Alliance for Regenerative Medicine

June 30, 2023

Dockets Management Staff (HFA-305)
Food and Drug Administration
5630 Fishers Lane, Rm. 1061
Rockville, MD 20852

Re: Docket No. FDA-2023-D-0026 for *Patient-Focused Drug Development: Incorporating Clinical Outcome Assessments Into Endpoints for Regulatory Decision-Making*

Dear Sir/Madam:

The Alliance for Regenerative Medicine (ARM) is pleased to submit comments to the US Food and Drug Administration (FDA) in response to the recently released guidance titled, *Patient-Focused Drug Development: Incorporating Clinical Outcome Assessments Into Endpoints for Regulatory Decision-Making*.

The Alliance for Regenerative Medicine (ARM) is the leading international advocacy organization championing the benefits of engineered cell therapies and genetic medicines for patients, healthcare systems, and society. As a community, ARM builds the future of medicine by convening the sector, facilitating influential exchanges on policies and practices, and advancing the narrative with data and analysis.

We actively engage key stakeholders to enable the development of advanced therapies and to modernize healthcare systems so that patients benefit from durable, potentially curative treatments. As the global voice of the sector, we represent more than 475 members across 25 countries, including emerging and established biotechnology companies, academic and medical research institutions, and patient organizations.

General Comments

ARM appreciates that within this fourth guidance document in the Patient-Focused Drug Development (PFDD) series, FDA clarifies a question that ARM posed in response to the third guidance document in the series—*Selecting, Developing, or Modifying Fit-for-Purpose Clinical Outcome Assessments.* ARM asked whether the third guidance document in the series would

1015 18th Street NW, Suite 1102
Washington, DC 20036
www.alliancerm.org
info@alliancerm.org
@alliancerm
@Alliance for Regenerative Medicine

replace or expand upon the 2009 guidance, *Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims.* In this fourth guidance, FDA states, "When final, the PFDD guidance series will replace the 2009 guidance document."

Because the content of the four guidance documents in the series complements the others, frequent referencing of previous PFDD guidance documents in the series within this fourth draft guidance is critical to foster the most comprehensive understanding among stakeholders in this space. We recommend additional referencing of the many existing FDA guidance documents that are relevant to the concepts within the current one. ARM particularly suggests referencing the third guidance document in the series on the topics of Item Response Theory and computerized adaptive testing where applicable within this guidance document.

We value that FDA created this highly detailed procedural guidance document on this topic since cell and gene therapy (CGT) development often utilizes clinical outcome assessment (COA)-based endpoints, and much of the information within the document provides quite beneficial guidance for sponsors in developing such endpoints. We specifically appreciate identification of situations in which multiple primary endpoints would be useful, as well as the provision of the following recommendations, which may assist sponsors in avoiding the challenges related to these topics:

- Ideally to statistically adjust for patients' baseline scores on a COA when evaluating a treatment benefit at fixed and predetermined time points; and
- If the value of the meaningful score difference (MSD) is not the same regardless of the baseline COA score, to use different values for MSD depending on the patient's baseline status.

ARM recommends FDA acknowledgement within the guidance document of the challenge of validating patient-reported outcomes (PROs) for some rare diseases and flexibility on the need for validation in such cases. We provide suggested language on this topic below.

FDA should further incorporate recommendations on statistical methods into the final guidance document. For example, the draft guidance document notes several approaches to constructing personalized endpoints but provides limited detail on appropriate statistical modeling for each approach. Thus, FDA could identify considerations for the selection of statistical methods to best support the achievement of research objectives. Another example of the type of content that should be addressed is that if a common score is based on different constituent symptom ratings, establishment of exchangeability would be necessary.

In Section IV, Other Considerations, which contains information on formatting and submitting patient experience data, we recommend that FDA also indicate how patient preference data will be used in benefit-risk evaluation during review of marketing applications. We encourage referencing and inclusion of content from the draft guidance document, *Benefit-Risk*

Assessment for New Drug and Biological Products. That guidance document indicates that if a methodologically sound and fit-for-purpose data collection tool is used to collect patient experience data in a drug development program, the collected data can provide direct evidence regarding the benefits and risks of the drug and their importance to patients. It also states that patient preference data can inform potential benefits that are most meaningful; acceptability of risk and uncertainty; and value and burden of risk mitigation efforts.

Also in this section, FDA should discuss how to measure the correlation between COAs and biomarkers on both a short-term and long-term basis and the usefulness of doing so in providing support of biomarker use. ARM also suggests that FDA should describe in this section the timing, format, and content for the submission of COA data within a marketing application and through a COA dossier. It would be helpful for the Agency to clarify when and whether a COA dossier is recommended for primary and secondary COA-based endpoints, as well as the timing of, and recommendations for content of, a COA-specific statistical analysis plan.

ARM appreciates the FDA for its consideration of these comments and the Agency's overall effort to provide guidance that will assist sponsors in the field of regenerative medicine. Below is a listing of line-by-line comments on this proposed guidance.

Sincerely,

Michael Lehmicke
Vice President, Science and Industry Affairs

Specific Line-by Line Comments: Section/Line	**Guidance Text**	**Rationale for Change or Comment**	**Proposed Change**
I. Introduction			
A. Overview of the Series of FDA Guidance Documents on Patient-Focused Drug Development			
Lines 49 – 51	"FDA encourages stakeholders to interact early with FDA and obtain feedback from the relevant FDA review division when considering the collection of patient experience data related to the burden of disease and the benefits, burdens, and harms of treatment."	Comment: ARM recommends explicitly identifying how sponsors should best engage with FDA on this topic, e.g., whether a Type D meeting would be appropriate. In addition, FDA should recommend how to utilize each meeting type for COA	

		discussions. Such information will assist sponsors in planning for meeting requests, which is important for efficient development due to challenges in obtaining meetings quickly.	
B. Purpose and Scope of PFDD Guidance 4			
Lines 62 – 64	"This guidance builds on Guidance 3 by focusing on endpoints constructed from fit-for-purpose[11, 12] COAs which are intended to reflect, directly or indirectly, how patients feel, function, or survive."	Comment: This statement refers to footnote 12, which states, "A COA is considered fit-for-purpose when the level of validation is sufficient to support its context of use. ..." ARM agrees with prior FDA acknowledgment of the difficulty of obtaining sufficient numbers of patients to validate PROs for some very rare diseases. We recommend moving the definition of fit-for-purpose COAs to the guidance text, followed by a statement regarding the flexibility that can be applied in these cases, as stated to the right.	"This guidance builds on Guidance 3 by focusing on endpoints constructed from fit-for-purpose[11, 12] COAs which are intended to reflect, directly or indirectly, how patients feel, function, or survive. A COA is considered fit-for-purpose when the level of validation is sufficient to support its context of use. For some rare diseases, validating PROs may not be feasible. In such instances, it may be acceptable to incorporate PROs that have not been validated into endpoints."
II. COA-Based Endpoint Considerations			
A. Endpoint of Interest: What Are You Measuring in the Target Study Population?			
2. Considerations for Constructing a COA-Based Endpoint			
b. Endpoints based on COA scores at a fixed time point or a summary of COA scores over time			
Lines 218-222	"In most situations in which a COA produces ordinal or continuous (interval or ratio scale) scores, the best and recommended endpoint will be the COA score at a predefined assessment point or summarized over some predefined post-baseline assessment period, and the most straightforward analysis will be a comparison of	Comment: While the comparison of randomized groups may be the most straightforward analysis, randomized trial design may be challenging or not possible for diseases with small or very small populations.	"In most situations in which a COA produces ordinal or continuous (interval or ratio scale) scores, the best and recommended endpoint will be the COA score at a predefined assessment point or summarized over some predefined post-baseline assessment period.~~, and t~~The most straightforward analysis, when feasible, will

	randomized groups with respect to the follow-up score(s) after adjusting for the baseline value (e.g., with a linear model to compare average follow-up scores)."		be a comparison of randomized groups with respect to the follow-up score(s) after adjusting for the baseline value (e.g., with a linear model to compare average follow-up scores. For products that treat rare diseases, randomized trial design may not be feasible."
d. Endpoints constructed by computing change from baseline or percent change from baseline COA scores			
Lines 275 – 281	"COA scores that are ordinal are challenging to interpret in terms of change from baseline because the difference between two ordinal scores cannot be assumed to have the same meaning across scores (e.g., for an ordinal score with 5 levels—when interpreting level 3 relative to level 1 and level 5 relative to level 3—both differ by two levels but might not correspond to the same degree of change in the underlying health state). Put another way, there might not be a linear relationship between the ordinal values and the true level of symptom severity or functioning being measured."	Comment: ARM agrees with this challenge. In addition, it may be beneficial to note that ordinal scales can also be challenging to use for slowly progressing diseases in which scores change minimally over the typical time period of a clinical trial. Guidance on endpoint selection in such cases would be helpful to sponsors.	"COA scores that are ordinal are challenging to interpret in terms of change from baseline because the difference between two ordinal scores cannot be assumed to have the same meaning across scores (e.g., for an ordinal score with 5 levels—when interpreting level 3 relative to level 1 and level 5 relative to level 3—both differ by two levels but might not correspond to the same degree of change in the underlying health state). Put another way, there might not be a linear relationship between the ordinal values and the true level of symptom severity or functioning being measured. Ordinal scales can also be challenging to use for slowly progressing diseases in which scores change minimally over the typical period of a clinical trial. In such cases, sponsors should consider the sensitivity of the COA score to small changes over a short period to select an endpoint that may best be able to demonstrate the benefit of treatment."

Line 289 – 291	"For situations in which it is not possible to conduct a randomized, controlled trial and a single arm trial is done instead (e.g., to evaluate some devices), a change-from-baseline endpoint might be the best available option."	Comment: ARM appreciates this acknowledgment that situations exist in which it is not possible to conduct a randomized, controlled trial. We recommend adding some rare diseases as another example.	"For situations in which it is not possible to conduct a randomized, controlled trial and a single arm trial is done instead (e.g., to evaluate some devices and for some rare diseases), a change-from-baseline endpoint might be the best available option."
e. Endpoint strategies when a disease affects multiple aspects of feeling and functioning			
Line 375	"Construct a Multi-Component Endpoint"	Comment: In this section, FDA should address use of existing rating scales that are commonly used, including characteristics of optimal rating scales and how to use individual components to address specific outcomes (and/or reference resources on these topics).	
Lines 458 - 461	• "Endpoint values are strongly dependent on the thresholds selected for meaningful improvement and/or worsening and choosing such thresholds can be challenging. Thresholds for each COA should be predefined and justified. Sponsors should also conduct sensitivity analyses that explore treatment effects over a range of thresholds."	Comment: The value of the endpoints are not dependent on the threshold range selected to represent meaningful improvement, but whether endpoint values are significant (fall within the threshold range) strongly depends on the threshold range selected. We think this is what is meant and suggest clarifying as stated to the right.	• "Whether ~~E~~endpoint values are significant (fall within the threshold range) ~~are~~ strongly dependsent ~~s~~ on the threshold~~s~~ range selected ~~for~~ as representing meaningful improvement and/or worsening.~~, and c~~ Choosing such thresholds can be challenging. Thresholds for each COA should be predefined and justified. Sponsors should also conduct sensitivity analyses that explore treatment effects over a range of thresholds."
Lines 463 – 468	• "There is the potential for bias when those completing or administering the COA are aware of the thresholds for being considered a meaningful improvement (or worsening). It is	Comment: It may not be possible for clinicians or researchers to be unaware of threshold definitions of change, such as when using a commonly used rating scale. For trials that utilize retrospective external	• "There is the potential for bias when those completing or administering the COA are aware of the thresholds for being considered a meaningful improvement (or worsening). It is

	important when possible that clinicians (for ClinRO measures), caregivers (for ObsRO measures), and/or any research staff (for PerfO measures) involved in assessment are not made aware of the threshold definitions and are masked to treatment assignment."	control, as many trials for rare diseases do, masking is not relevant.	important when possible that clinicians (for ClinRO measures), caregivers (for ObsRO measures), and/or any research staff (for PerfO measures) involved in assessment are not made aware of the threshold definitions, although this is not always possible. ~~and are m~~Masking~~ed~~ to treatment assignment may be helpful when relevant."
III. Evaluating the Meaningfulness of Treatment Benefit			
A. Factors Affecting the Interpretability of COA Scores			
Lines 704 - 706	"For example, if a treatment is shown to reduce scores on a performance outcome measure by an average of 2 points on a 15-point scale, it would be helpful to know whether a 2-point difference corresponds to something that patients would notice as important in their daily lives."	Comment: ARM agrees with this statement and suggests strengthening the wording of the recommendation to emphasize the importance of patient experience information.	"For example, if a treatment is shown to reduce scores on a performance outcome measure by an average of 2 points on a 15-point scale, it would be ~~helpful~~ important to know whether a 2-point difference corresponds to something that patients would notice as important in their daily lives."
1. How Closely Does the Measured Concept of Interest Correspond to the Patients' Experiences?			
Lines 725 – 732	"Other COAs might measure a concept of interest that is more indirectly related to the patient's health-related experiences, such as an ObsRO measure of the patient's pain behavior (which is indirectly related to the patient's actual pain) or a PerfO measure of leg strength (which is indirectly related to activities that require lower limb function such as walking or climbing stairs). For these types of measures, it may be more challenging to infer how different scores on the measure correspond to	Comment: Obtaining patient input on which functional activities are most important to them and surveying patients, similar to the approach mentioned in lines 1033 – 1035 of the guidance, on their ability to complete these tasks may assist in determining what level of strength, for example, correlates with the functional activities of interest.	"Other COAs might measure a concept of interest that is more indirectly related to the patient's health-related experiences, such as an ObsRO measure of the patient's pain behavior (which is indirectly related to the patient's actual pain) or a PerfO measure of leg strength (which is indirectly related to activities that require lower limb function such as walking or climbing stairs). For these types of measures, it may be more challenging to infer how different scores on the measure correspond to

	different experiences the patients might have; this means that additional empirical support is needed to translate scores on the measures to corresponding patient experiences in their daily lives."		different experiences the patients might have; this means that additional ~~empirical~~ support may be ~~is~~ needed to translate scores on the measures to corresponding patient experiences in their daily lives."
B. Approaches for Collecting Evidence to Support Interpretability of COA-Based Endpoints			
1. *Interpreting in Terms of Meaningful Score Differences*			
a. Choice of anchor variables			
Lines 831 – 835	"Sometimes it may not be possible to find an anchor that is a direct reflection of the patients' experiences related to the concept of interest measured by the COA-based endpoint. In such cases, sponsors can consider using multiple, less directly related anchors to aid in the interpretation of a meaningful difference in scores."	Comment: ARM recommends FDA clarify whether multiple anchors refers to multi-item anchors (e.g., a Patient Global Impression of Change scale with multiple items) and/or other types of anchors.	
IV. Additional Considerations			
A. Other Study Design Considerations			
4. Considerations When Using a Nonrandomized Design, External Controls, or Nonconcurrent Control			
Lines 1362 – 1364	"Whenever possible, COA-based endpoints should be assessed in the context of randomized, controlled clinical trial designs. Sponsors considering COA-based endpoints in nonrandomized, external control, or nonconcurrent control (randomized groups but at different calendar times) trial designs should be aware of the significant potential for bias in estimating treatment effects:"	Comment: The 2020 final guidance document, *Human Gene Therapy for Rare Diseases*, indicates, "For rare diseases, there may be a limited number of patients who may qualify for enrollment into a clinical study." We recommend similar acknowledgement of this challenge to the use of randomized, controlled trials for rare diseases.	"Whenever possible, COA-based endpoints should be assessed in the context of randomized, controlled clinical trial designs. Consideration of the use of alternate trial designs, such as the use of external controls, may be appropriate in clinical trials for rare diseases, which may have a limited number of patients who qualify for enrollment into a clinical study. Historical information can potentially serve as a control group in certain situations, such as when the disease is known not to

			improve in the absence of an intervention or with available therapies. Sponsors considering COA-based endpoints in nonrandomized, external control, or nonconcurrent control (randomized groups but at different calendar times) trial designs should be aware of the significant potential for bias in estimating treatment effects,~~:~~ outlined below. Such bias may be able to be addressed with the use of appropriate analytic methods."
7. *Minimizing Participant Burden*			
Lines 1463 – 1465	"With respect to COA-based endpoints, patient communities can provide input on the relevance, type, length, and frequency of COAs."	Comment: The frequency of patient assessment and number of trials post-market may be more than patients and caregivers are willing to engage in. We suggest allowing patient community input to inform the number and frequency of post-market confirmatory trials that are reasonable to expect patients to participate in.	"With respect to COA-based endpoints, patient communities can provide input on the relevance, type, length, and frequency of COAs, including in post-market confirmatory trials."