

July 3, 2023

Dockets Management
Food and Drug Administration
5630 Fishers Lane, Rm. 1061
Rockville, MD 20852

Re: Docket No. [FDA-2023-D-0026](#)

Draft Guidance

Dear Sir or Madam,

This letter is in reference to FDA's draft guidance: **Patient-Focused Drug Development: Incorporating Clinical Outcome Assessments Into Endpoints for Regulatory Decision-Making** (PFDD Draft Guidance 4).

In response to the FDA's Open Comment Period, these comments were drafted by members of the International Society for Quality of Life Research (ISOQOL) Special Interest Group for Regulatory and Health Technology Assessment Engagement (R&HE SIG). The comments shared in this document are a collection of comments contributed voluntarily by R&HE SIG members. The comments below do not necessarily reflect the views of ISOQOL or ISOQOL Leadership or all ISOQOL R&HE SIG members, nor do they necessarily reflect the views of the SIG members' respective organizations, employers, or affiliations.

The purpose of the R&HE SIG is to routinely and actively engage with regulatory agencies and health technology assessment (HTA) bodies to address critical applied measurement issues raised by these entities or that emerge from their public communications (e.g., guidance documents, reflection papers, workshops). This collaboration embraces a broad measurement perspective and enables us to build scientifically-based patient-focused recommendations regarding the use of clinical outcome assessments (COAs) in the evaluation of therapeutic interventions aimed at advancing human health and well-being.

ISOQOL is a global community of researchers, clinicians, health care professionals, industry professionals, consultants, and patient research partners advancing health related quality of life research (HRQL).

Together, we are creating a future in which patient perspective is integral to health research, care and policy.
www.isoqol.org

PFDD Draft Guidance 4 represents another positive step forward in FDA's development of the 4-guidance series on PFDD. The ISOQOL R&HE SIG appreciates this opportunity to engage with the FDA by providing these comments to the docket for consideration.

General comments:

- The continued focus in this Draft Guidance, as in Guidances 1-3, on the need to provide rationales for decision-making and to ensure patient experience data are planned and collected to answer research questions (i.e., collecting patient experience data should not be a checkbox exercise) is much appreciated.
- The discussion on interpreting meaningful benefit (Section III) is very methodologically complicated. While it is appreciated that the FDA is focused on what represents meaningful benefit to patients (beyond statistical significance), the methods to understand this benefit are becoming increasingly complex, quite disparate across review teams at the FDA, and are confusing to many. This Draft Guidance introduces many types of methods, some new and uncommon to drug development. While surely FDA is attempting to exhibit flexibility in the methods in this guidance, there seems to be less flexibility (and many divergent requests) across review teams in practice. There are concerns that in the effort to increase flexibility, there is less direction on what is critical for the FDA to base decisions on, and therefore more risk to sponsors. It may not yet be possible in the Final Guidance to better harmonize and take a more standardized approach to evaluating meaningful benefit, but we encourage the Agency to focus on meaningful benefit as a critically important area that needs to be better defined, with more standardized methods, and internal alignment across review divisions, so that sponsors have a better idea in advance of what data are critical to collect vs what is nice to have. Public meetings to discuss methods for meaningful benefit interpretation are crucial. And, somehow, simplification of evaluating meaningful benefit must be the goal and outcome of these meetings.
- The Draft Guidance introduces some new terms and omits others that had become commonly used. For example, MSD and MSR are newly introduced in this Draft Guidance. FDA public comments have indicated that the attempt is to be more descriptive in terms so that they describe better what is meant. With that goal in mind, it seems that maintaining (and including in the Final Guidance) the term “meaningful within person change” (MWPC), which had become commonly used across instrument developers and sponsors, would be useful as it conveys specifically what is being evaluated when a change endpoint is being considered as meaningful. Further, the MSD term is unclear whether it is describing meaningful within person change, “meaningful” group level change (though it is debatable whether group level changes can be appropriately described as “meaningful” to patients), or differences across groups at a single timepoint. If it is FDA’s intention to allow MSD to include multiple meanings, it could be useful to further define those meanings with their own abbreviations to avoid the creation of unique abbreviations across the field.
- New to this Draft Guidance (at least from a CDER perspective) is a focus that appears to be moving away from changes from baseline and moving toward group level differences at a specified



timepoint. While statistically, when controlling for baseline scores, mean changes compared across groups and mean group differences at timepoint t will provide the same result, the endpoint that is used in the trial is what must ultimately be communicated in labeling. It would be valuable to learn from patients whether they would prefer labeling and description of treatment benefit that provides for communication of group level changes from baseline or group level differences at a specified timepoint. As some members (who are patients themselves) have commented, understanding change over time may be more patient-relevant than reporting on group differences at one timepoint to better understand the patient journey while on treatment over time. While various endpoint constructions such as group differences at one timepoint might be appropriate in certain circumstances, it seems appropriate for the FDA to continue encouraging in the final guidance supportive analyses that look at proportion of patients who have various levels of potentially meaningful change (e.g., looking at responders across a range of possibly meaningful thresholds) across programs regardless of whether key endpoints are mean change from baseline or group level differences. An emphasis on CDF curves (proportion of responders by change scores) by treatment arm seems appropriate and relevant to recommend as one standardized way that all sponsors could at a minimum evaluate meaningful within person change and communicate treatment benefit that is relevant to patients. Lines 1159-1162 attempt to address CDF curves but could be made stronger to further encourage more use of these rather than stating they “can be plotted.” Further, efforts by the FDA and sponsors to better understand what claims of benefit (e.g., change from baseline vs group level change at a timepoint) are more meaningful to patients would be informative and if available, would provide useful information if added to the PFDD Final Guidance 4.

Specific comments by line number:

- Line 123-147: this is a very nice bulleted list of components of an endpoint. Often the measure and the endpoint are conflated in protocols and/or endpoint planning discussions (e.g., “the endpoint is the EORTC-QLQ C30”), which can cause challenges if the details of the endpoint are not considered early and thoughtfully. This section does a nice job of providing the elements of an endpoint to consider when making endpoint decisions. Please retain this text in the PFDD Final Guidance 4.
- Line 204-209: thank you for the helpful advice about separating screening from baseline assessments. Please retain this in the Final Guidance. It may be helpful to add a further comment on whether the “later pre-randomization assessment” value must also meet the screening criteria or not. In some cases, sponsors have had screening and baseline values separated, but have implemented the same inclusion/exclusion criteria on the baseline value as well (i.e., exclude patients who successfully screened in but for whom the baseline value did not meet the screening criteria). While this approach may defeat the purpose of separating the screening and baseline values, some sponsors have worried that screening and baseline values are required by FDA to both meet the I/E criteria.

- Line 259-260: It is not clear why data used to derive a meaningful score threshold should be derived from data different from the registrational trial. In cases where a single value threshold is necessary for prespecifying an endpoint (e.g., a responder analysis or time to event analysis) in a registrational trial, the importance of using non-registrational trial data for the threshold derivation is clearer. However, for evaluating meaningful change in patients in the trial as supportive analyses (e.g., secondary endpoint is mean change from baseline and exploratory analyses are intended to evaluate treatment effects over a range of meaningful change thresholds) the need to use non-registrational trial data is unclear. In fact, the registrational trial data may be very appropriate to use in order to examine whether the treatment effects seen in those patients in that trial are meaningful to those same patients. It may be ideal if the range of thresholds might be evaluated across a range of data sources (e.g., phase 2 trial data, phase 3 trial data, exit interviews, etc), however, there are many occasions where the only data available for examining meaningful treatment benefit is the registrational trial. It would be useful if the Agency could address in this section the situation of when registrational trial data might be appropriate to use to examine meaningful patient benefit vs when it may not be.
 - In addition, some sponsors report receiving inconsistent advice from the Agency on whether a range of meaningful change values is appropriate versus a single value. Providing more clarity in the Final Guidance around when a range vs when a single value is appropriate would be appreciated.
- Line 316-317: some have expressed confusion about how to identify when a treatment effect is expected to be multiplicative rather than additive. Providing examples of specific evidence that may be useful for Agency review when a sponsor is considering a percent change from baseline endpoint would be appreciated, particularly given that common CDER advice in recent history has been to avoid percent change from baseline endpoints. Describing example situations when CDER (or other Centers) might agree with a percent change endpoint would be helpful.
- Line 391-392: as written the text may be misunderstood that an overall symptom index score should be weighted in a way that provides different weighting for each item. While a simple sum or average implicitly has a weighted combination of scores (i.e., all items equal weight), not all readers of the guidance appreciate that and there has been some confusion on this point. It may be useful to update in the Final Guidance to read something similar to: “An overall symptom index score created by a well-justified weighted combination of responses to separate items that each assess a different type of symptom. This weighted combination could be an equal weighting of all items or could be an unequal weighting of items based on an appropriate rationale.”
- Line 487-493: It was surprising to those who work with CDER primarily to see Goal Attainment Scaling (GAS) given some seeming prominence in this draft guidance. No voluntarily participating



member to this document is aware of any proposal for GAS being agreed upon by CDER, with strong pushback provided. If other Centers are accepting this method, while CDER is not, it would be very useful to make note of that in the guidance. Or, at the very least to more prominently display and connect the very real challenges with GAS to the section on GAS. Some challenges are described further down in the text, but the link between the challenges and GAS is unclear, which may result in some sponsors overestimating the acceptance and ease of use of GAS.

- Line 577-582: thank you for the suggestion to have an end of day assessment to ensure events in an event-driven diary are not omitted. It may be helpful to add to this something like: “patients should be asked in the end of the day reporting to either report missed events or, if none to report, to indicate such (e.g., affirmatively mark: I have no other episodes to report)”. While many sponsors implement this already, taking the opportunity at the end of the day to confirm that there are no missing events can help minimize concerns about whether the data are accurately capturing all events in a given day or whether events are potentially missing.
- Line 704-732: thank you for highlighting the value of linking change on performance outcomes to something meaningful in patients’ daily lives and that PerfOs may require more evidence to understand what is the true treatment benefit. Performance outcomes are valued by some as “objective” measures that do not suffer from the subjectivity of PRO measures, however, how these measures translate into meaningfulness to patients can often be very challenging. Moreover, with that “objectivity” often comes the loss of interpretability of what the PerfO means in patients’ daily lives. We also see these challenges with digital monitoring assessments (e.g., accelerometer-based endpoints). Please keep the current text in the Final Guidance and consider including examples related to PerfOs and digital monitoring assessments to help inform readers about how these types of endpoints may be linked with meaningful outcomes for patients (e.g., how to answer the question: what does it mean to see improvements on a 6MWT?).
- Line 846-848: thank you for suggesting that anchor response categories might be evaluated as meaningful (or not) in cognitive debriefing interviews. These are good opportunities to leverage a study for multiple purposes if considered in advance. Please retain in the Final Guidance.
- Line 1027-1028: Thank you for including the current agency thinking on the need to evaluate meaningful score changes on raw scores (in addition to transformed scores). Please retain this information in the Final Guidance.
- Line 1122-1131: It is surprising to see that anchor analyses and their resultant range of meaningful change values, which have traditionally been used to describe meaningful within-person change values, are suggested for use to evaluate group level differences. There are members who disagree that this is an appropriate use of anchor-derived values. Would FDA please consider a discussion of this with the public as FDA seeks to improve upon and standardize methods to

ISOQOL

555 East Wells Street, Suite 1100

Milwaukee, WI 53202 USA

Phone: +1 (414) 918-9797 • Fax: +1 (414) 276-3349

www.isoqol.org

evaluate meaningful treatment benefit?

- Lines 1159-1162: this section addresses the use of CDF curves but should be made stronger to further encourage more use of these as standardized approaches to evaluate treatment benefit rather than simply stating they “can be plotted.”

Thank you for the opportunity to provide comments, and thank you for your dedicated efforts to improving patient focused drug development.

Respectfully submitted,

Co-Chairs on behalf of the ISOQOL R&HE SIG:

Ashley F. Slagle, MS, PhD
Principal, COA Scientific and Regulatory Strategy
Aspen Consulting, LLC

Robyn von Maltzahn, MSc
Senior Director and Vaccines/Oncology Patient Centered Outcomes Head
GSK