength of a COA is dependent on a number of factors, including its ability to detect change and whether that change can be interpreted as clinically meaningful to patients. However, the heterogeneous course of many diseases has proven a substantial challenge for the effective use of COAs. For example, in multiple sclerosis, the range and severity of symptoms associated with the disease are different for each patient, and may vary for each patient over time. While a number of COA tools have been developed and refined in the last few decades, the development of tools that are sensitive, precise, and measure multiple dimensions of MS that are important to patients has proven difficult.

With these challenges in mind, stakeholders have become increasingly interested in the use of more personalized approaches for outcome measures which are individualized to evaluate symptoms and functions that matter most to patients.

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. 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. 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.

Another key challenge is how best to formally define the concept of “personalized COAs.” Although specific methodological approaches are often used to illustrate what they are and how they might be used, these methodological approaches may engage patients in different ways to achieve
personalization. The lack of an inclusive definition creates confusion that can hamper communication and progress in this emerging field.

**Personalized COAs in Medical Product Development: Challenges and Opportunities**

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 symptoms’ approach (such as what is described in the FDA draft guidance on migraine), Goal Attainment Scaling, and Computer Adaptive Testing (CAT). (Note: this list is not exhaustive and includes only some examples of methods to “personalize” COAs; other existing or novel methods should be considered in the broader context.)

**Most Bothersome Symptoms**

This approach allows patients themselves to select the outcomes (e.g., signs, symptoms or functional impairment) that matter most to them, in recognition of the fact that patients may be differentially impacted by outcomes specific to the disease or condition. In its 2014 draft guidance “Migraine: Developing Drugs for Acute Treatment”, FDA proposed the use of such an approach to demonstrate efficacy in a clinical trial. 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. The approach described in the draft guidance recommends two co-primary endpoints: 1) having no headache pain at 2 hours after dosing, and 2) a demonstrated effect on the migraine-associated symptom designated by each individual patient at 2 hours after dosing.

The potential for this approach to be modified for implementation in other disease areas is also a subject of substantial interest, though there are ongoing questions over how best to do this in a rigorous and methodologically sound way. (See Attachment I.)

**Goal Attainment Scaling**

Goal Attainment Scaling was developed in 1968 for the evaluation of mental health services, and has been widely used in fields such as geriatrics and rehabilitation medicine (e.g. the treatment of spasticity associated with cerebral palsy or stroke) to evaluate the effect of an intervention on an individual basis. The method 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. While Goal Attainment Scaling has been used in a number of settings and in different disease areas, it has not been consistently evaluated for use in clinical trials, and more work is needed to verify its applicability in medical product development. (See Attachment II.)

**Computer Adaptive Testing**

While CAT has been used most notably in educational testing, the approach has more recently been applied to health outcomes, such as the Patient Reported Outcomes Measurement Information System (PROMIS™) measures and the European Organisation for Research and Treatment of Cancer CAT (EORTC CAT).

In CAT, the computer administering the test selects questions or “items” from an item bank based on a patient’s response to previously answered questions. This is achieved through the application of Item Response Theory (IRT), a psychometric method that produces scores associated with the patient’s
answers to questions. Although patients receive different questions based on their individualized responses, scores are standardized and can be compared using a common scale. The goal of CAT is to improve measurement precision for each individual for the specific domain of interest (e.g. physical functioning, depression) being measured. (For additional background information on this approach, please see Attachment III.)

Challenges in personalizing COAs
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. This includes questions over whether personalized COAs, however defined, can be sufficiently evaluated for use in trials to support approval and labeling, and if so, how to determine when they are most appropriate. It is unclear how to adequately reconcile the personalization of COAs with the broader need for standardization in the clinical trial setting. There are also statistical challenges associated with the use of these methods. It is not always clear, for example, how best to analyze the data derived from individualized measures, where a major issue is whether the data can be pooled and if so, how best to accomplish this. For Goal Attainment Scaling and the most bothersome symptoms approach, it can be challenging 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. For Computer Adaptive Testing, it is not always evident how best to ensure that endpoints derived from CATs are equivalent across and within patients in a trial so that scores and interpretation are compatible across patients. While there is great potential for personalized COAs to measure outcomes that are the most important and meaningful to patients, these challenges underscore the need for more clarification and consensus in the field.

Regulatory 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 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. Advancing the discussion on potential methodological and analytical approaches to personalized COAs will be an important contribution to this ongoing process, and will help to clarify and prioritize for action the outstanding questions that will need to be addressed through future research.

Meeting Objectives
The purpose of this workshop is explore the challenges associated with the development, implementation, and evaluation of personalized COA approaches, identifying best practices, and developing specific recommendations to advance the use of personalized COAs in medical product development. Discussion will encompass special considerations for small and heterogeneous study
populations, as well as how approaches may differ across the four types of COA (patient-reported outcomes, observer-reported outcomes, clinician-reported outcomes, and performance outcomes).

**Session I: Evaluating the Most Bothersome Symptoms (MBS) Approach for Developing Personalized COAs**

**Objective:** This session will consider one example of an existing personalized COA approach - the Most Bothersome Symptoms Approach - highlight statistical and methodological gaps in this approach, and identify strategies for determining when this approach is the most appropriate for a given context 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 do we establish baseline scores?
- How do we determine meaningful change under this approach?
- How do we ensure that the scores and interpretation are compatible across patients? Between study arms?
- How do we best analyze the data using an MBS approach? What are some of the best practices when analyzing such data?
- What are some of the best practices other than the MBS approach when analyzing such data?
- How do we handle heterogeneity within the patient over time?
  - When we focus on “most bothersome symptoms”, how do we ensure other symptoms do not worsen?
  - How do we handle the situation when the symptoms naturally relapse and remit so the “most bothersome symptom” at baseline has changed?
- What are the advantages and disadvantages of the most bothersome symptoms approach for use in drug development?

**Session II: Evaluating Goal Attainment Scaling for Developing Personalized COAs**

**Objective:** This session will consider one example of an existing Personalized COA approach - Goal Attainment Scaling - highlight statistical and methodological gaps in this approach, and identify strategies for determining when this approach is the most appropriate for a given context 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 do we establish baseline scores?
- How do we determine meaningful change under this approach?
- How do we ensure that the scores and interpretation are compatible across patients? Between study arms?
- How best to analyze the data? What are some of the best practices when analyzing such data?
• How do we handle heterogeneity within the patient over time? For example, how do we handle the situation when the symptoms naturally relapse and remit so the “goal” at baseline has changed?
• What are the advantages and disadvantages of the goal attainment approach for use in drug development?

**Session III: Evaluating Computer Adaptive Testing for Developing Personalized COAs: A Domain Based Approach**
**Objective:** There is significant interest in developing item banks that can administer items tailored to the level of the underlying construct of individual patients, often referred to as the domain approach. This includes approaches such as computerized adaptive testing (CAT) and hybrid CATs. However, there are ongoing questions over how best to ensure that endpoints derived from the CATs are equivalent across and within patients in a trial. There also are ongoing questions over how best to ensure the results of a study sample could be generalized to the patient population. This session will explore those challenges and identify specific recommendations for addressing them.

**Questions to address:**
• When is this set of approaches useful in clinical trial settings (e.g., special populations such as pediatric, rare disease, etc.)?
  o How do we operationalize this set of approaches?
  o How feasible is this set of approaches?
  o How do we establish baseline scores?
• How do we determine meaningful change under this set of approaches?
• How do we ensure that the scores and interpretation are compatible across patients (i.e., the same concept is being measured at all-time points and in all patients)? Between study arms?
• How best to analyze the data? What are some of the best practices when analyzing such data?
• What are the advantages and disadvantages of CAT domain-based approaches for use in drug development?

**Session IV: Reflecting on the Methodological Examples for Developing Personalized COAs**
**Questions to address:**
• How can we best determine *when* a personalized COA approach is appropriate and feasible for use in drug development (e.g., special populations such as pediatric, rare disease, etc.)?
• How can we best determine *which* personalized COA approaches are the most appropriate in a given context of use?
• What are additional advantages and disadvantages that have not been discussed for any of these methods for use in drug development?
• What are the key analytical and methodological implications of personalized COA approaches?
• How would the personalized COA approaches discussed today be applied to other types of COA for use in drug development?
• What are some of the additional emerging personalized COA approaches that warrant further exploration for use in drug development?
• What key questions remain regarding the validity/rigor/appropriate applications of these methods?

Session V: Identifying Potential Paths Forward
Objective: Reflect on the day’s discussion, identify the areas of consensus and potential paths forward, highlight major gaps in knowledge and prioritize them for future research/discussion
3 Ibid.
6 Ibid.
7 Ibid.
10 Ibid.
11 Ibid.
14 Ibid.