

**2022–2023
Board of Directors**

September 28, 2022

PresidentJan Elias Hansen, PhD
Genentech
San Francisco, CA,
USA

Docket Number: FDA-2022-D-1385

President-ElectBrian O'Rourke,
PharmD
Brian O'Rourke
Healthcare Consulting
Ottawa, Canada

Dear FDA:

ISPOR – the professional society for health economics and outcomes research - is pleased to respond on behalf of its membership to your consultation entitled “**Patient-Focused Drug Development: Selecting, Developing, or Modifying Fit-for-Purpose Clinical Outcome Assessments; Draft Guidance for Industry, Food and Drug Administration Staff, and Other Stakeholders.**”

Past PresidentIsao Kamae, DrPH, MD
University of Tokyo
Tokyo, Japan

ISPOR is a scientific and educational society with many of its members engaged in evaluating health technologies, including pharmaceuticals, medical devices, and other interventions. We have a large membership living and working in 110 countries globally, across a range of disciplines, including health economics, epidemiology, public health, pharmaceutical administration, psychology, statistics, medicine, and more, from a variety of stakeholder perspectives, such as the life sciences industry, academia, research organizations, payers, patient groups, government, and health technology assessment bodies. The research and educational offerings presented at our conferences and in our journals are relevant to many of the issues and questions raised in this request for information.

DirectorsEberechukwu
Onukwugha, PhD
University of Maryland
Baltimore, MD, USALotte Steuten, MPhil, PhD
Office of Health Economics
London, UKDavid Thompson, PhD
Open Health Evidence &
Access
Boston, MA, USABeth Devine, PhD
University of Washington
Seattle, WA, USAMarlene Gyldmark, MPhil
Roche Diabetes Care
Basel, SwitzerlandAndrea Manca, MSc PhD
University of York
York, UKDong-Churl Suh, PhD
Chung-Ang University
Seoul, Korea**Treasurer (2020-2023)**Sean D. Sullivan, BSc
Pharm, MSc, PhD
University of Washington
Seattle, Washington**CEO & Executive Director**Nancy S. Berg
ISPOR
Lawrenceville, NJ, USA

The response to this consultation was led by the Policy Outlook Committee of our most senior advisory body, the Health Science Policy Council. To engage our membership, we consulted with interested members of our Patient Roundtables, Task Forces, as well as our Rare Disease, and Patient-Centered Special Interest Groups, and solicited our general membership for comments. The attached document provides both summary and line-by-line responses based on their comments. We hope they prove useful.

ISPOR would be happy to answer any questions about our response, and to participate in any follow-up consultations on the relevant program items mentioned within the report.

Sincerely,

Nancy S. Berg
CEO & Executive Director
ISPOR

Patient-Focused Drug Development: Selecting, Developing, or Modifying Fit-for-Purpose Clinical Outcome Assessments; Draft Guidance for Industry, Food and Drug Administration Staff, and Other Stakeholders.

ISPOR Summary Comments, September 28, 2022

We applaud the collaboration and harmonization of CDER, CBER, CDRH guidance in this document. It provides a very consistent regulatory approach to this subject that is most useful to researchers and sponsors.

We recognize that selection, development, and use of COAs are an integral part of PFDD. However, patient engagement and patient experience data (PED) also contribute to medical-product development in other ways (e.g., Go/No-go decision making, refining clinical development plans, protocol development, improvements to trial recruitment, addressing diversity and equity, etc.). With such a heavy emphasis on COAs in this guidance, it could provide the perception that FDA considers COAs as the only activity under PFDD. It is important that the FDA communicates clearly that COAs do not represent the totality of PFDD.

We recommend the FDA emphasize throughout the draft Guidance that the term “patient perspectives” about the impact’s disease and/or treatment have on a patient’s health, and life must come from patients. The impacts on patients’ lives cannot be based on the views of other stakeholders (e.g., clinical specialists as a proxy reporting on patient perspectives based on their experiences caring for patients). For example, the guidance says to “consider patient views/perspectives,” but that is not the same as “gather views/input directly from patients.” It is clear in this document that the FDA understands what is meant by patient perspectives and that it connotes direct patient engagement. Given that this is still a nascent field with many stakeholders who do not understand or have not yet embraced patient engagement, it would help general understanding of the field for this language to be as clear as possible - that “patient perspective” must include direct patient input. It may seem repetitive, but at this point in the evolution of PFDD, that may be necessary to move the field forward.

The guidance describes the important principles for developing or modifying COAs. However, the examples and discussions are often very PRO specific, although the principles are relevant to all types of COAs. As a result, some of the discussion does not apply well to PerfO type COAs, where for example, the health concept identified for a potential treatment benefit can be broader than the PerfO’s specific concept of interest for measurement. Instead, Figure 3 indicates that the concept of interest for measurement should be the same as the meaningful health concept. A revision would clarify that for some COAs, particularly PerfOs, the meaningful health concept and concept of interest for measurement will be different. Clinical validation of new COAs would then need to demonstrate that an effect on the concept of interest shown by the specific COA (eg, a PerfO) enables inferring effects on the broader health concept.

The guidance identifies the traditional four types of COAs in Section II-A, and later discusses how digital health technologies (DHTs) can be used to collect the data for the COAs. The use of technically sophisticated DHT methods, such as actigraphy, enable measuring clinical outcomes that are potentially highly informative of patient health concepts in ways not previously feasible. For example, a study subject might wear the DHT during their normal daily

activities to passively record and report the total number of steps per day, or to report the amount of time per day the patient walks in episodes of at least a criterion pace. These measurements do not fit the definition of a PerFO type COA, and so are entirely excluded from this guidance. This guidance is missing an important opportunity to explicitly identify this type of COA, and provide some understanding of how this type of COA fits into the framework. This type of COA should be given an explicit category name, or alternatively explicitly designated as part of an expanded definition of PerFO if preferred.

Overall, we ask that the Agency consider more specific, prescriptive guidance about the appropriate methods that may be used to assess whether a COA is influenced by processes not part of the concept of interest. For example, providing examples of when to implement cognitive interviews or how to establish appropriateness of recall period with a literature review would be instructive. Furthermore, the Agency may consider a separate section to cover implementation issues (e.g., missing data, mode of assessment, etc.) as these may not be specific to the development of a specific assessment but relate to larger trial-related issues.

We suggest that the Agency provides clarification when introducing interfering influences and how to address them, as there are situations where complete removal of an influence is not possible (for example, double-blinding can be challenging for rare disease studies or in cases when a transition of an originally paper-based COA to eCOA is preferred as it increases accessibility to patients). Additionally, when the guidance suggests that users provide evidence of usability testing or literature reviews of appropriateness of a COA, we ask that the Agency clearly state how and to what degree this evidence should be documented. The guidance should also consider providing applicable suggestions for interfering influences with the use of DHTs, and particularly, considerations when using DHTs to record historic events.

We realize that there are many requests for clarification and/or further detail in this response. We know that sometimes the science is not 100% clear about the best approaches, or that they may be situation-dependent, and the Agency may be hesitant to be too prescriptive in such cases. Nevertheless, these responses reflect a desire for information about decisions that researchers and sponsors must make in the course of their work. In many cases, even recognizing that there may be several feasible approaches, perhaps depending on the specific case, would be useful.

Finally, as a complement to your definitions on page 5, you may want to reference an ISPOR task force report that also provided definitions of those terms:

Citation (already in your reference section): Walton MK, Powers JH III, Hobart J, et al. Clinical outcome assessments: a conceptual foundation – report of the ISPOR Clinical Outcomes Assessment Emerging Good Practices Task Force. *Value Health*. 2015;18(6):741-752.

Clinical Outcome Assessment (COA): A clinical assessment instrument that is used as the measure of patient outcome in a clinical trial. There are four types: PRO, ClinRO, ObsRO, and PerfO.

Clinician Reported Outcome (ClinRO): A type of COA in which a member of the investigator team is the rater. The investigator's professional training is relied upon to judge what rating or score will be reported.

Observer Reported Outcome (ObsRO): A COA in which observations can be made, appraised, and recorded by a person other than the patient who does not require specialized professional training. The rating is nonetheless influenced by the perspective of the observer.

Patient Reported Outcome (PRO): A COA in which the report comes directly from the patient. The patients' responses to questions about their health condition are recorded without amendment or interpretation by anyone else.

Performance Outcome (PerfO): A COA in which the patient is assessed by performing a defined task that is quantified in a specified way. Although a member of the investigator team may administer the PerfO task and monitor the patient's performance, the investigator does not apply judgment to quantify the performance

Clinical Assessment: An assessment that is susceptible to variation unrelated to the patient's true clinical status owing to 1) variation in patient volition (which may cause rating differences for an individual patient between successive administrations or between patients), or 2) dependence on the judgment of a rater (which may differ between raters owing to differing experiences or perspective).

Outcome Assessment: A measuring instrument that provides a rating or score (categorical or continuous) that is intended to represent some aspect of the patient's medical status. Appropriate outcome assessments may include both COAs and biomarkers.

ISPOR Detailed Comments

I. Introduction

A. Overview of FDA Guidances on Patient-Focused Drug Development

General. As part of the purpose of prior PFDD guidances, the FDA underscored that high-quality measurement is important for several reasons, including making sure to measure what is important to patients during clinical development programs. It is crucial that they really emphasize this point in the introductory section of this guidance too. It's important that the FDA directly acknowledges the need to measure what is important to patients during clinical

development programs.

B. Purpose and Scope of the Guidance

General. This section of the guidance touches on how understanding a disease/condition is important, and a component of that is understanding patient/caregiver perspectives on the disease, its impacts, and therapeutic needs and priorities.

Given this approach to the natural history, patient/caregiver perspectives on the disease, its impact, and therapeutic needs and priorities, the guidance should include these as necessary in the development of further studies.

Lines 91-93: Several best practice publications have described recommendations for developing and evaluating COAs, as well as analyzing and reporting COA data. Readers are directed to relevant publications throughout this guidance. These best practice publications have been released over a number of years by different authors. In the case of conflicting recommendations between publications and over time, how will sponsors reconcile the best practices to follow endorsed by FDA?

II. Overview of COAs in Clinical Trials

A. Types of COAs

Line 157: Among these 4 types of COAs, it is unclear how 'passive' or 'continuous' data collection would be captured or classified. For example, daytime sleepiness or stride velocity, where these data are measured over a long period of time rather than being triggered based on a standardized set of tasks, would not seem to fit the definition of PerfO or any other COA. We recommend more clearly defining where endpoints based on wearables/sensors fit, and if DHTs do not fit into one of the other COAs, where should we place them.

Lines 160-165: As with ClinROs, ObsROs should also allow for the incorporation of an element of "professional" judgment (in this case – the judgment of the person closest to the patient.)

Line 170: "This may include use as a measurement tool for COAs in clinical investigations": Recommend adding "and epidemiological and other types of studies" after the word "clinical".

B. The Concept of Interest and Context of Use for a COA

Lines 192-247: Line 237 is mixing the health concept(s) for the disease with the 'concept of interest selected for assessment' in a manner that does not clearly illustrate how the concept of interest for assessment (measurement) can be quite different for some types

of COAs, such as PerfOs. While the near equivalence is generally the case for PROs, they can be quite different for PerfOs. It would be beneficial to specify the differences in the text and Figure 2. For example, A patient with a neurological disorder affecting upper limb movement has many identifiable affected actions that are part of daily life, and might be asked as items in a PRO. However, it may also be believed that for a particular disorder a PerfO (eg nine-hole-peg test) may be good, reliable, sensitive, COA that will enable inferring how the patient performs on the health concepts. A score on the nine-hole-peg test, however, does not inherently describe the level of a patient's meaningful function. In this case, the peg test could be a very effective COA when the relationship between scores on the test and meaningful patient function is established. The current text does not convey that this may be a valuable approach in many cases.

Line 193-196: COA assessing preventive of prophylactic effect. The guideline refers to a preventive outcome "... prevent the onset of a symptom or a loss of a specific function". In the case of vaccines, the context is not applicable as written. Vaccines are not expected to have targeted effect on symptoms. As it is described in Curran et al. 2021 the vaccine effect may be shown as an overall reduction in disease impact and/or reduction in disease impact in breakthrough subjects.

Line 203: "For some diseases/conditions, important concepts of interest might have already been developed and used in studies based on input from patients, caregivers, clinical experts, and other sources": Consider rephrasing to: For some diseases/conditions, COA measures assessing important concepts of interest might have already been developed and used in studies based on input from patients, caregivers, clinical experts, and other sources. For other diseases/conditions, important concepts of interest might have already been explored and assessed in studies based on input from patients, caregivers, clinical experts, and other sources.

Line 211: Minor comment on phrasing: the text suggests selecting a COI that has the ability to be modified by the treatment. However, until the treatment's effects are evaluated in clinical trials what the treatment can or cannot do is unknown, thus the recommendation, as phrased, is not easily followed. The intended meaning of the line would be clearer by stating the COI selection should be selected considering what the investigation treatment is thought to potentially affect.

Line 215-226: The guidance helps clarify that diseases may have both primary (core) manifestations and downstream features of "how a patient feels or functions". The section gives an example of this relationship and highlights using a COI that evaluates a core concept. Many diseases have separable primary manifestations, such as a neuromuscular disease which can impact leg function, arm function, or other body actions. They are all primary, not downstream. An acceptable COI might evaluate just one of those manifestations or might evaluate several in a single tool.

Line 223 "It may be important to assess a core concept such as dyspareunia defined as

pain with intercourse": Consider using examples that are relevant for the entire population within the context of use. An example like dyspareunia might not be relevant for a subgroup of patients with endometriosis that may not engage in sexual activity.

Line 232: We very much support the statement that: Patient and/or caregiver input can be used to identify which aspect(s) of a concept is most impactful for patients.

Line 236: The term 'health concept' has replaced 'meaningful aspect of health (MAH)' ostensibly to avoid confusion (footnote 14) given the wider use of aspect throughout the guidance. However, MAH is a widely used term and is increasingly used in the literature. Additionally, the word concept is also widely used (in the draft guidance and beyond) and so the change could cause confusion between MAH and concept of interest of for measurement (COI). Retaining the terms MAH and COI is suggested.

Lines 270-272: This point should also indicate whether specific training is needed for COA collection.

C. Deciding Whether a COA is Fit-For-Purpose

Lines 280-283 and 285: Lines 280-283 state that fit-for-purpose in the regulatory context means the same thing as validity. However, Line 285 states two considerations in regulatory decision whether a COA is fit-for-purpose and does not mention validity of the COA. We recommend that FDA be more explicit about measurement properties being a consideration for determining fit- for-purpose – not buried within the 2nd consideration for fitness-for-purpose.

We also recommend that FDA provides more crosswalk from the 2009 guidance and 2018 discussion guides to current draft.

Lines 297-298: "Regardless of whether sponsors propose to use an existing COA, a modified COA, or a newly developed COA, sponsors should present a well-supported rationale for why the proposed COA should be considered fit-for-purpose": Clarify if this is meant to imply that regardless of the endpoint hierarchy of the COA, the sponsor would be required to submit a PRO dossier to support a COA as fit-for-purpose? Are there cases in which the rationale included in the study protocol or submission document would be adequate justification? In cases in which PRO evidence dossier is not required, which study/or submission documents would the FDA expect the sponsor should provide this rationale?

Lines 306-313: The agency plans to evaluate the broader impact on the target patient population and medical product development of collecting additional evidence, when determining a COA's fit for purpose. This description leaves substantial room for interpretation of the phrase "broader impact ... of collecting additional evidence".

Furthermore, a reference to Leptak, et al (2017) fails to provide clarity and may create confusion as it refers to a publication discussing biomarker development. - This statement calls for further clarity to specify what the Agency means by “broader impact”, ie, what are the aspects of impact that will be evaluated (e.g., usability for future clinical practice, population health implications, availability of the instrument for broad use, patient burden, delays to approval)? Also, given that the intended audience of this Draft Guidance may not have expertise in biomarker development, it would be very helpful if the agency could clarify which specific part of the Leptak, et al (2017) article this part references.

Lines 309-310: “The evidence for a particular part of the rationale is weighed relative to the degree of uncertainty about that part.” This is not clear as to how to judge the relative uncertainty and the level of evidence support given a level of uncertainty. Can the Agency please clarify?

III. A Roadmap to Patient-Focused Outcome Measurement in Clinical Trials

A. Understanding the Disease or Condition and conceptualizing Clinical Benefits and Risks

Line 326: Figure 2 is a suboptimal guide for some COAs. This figure appears to combine the concepts of how a patient feels or functions with the concept of interest for actual measurement. This may be frequently the case for PROs, but not for other types of COAs. It is more helpful for sponsors to think about these as two elements, not a single element. For example, in some neurological disorders a patient may have difficulties with activities utilizing fine motor control of the forearm and hand. Identifying specific activities as suggested in figure 1, then combining related ones into categories can be an ADL concept that is meaningful to patients. A useful COA might focus on assessing those activities directly, in which case the meaningful health concept also is the COI. However, a different approach may identify the underlying neuromuscular function causing the activity problems and make evaluating that the concept of interest to be measured. A COA can be developed that evaluates the neuromuscular difficulty, and plan for that COA in a clinical trial. If shown to be reliable and to validly enable inferring the meaningful function from changes in the COA, it would be the basis of a good study endpoint. Although this is a somewhat indirect approach to learning of impacts on meaningful health concept, it has potential to be very useful in taking advantage of the attributes of PerfO type COAs. For example, the “9-hole peg test” might be selected, a well-known PerfO type of COA. Demonstrating the test is reliable and the test results can be used to infer changes in the ADLs of interest would make the test an acceptable COA for a clinical trial. This is frequently the situation with PerfO types of COAs. The clinical validation process of demonstrating the relationship of the COA to the meaningful health concept is clearer for such PerfOs when this two-stage thought process is understood.

Line 326: We suggest that FDA considers including 2 roadmaps -one higher level (as in Figure 2) and one with more detail (like Figure 6 of the discussion guide) - this would benefit both a more general audience that only requires a higher-level overview of the roadmap, and more technical users of the guidance that would require more details fleshed out

B. Select/Develop the Outcome Measure

Lines 343-345: FDA can further enhance understanding of what a COI may be if such a circumstance is described and the guidance states that when there are several primary manifestations of a disease that are all direct effects of the pathophysiology it not a necessity for the COI to encompass all of them. A subset of the primary manifestations is acceptable, and for some treatments or stages of disease, may be preferable to a tool that attempts to measure every core manifestation. It would be underscored by including in the description of the conceptual model and figure 1 that an acceptable COI could also focus on a subset of the health concepts illustrated.

Line 370: Section III-B-2 as written promotes consideration of only COAs that directly measure the meaningful health concept. This has the potential to discourage use or development of PerfOs, or other indirect COAs. Please revise the text and the figure (along with relevant earlier text of In 192-247) to be clear that the thought process leading to a COA to select or develop can separate the meaningful health concept and the specific concept of interest to be measured. In addition, section III-B-2 should also be revised to make the two-stage thought process clearer and how it better supports arriving at PerfO COAs for selection or development.

Line 376: We suggest FDA clarifies that the FDA compendium is also a useful search source in addition to the COA qualification program and the MDDT. Will the compendium continue to be maintained?

Lines 388-390: Rationale & evidence needed to support validity of a COA with similar concept of interest & context of use. If an appropriate COA exists for the same concept of interest, same context of use, and it has been accepted by FDA for use in labeling in the past, can the Agency clarify what type or additional level of evidence is needed to support its qualitative and quantitative validity by a new sponsor? Are sponsor companies required to provide a detailed dossier supporting its validity? Certain components of the evidence used by a prior sponsor to have FDA accept the COA previously may not be publicly available. If so, requiring the new sponsor to supply a fully detailed evidence dossier can pose significant burdens to Sponsors, while hindering drugs trial progress and approval.

Lines 398-440: Rationale & evidence needed to support validity of a COA. The sponsors should “summarize all existing information and evidence that support the rationale for the use of COA and assess how well the rationale is supported by the available

information” The agency should provide clarity as to what extent (and in how much detail) all existing information should be summarized and what is considered a well-supported rationale for its sufficiency.

Lines 416-417: The statement “if the existing evidence leaves too much uncertainty about the appropriateness use in the new context of use...” lacks specificity. It appears that some uncertainty is acceptable, however too much of it would not be acceptable. Unfortunately, this statement provides substantial room for interpretation. The agency should provide additional clarity of how much uncertainty is too much uncertainty. Please take into account that in a drug development process, uncertainty is always present. In the early or mid-stages of clinical development program, which is when decisions about selecting COA instrument take place, uncertainty abounds

Line 433: “Some modifications are unlikely to alter the scores or their interpretation (e.g., changing the display on a tablet-based administration from one item per screen to three items per screen),” It would be helpful for the Agency to provide further guidance or references on what types of modifications do not alter scores or their interpretation., such as any formatting alterations as in the examples already provided.

Lines 475-480: This text highlights the potential for observational studies to importantly contribute to the development of new COAs. This is often not well recognized, and inclusion of this text has potential to be very helpful. That said, the agency should also consider practical aspects of a drug development possess. Sometimes (e.g., in rare diseases or oncology) it is impractical to enroll patients into an observational study. Furthermore, occasionally, clinical trial programs go for registration based on phase 2 data. In some instances, sponsors (e.g., small biotech companies developing treatments in the areas of high unmet need) may not have sufficient resources. Therefore, testing COA’s may be possible in neither a standalone, nor an early clinical trial. We suggest the agency remains open to instances when it is not possible to generate evidence outside of a registrational trial. With this in mind, it would be helpful if the agency could discuss the key principles of conducting validation vs confirmatory clinical trials, so the sponsors could have guidance on which principles to follow.

Lines 488-520: Section "Special considerations for selecting or developing COAs for pediatric populations": While we recognize that pediatric populations entail many specific considerations for regulatory purposes, it may be worth recognizing there are other special populations where similar considerations may apply (e.g., Alzheimer’s disease if the disease progresses as trial progresses) 4a. Can FDA provide recommendations on how to deal with ObsRO/PRO in such special populations more generally? A particular question relates to changes in response capability over time. For example, in pediatrics, once patients are able to self-report, is it recommended to switch from ObsRO to PRO (e.g., ObsRO at BL and PRO in subsequent assessments), or is it recommended to keep the same respondent for the entire trial (despite the risk of missing data because caregivers don’t come to site with patients anymore when patient become adults or

despite the risk of jeopardizing data quality/reliability)?

Lines 521-531: DHTs can be used to implement a COA such as collecting responses to items from a PRO measure... the sponsor should define and provide rationale to justify the use of the DHT. Is the agency referring to eCOA as a DHT in the context of collecting PRO data? This guidance does not seem to appear in the list of guidances or recently released (draft) guidances. The guidance asks for justification on use of DHT for measuring the concept of interest. Would it be preferable to frame this section as the need to justify the means of data capture? In addition, 5. COA Accessibility and Universal Design seems to assume that COA data will be collected using electronic data capture with detailed specification of principles for good interface design.

Lines 521-531: We would recommend this section mention that DHTs can also support the collection of Clinician Reported Outcomes (ClinRo) data.

Line 533: "A portion or all of the target patient population may benefit from accessibility features and universal design considerations": We propose referencing FDA's Good Machine Learning Practices as an existing source of principles if developing COA using AI/ML in order to control for potential bias in models that could impact universal design. The principles are high level and including this guidance helps raise awareness and understanding of the role ML/AI can play in development of COA.

FIGURE 3:

General. This figure depicts that there is a concept not measured, Symptom A. We suggest that FDA adds here that they would expect to see rationale on why Symptom A was not selected to discourage cherry-picking of symptoms to measure (or ignore). We also would be concerned and want to avoid the rationale that it did not get measured because it is "too hard" to measure. This could discourage investment in COA development for the concepts patients prioritize as most important but might not ever be measured.

Figure 3 might be confusing in that those triangles between the scores and the arrows might be seen as backwards arrows, but, per your explanation in ll. 602-609. we assume they are not meant that way. If those triangles are seen as arrows. with the clinical trial sample's data flowing, right to left, back into the diagram, it may seem to leave out some important elements. While we recognize this is a hypothetical illustration, we believe it could be misleading as is and suggest you replace the triangles with figures that make the left-to-right flow of the measurement model very clear. Otherwise, it may be seen interpreted that the diagram implies conceptual-model and measurement-model development happen simultaneously, when they must be sequential. It could also imply measurement-model operationalization happens only in trials, while, ideally, much of the data collection for COA development is best performed before the trials begin, outside of trials We have the following other suggestions about the diagram

- There is no consideration for how the conceptual model is related to the measurement model in terms of either methods or process.

- We believe it would make sense to split the diagram into sections that align with the sequencing of the roadmap outlining progression and type of work over time. This is needed in order for COA developers to be able to segment their regulatory submissions in a way that will be aligned within the industry and useful for FDA review and comment.
- Where the header states COI, it would be helpful to indicate here not only the COI but also the aspect(s) of the COI.
- Where the header states “selected” COA and scores, indicate selected “type of COA tool.”
- The figure does not indicate where concept validation takes place. We believe this is needed and should happen before the conceptual model is finalized with confirmation of content validity before moving on to the measurement model. It should indicate that unless the concept(s), the aspect(s) of the concept(s), the response options, and scoring are confirmed by patients through appropriate qualitative research, it does not make sense to move to the measurement model. We suggest this be indicated in the diagram.

C. Developing a Conceptual Framework

General. Please provide additional detail regarding methodology for the structuring a concept of interest, including the different aspects of the concept.

IV. Developing the Evidence to Support the Conclusion that a COA is Appropriate In a Particular Context of Use

A. The Concept of Interest Should be Assessed by [COA Type], Because...

Line 634: We suggest the Agency provide more clarity surrounding the expectations for the use of Table 1. It would be helpful to provide more information on how the FDA expects sponsors to use this table. In particular:

- Should stakeholders implement this table to be supplementary to the COA dossier or should there be a summary of the evidence from this table in the COA dossier?
- Do sponsors need to provide evidence of all eight components or are there cases where evidence of only certain components is required?

B. The COA Measure Selected Captures All the Important Aspects of the Concept of Interest

Line 650: The draft guidance states that, "the COA Measure Selected Captures All the Important Aspects of the Concept of Interest.": It would be instructive to further specify the source of evidence to evaluate this component (e.g., qualitative studies, literature review, expert input, etc.). Per the 2009 guidance, this would have come under content

validity, however as it is less clear in this new guidance, we would suggest providing more clarity.

D. Scores of the COA Are Not Overly Influenced by Processes/Concepts That Are Not Part of the Concept of Interest

Lines 682-683: Much of this section pertains more to implementation of PROs rather than evidence requirements for selecting, developing or modifying fit-for-purpose COAs. Consider a separate section to cover implementation issues (e.g., missing data, mode of assessment, or expectation bias). These issues are not specific to the development of a specific PRO but are related to the rigor with which PRO data is collected. The issue of expectation bias is also not an issue to a specific PRO but in the way in which a trial is developed. This advice also conflicts with the FDA ASCO workshop of June 29 focusing on how to bring PRO data effectively into single arm and non-randomized studies.

Line 701: If cognitive interviews are the method selected to assess interpretation, it would be helpful to clarify how to determine group sample sizes to support that the score is not influenced by demographic characteristics. Furthermore, in reference to using qualitative (e.g., cognitive interview) methods to evaluate potential differences between groups, we would suggest providing a reference to ISPOR's good practices documents or other literature.

Lines 715-724: The guidance notes that it is important for scores of the COA to not be overly influenced by processes/concepts that are not part of the concept of interest. The guidance recommends performing qualitative cognitive interviews and quantitative methodologies such as differential item functioning (DIF). This recommendation is in supplement to appropriate cultural adaptation and linguistic validation methods. Can the Agency clarify whether sponsors are expected to show proof of DIF for their proposed COAs? If so, is this recommendation meant to apply across a range of demographic characteristics rather than just translations? If this is not considered to be a requirement by the Agency, would appropriate adaptation and linguistic validation, as well as sufficient qualitative interviews be sufficient? Moreover, the increased sample size needed to demonstrate DIF may be not practical (e.g., in some rare diseases or in oncology). Furthermore, it would be helpful to provide clarification on what group difference is "large enough" to impact score without needing to conduct a DIF study. We suggest that FDA provides more references and detail.

Line 729: "For COAs that involve a recall period (e.g., past 24 hours, past 7 days), sponsors should provide support for the appropriateness of the recall period to be used." Could the Agency clarify whether the use of diaries and/or e-diaries to help complete a COA with a lengthy recall period be an acceptable mitigation strategy?

Line 732: In consideration of the resources required and burden on participants, it may

be unrealistic to test different recall periods in empirical studies. We suggest that the Agency clarify if this is a threat to reliability and/or validity and provide clarification on how to determine the degree of uncertainty. Furthermore, might the Agency consider allowing some flexibility in this language, as it may be too prescriptive?

Lines 734-738: The guidance notes that “cognitive interviews can provide justification that a given recall period is inappropriate, but cognitive interviews cannot provide evidence that respondents can recollect with sufficient accuracy”. Based on precedent, the appropriateness of a recall period has been justified using qualitative interviews with patients. The Agency’s current preferred methodology to present evidence on duration of a recall period to avoid recall bias is not clear as written in the draft guidance. The guidance recommends an empirical study to support accuracy of different recall periods and/or literature reviews of recall accuracy for the same or related concepts of interest. Can the Agency provide an example of how a literature review on recall period accuracy should be conducted? Furthermore, should the literature review be disease specific and focused on a specific context of use?

Lines 742-755 We suggest providing additional context for respondent fatigue/burden when multiple measures are utilized. The burden of PRO completion is more often considered in light of multiple PRO measures rather than focusing on the fatigue and burden represented by a single measure. Additionally, it would be helpful to provide guidance on how to define appropriate numbers of items while considering fatigue/burden. This is a consistently a difficult issue with no clear approaches to resolve beyond “rule of thumb” type recommendations. We are challenged between creating a set of PRO measures that comprehensively assess key attributes of disease and treatment symptoms and impact and having an efficient set.

Lines 760-762: There is a need for more clarity on the use of BYOD (bring your own devices) as it has drawn opposing perspectives, some arguing for and others against the use of BYOD device in the clinical trials. We suggest more clarity through outlining conceptual and practical considerations related to the use of BYOD: (for e.g., principles for selecting this approach, best practices for the execution, key concerns, and errors to avoid).

Lines 764-765: It would be valuable to clarify the FDA’s viewpoint on the transition to eCOA collection. The guidance notes “using a mode of collection different from what was originally used for the COA may raise concerns about comparability of assessment to prior experience”. The transition from paper collection to eCOA collection has been long supported by the FDA. In the past decade, there have been multiple studies demonstrating consistency in psychometric properties when data is collected by paper or eCOA. Many of the legacy PROs in current use were created for paper-based collection but have been consistently and successfully moved to eCOA. There are also best practice documents that advocate for eCOA over paper completion.

Lines 768-771: It would be helpful for the Agency to clarify regarding comparability between different COA modes. For examples: In which conditions is the adaption between different modes acceptable/not acceptable? What criteria will the Agency use to deem different modes as acceptable? During COVID to optimize clinical trial experience for patients, there is a desire to promote flexible approaches for completing COA measures. Can the Agency clarify how to optimize consistent data collection approaches while optimizing patient acceptability? Finally, can the Agency clarify if theirs' concern with multiple modalities is different for "within person" versus "within trial" changes in mode of COA completion. For example, would concerns differ if each patients selected the mode most suitable for their own needs and used it consistently, compared to switching modes throughout the trial?

Line 779-783: Please provide clarity on how usability testing evidence should be documented and to what degree. This is referencing the guidance noting "If a data collection platform has already demonstrated usability in a group of participants...it may not be necessary to conduct a new equivalence study..."

Line 788¹: We recommend broadening this section (section 5) to include different sources of bias: Since there are other prominent forms of bias beyond expectation bias, this section should be broader. For example: [ensuring that] "Different Sources of Bias Do Not Unduly Influence Assessment of the Concept of Interest".

For example:

- Caregiver's report of patient status being influenced by the caregiver's own well-being (ClinRO)
- Parent's report being influenced by social factors, understanding of concepts, or child behavior (ObsRO)
- Patient self-report being influenced by current mood while completing COA(PRO)

¹ Schulz R, Cook TB, Beach SR, et al. Magnitude and Causes of Bias Among Family Caregivers Rating Alzheimer Disease Patients. *Am J Geriatr Psychiatry*. 2013;21(1):14-25. doi:10.1016/j.jagp.2012.10.002

Bennetts SK, Mensah FK, Westrupp EM, Hackworth NJ, Reilly S. The Agreement between Parent-Reported and Directly Measured Child Language and Parenting Behaviors. *Front Psychol*. 2016;7. doi:10.3389/fpsyg.2016.01710

MIDDLETON L, DENNEY D, LYNCH S, PARMENTER B. The relationship between perceived and objective cognitive functioning in multiple sclerosis. *Arch Clin Neuropsychol*. 2006;21(5):487-494. doi:10.1016/j.acn.2006.06.008

Vardy J, Tannock I. Cognitive function after chemotherapy in adults with solid tumours. *Crit Rev Oncol Hematol*. 2007;63(3):183-202. doi:10.1016/j.critrevonc.2007.06.001 820: Practice effects: It should be noted that since the task performance is also an element of other COAs types e.g., the Clinical Dementia Rating (a ClinRO), practice effects may need to be considered for other COA types, where task performance is scored in other ways

Line 815: We recommend loosening the language to recognize situations where double blinding is challenging. The 7th FDA COA-CCT workshop (PROs in open-label trials) discussed that there is no evidence in the literature that an open-label cancer trial or a trial design creates biases in PRO data. It may be difficult for trials in oncology or other rare diseases to be blinded and additionally, patients are likely to experience inadvertent unblinding. Loosening the language allow for sponsors to provide evidence that the study design is appropriate and that PRO scores are interpretable in such contexts.

Lines 832-835: Practice Effects. We suggest clarifying that even in the case of a Randomized Controlled Trial, complete removal of practice effects may not be possible. Additionally, the Agency recommends reviewing natural history or non-affected cohorts outside of a trial to show that practice effects do not overly influence the assessment of the concept of interest. However, such evidence may not be available for all COAs.

Line 835: We recommend referencing existing guidance or considerations that are applicable when using DHTs that record historic events which might influence patient outcomes

E. The Method of Scoring Responses to the COA Is Appropriate for Assessing the Concept of Interest

General. FDA should be more specific about where there are still gaps in evidence regarding paper vs. electronic data capture methods, since there has been recently been a significant amount of evidence demonstrating that there is no major difference across a number of use cases. What are the gaps that remain? Additionally, in this section, the FDA suggests developers should avoid sliding scales due to “known limitations”. However, no preferred alternatives are mentioned in the guidance. We ask that alternatives be further expanded on.

General. This section may be too technical for some of the intended audience - providing more examples about how these models can be applied, may be helpful.

General. In general, we would appreciate more discussion around FDA's stance on IRT and what evidence is expected to be provided when sponsors use this approach.

Line 861: Please clarify why the specific models (reflective indicator and composite indicator) were highlighted and the implications for their use in evidence packages. In previous guidances, the type of models used were less prescriptive.

Line 908 “When using IRT models to design, evaluate, or score a COA, additional information concerning the items and scale can be provided.”: Please clarify what types of additional information are needed.

Lines 880-881: Unidimensionality assumption. The guideline states “Statistical evidence including, but not limited to, confirmatory factor analysis can be provided to support the reasonableness of the assumption of unidimensionality.” When would be considered “reasonable” that the assumption of unidimensionality is met? Would the FDA provide clarifications about which criteria would be considered according to the psychometric model used to validate the COA (e.g., Classic Theory, IRT models)?

Line 968: “Approach to Missing Item or Task Responses”. When using DHTs, missed and/or skipped questions may be eliminated based on the rules of the technology being used. What is the Agency’s guidance in this area?

F. Scores From the COA Correspond to the Specific Health Experience(s) the Patient Has Related to the Concept of Interest

Lines 990- 1018: Example of evaluating health experience via PRO and PerfO. The example of lower- limb related function illustrates the potential balance between selecting between PRO and PerfO. Similar illustrative examples may be very helpful in the area of cognitive function, when patient performance may be less observable than in the case of limb function. Alternatively, examples of approaches related to capturing a child’s cognitive and physical function in neurodegenerative diseases would be very helpful. We encourage the agency to provide further clarity in these areas.

Line 1033: "When a sponsor is collecting convergent evidence, FDA notes that the correlation coefficient cutoffs based on Cohen (1988) may be too low to be considered as a moderate and/or strong correlation": Does the FDA have recommendations for cutoffs to use or what supportive evidence/rationale do sponsors need to show to support the cutoffs selected?

Line 1039: "When interpreting correlation coefficients, sponsors should consider the size of the corresponding coefficient of determination and how the distribution of the variables might impact the magnitude of the correlation (e.g., attenuation due to restriction of range)": Please consider clarifying this statement as it is unclear what FDA's intentions are.

G. Scores from the COA Are Sufficiently Sensitive to Reflect Clinically Meaningful Changes Within Patients Over Time in the Concept of Interest Within the Context of Use

General. To make it clear that Section G.1 is discussing responsiveness per se and not necessarily clinically meaningful change, it may be useful to elevate part of the comment in footnote 29 to the main text.

H. Differences in the COA Scores Can Be Interpreted and Communicated Clearly in Terms

of the Expected Impact on Patients' Experiences

General: It is unclear how Section H relates to Section G, or if Section H should be part of Section G.

Line 1118: This is particularly important for rare diseases. For instance, for some of these patients, stabilization of function or tumor size can be clinically meaningful and significant.

APPENDIX D: Performance Outcome Measures

General: Conceptual Framework for a Concept of Interest Assessed by a Clinician-Reported Outcome Measure is most applicable

Line 1623: Potential interference of functions or abilities that are not part of the concept of interest is described as an issue specific to PerfO assessments. However, this issue may be relevant for all COA types. For example, for PROs biases towards emotionally valenced material, negative affect and catastrophizing may impact self-report across different functions and abilities in affective disorders, which are not mood and affect concepts; or for ObsRO the impact of challenging behaviors in patients may influence the report of caregivers and family members across different functions and abilities in neurodevelopmental or neurodegenerative disorders. This should be a part of the discussion from line 788 and go beyond expectation bias e.g., "Different Sources of Bias Do Not Unduly Influence Assessment of the Concept of Interest".

Line 1639: Draft guidance states "Consider gamification to make the task more appealing". The science of gamification of tasks is not well developed and gamification may both fail to solve the issue of refusal to complete a task, but also introduces a variety of new factors such as additional demographic and sociocultural bias, changing levels of appeal and engagement over time and with repeated assessment, novel design elements and aging of design etc. that may complicate development and validation of tasks. The guidance might more reasonably caution against gamification as a solution, without undertaking substantial additional work. In general, it is the nature of the relationship between trial participants and clinical researchers and a shared understanding of the value of the data that is most important to voluntary completion. Even very burdensome (e.g., a 2-hour neuropsychological test battery) or very invasive (eg, lumbar puncture) assessments are acceptable to patients when shared value is well understood. If we devalue PerfO assessments and consider them as add-on assessments and low burden, assume we do not need to engage patients in the importance of the research endeavor, or reframe assessments as entertainment, missing data may be more likely. Gamification may even contribute to a perception of low value. Note too that since all assessments are performed voluntarily and consent can be withdrawn at any time for any element of a clinical trial, the term "Voluntary" does not appear to describe the issue well or as specific to PerfO assessments. I would recommend deleting this section given that the issue of burden and fatigue is described above and to simply add the comment on missing data.



The professional society for health
economics and outcomes research

Improving healthcare decisions

505 LAWRENCE SQUARE BLVD SOUTH
LAWRENCEVILLE, NJ 08648

P +1-609-586-4981
F +1-609-586-4982

info@ispor.org
www.ispor.org

APPENDIX E: Example Table to Summarize Rationale and Support for a COA

General. Consider adding a column to provide evidence to discuss degree of uncertainty and level of evidence needed

ISPOR would like to acknowledge Bryan Bennett, Eleanor Perfetto, Marc Walton, Naomi Suminski, and Rinchen Doma for their time and expertise in assisting with this response to the draft guidance.