U.S. FOOD & DRUG ADMINISTRATION

Public Meeting Discussion Documents:

Draft COA Evidence Outline

Draft COA Evidence Dossier Template

September 2025

INTRODUCTION TO DISCUSSION DOCUMENT

The *2009 Guidance for Industry, Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims* included an Appendix titled, "Information on a PRO Reviewed by the FDA".

When the draft PFDD Guidance Series was published, FDA received many comments requesting that a similar appendix be provided somewhere in the PFDD Guidance Series. Similar requests were received during the public meetings where PFDD Guidance documents have been discussed.

In an effort to address these requests, as well as to follow the evidence-based rationale described in the PFDD Guidance Series, reduce stakeholder burden, improve review efficiency, reduce the number of information requests (IRs) sent, and reduce the submission of unnecessary information, FDA is sharing this discussion document which includes an updated outline of areas that should be addressed in COA documents provided to the FDA for review. This discussion document also includes a detailed evidence dossier template that industry may find useful and that may facilitate the submission of evidence to FDA to support a Clinical Outcome Assessment.

It is important to note that sponsors may not need to complete all of the subsections within the templates. For example, the PFDD Guidance Series covers methods that may not be used in a specific development program (e.g., bookmarking). Those parts of the template can be marked N/A.

These documents may also not outline everything that is needed. For example, if sponsors are using a novel approach, or a key element is missing in the template language to support the evidence, it should be included.

What has changed?

The draft evidence dossier template provides insights into FDA's thinking. Based on interactions with sponsors since the 2009 guidance, this document provides more detail with the goal of helping to get sufficient information with adequate levels of detail in submissions the first time. FDA has also reordered some of the sections to better align with review practices and removed some subsections.

Some of the revisions and additions that can be found in the Evidence Dossier Template include:

- The content has broadened to include all types of COAs.
- FDA has added a section for the Executive Summary (Section I). This section should include a table of rationale components.
- While there is a new section on meaningful aspects of health, the content of this section is very similar to the section of the 2009 Appendix which discusses the Conceptual Framework.

- There is encouragement to use links to trial documents elsewhere in a submission, when appropriate. This includes links to materials that were previously summarized within the dossier.
- Areas to:
  - Explicitly describe the concept of interest
  - Include a summary of the COA-related regulatory history (Appendix H)
  - Include, once trial data are available, information from blinded and unblinded analyses

Table mapping the 2009 Appendix to the current draft documents

2009 PRO Appendix Section	***Draft 1* COA Dossier Section**
I. Instrument	IV. Description and Justification for COA Appendix A
II. Targeted Claims	II. Context of Use Information
III. Endpoint Model	II. Context of Use Information
IV. Conceptual Framework	III. Meaningful Aspect of Health (MAH) IV. Description and Justification for COA (*if concept of interest (COI) is different from MAH*)
V. Content Validity	IV. Description and Justification for COA
VI. Assessment of Other Measurement Properties	IV. Description and Justification for COA V. Description and Interpretation of COA-based Endpoint(s) Appendices E, F, G, I
VII. Interpretation of Scores	V. Description and Interpretation of COA-based Endpoint(s)
VIII. Language Translation and Cultural Adaptation	IV. Description and Justification for COA
IX. Data Collection Method	IV. Description and Justification for COA Appendix B
X. Modifications	*Not a standalone section. Evidence should be provided in IV.C., as appropriate/as needed* Appendix A
XI. PRO-Specific Plans related to Clinical Trial and Analysis	II. Context of Use Information V. Description and Interpretation of COA-based Endpoint(s) Links and directions to specific portions of submission(s)
XII. Key References plus Appendix A: User Manual Appendix B: Item Tracking Matrix Appendix C: Transcripts	Appendices B, C, D, J

DRAFT COA DOSSIER OUTLINE

OUTLINE OF A CLINICAL OUTCOME ASSESSMENT (COA) EVIDENCE DOSSIER

I. EXECUTIVE SUMMARY
- A. Summary of COA-Based Endpoint Approach
- B. Summary of the Rationale to Support the Fitness-for-Purpose of Each COA
- C. Summary of Qualitative and Quantitative Evidence to Support the Evaluation of the Meaningfulness of COA-based Endpoint Scores

II. CONTEXT OF USE INFORMATION
- A. Clinical Trial Population
- B. Clinical Trial Design
- C. Clinical Trial Objective/Hypothesis Corresponding to COA
- D. Endpoint Definition(s) and Positioning
- E. Desired Claim Related to COA-Based Endpoint

III. MEANINGFUL ASPECT(S) OF HEALTH (MAH)

IV. DESCRIPTION AND JUSTIFICATION FOR COA
- A. Measured Concept of Interest (COI)
- B. Description of COA
- C. Evidence-Based Rationale for Interpreting Scores from the COA as Reflecting the COI within the COU

V. DESCRIPTION AND INTERPRETATION OF COA-BASED ENDPOINT(S)
- A. Computation of Proposed COA-Based Endpoint(s)
- B. Rationale to Support Endpoint Selection
- C. Evaluating the Meaningfulness of COA-Based Endpoint Scores

VI. APPENDICES

Appendix A: COA Copies/Screenshots

Appendix B: COA-Related Manuals and Training Materials

Appendix C: Item Tracking Matrix

Appendix D: Transcripts

Appendix E: Psychometric Analysis Plans (PAPs) and Psychometric Analysis Report

Appendix F: Standalone/Observational Study Protocol(s)

Appendix G: Results of Usability Testing

Appendix H: Summary of COA-Related Regulatory History

Appendix I: Additional Results of Analyses

Appendix J: Copies of Literature Cited/Referenced

DRAFT COA EVIDENCE DOSSIER

DRAFT COA EVIDENCE DOSSIER

[COA Name] Evidence Dossier

To support the use and interpretation of one or more proposed COA scores and their corresponding COA-based endpoint, a COA Evidence Dossier(s) should be submitted. Sponsors may decide to submit a single Evidence Dossier or an Evidence Dossier per each relevant COA.[1]

The following sections correspond to areas that should be addressed in the COA documents provided to the FDA for review as part of submissions related to medical product development and regulatory decision-making. The extent of evidence provided in each section will vary depending upon the COA measure as well as the stage of development for the COA and/or medical product.

1. The submitted evidence dossier should include all sections in this template. If a section is not applicable to the development program, that section may be left blank with "NA" accompanied by a brief explanation of why the section is not relevant.
2. Refer to the content of specific guidances[2] for additional information concerning the types of evidence needed in a given section of the dossier.
3. All embedded hyperlinks should be active, and the PDF should be optimized for full-text searchability. A complete table of contents, a list of tables, and a list of figures should be included.
4. In addition to the general considerations discussed herein, a study may need to meet specific statutory and regulatory standards governing the collection, processing, retention, and submission of data to the FDA to support regulatory decisions regarding a marketed

[1] For medical device submissions, the recommendations herein should be implemented consistent with the least burdensome principles outlined in the guidance for industry and FDA staff *The Least Burdensome Provisions: Concept and Principles* (February 2019) and in the guidance for industry and FDA staff, and other stakeholders *Principles for Selecting, Developing, Modifying, and Adapting Patient-Reported Outcome Instruments for Use in Medical Device Evaluation*.

[2] See the FDA guidances on patient-focused drug development 1-4 and the guidance for industry and FDA staff and other Stakeholders *Principles for Selecting, Developing, Modifying, and Adapting Patient-Reported Outcome Instruments for Use in Medical Device Evaluation* (January 2022).

or investigational medical product. This document focuses on more general considerations that apply to many types of studies, and sponsors should consult with the review division and relevant guidance regarding any other applicable requirements.

5. For submission of a marketing application to CDER and CBER, the evidence for the COA measure should be submitted in Module 5.3.5.3 of the electronic common technical document,[3] and a reviewer's guide providing an explanation of the format of Module 5.3.5.3 and links to other relevant documents in other sections of the eCTD (e.g., study protocols, study reports, and data folders with the define.xml files) should be included. Full articles for all cited references should be submitted to Module 5.4.

[3] Guidance for industry *Providing Regulatory Submissions in Electronic Format — Certain Human Pharmaceutical Product Applications and Related Submissions Using the eCTD Specifications* (September 2024).

I. EXECUTIVE SUMMARY

A. Summary of COA-Based Endpoint Approach

This high-level summary should cover the following:

1. Intended context of use (COU).
2. Description and justification for the meaningful aspect(s) of health (MAH).
3. Concept of interest (COI), if not identical to the MAH.

The table below should include all relevant COAs in the program. If there are multiple COUs, include one table per COU.

Table 1. Summary of COAs by Context of Use

		COAs				
MAH	**COI for Measurement**	*Type*	*Name*	*Score*	**COA-based Endpoint**	**COA-based Endpoint positioning**

B. Summary of the Rationale to Support the Fitness-for-Purpose of Each COA

Provide a rationale that explains how and why the COA is expected or intended to work.

A table summarizing the components and corresponding supporting evidence that make up the rationale above should be included (see example below). The rationale and components should be tailored to the proposed interpretation of COA scores in the COU. The Table should include the components and a high-level summary, including statistical results if applicable, of supporting evidence with a link to the corresponding section of the COA dossier (i.e., Section IV) where the full justification and relevant results (including tables and figures) can be found.

Language and specific components shown in table below refer to a hypothetical example and are for the purpose of illustration only. Sponsors should include components that are relevant to their development program using language that is clearest for their situation.

Table 2. Example Summary of Components That Support the Rationale for a COA to Measure a Concept of Interest in a Specific Context of Use

No.	Component	Support[a]
A	**The concept of interest, [FILL IN], should be assessed by a [PRO/ObsRO/ClinRO/PerfO] because . . .**	
B	**[NAME OF MEASURE] includes all the important parts of [CONCEPT OF INTEREST].**	
B.1	*Other subcomponent?*	
C	**[NAME OF MEASURE] is administered as intended by the measure developer.**	
C.1	*Other subcomponent?*	
D	**[RESPONDENTS PROVIDING INFORMATION] understand the instructions and [ITEMS or TASKS] as intended by the measure developer.**	
D.1	Respondents support that they are able, without difficulty, to select an answer that matches their experiences.	
D.2	*Other subcomponent?*	
E	**The method of scoring responses to the [NAME OF MEASURE] is appropriate for assessing [CONCEPT OF INTEREST].**	
E.1	Method of combining responses on items/tasks (i.e., scoring) is appropriate to assess [CONCEPT OF INTEREST].	
E.2	The assumption(s) of the psychometric model used to generate COA scores are sufficiently met.	
E.3	Empirical assessment of missing data rule shows the method for handling missing [ITEM or TASK] responses in scoring is appropriate for assessing [CONCEPT OF INTEREST].	
E.4	Scores from the COA are sufficiently sensitive to reflect changes in [CONCEPT OF INTEREST] within the COU.	
E.5	*Other subcomponent?*	
F	**Scores from the [NAME OF MEASURE] are not overly influenced by processes/concepts that are not part of [CONCEPT OF INTEREST].** [Select and comment on appropriate rows for the type of COA]	
F.1	[ITEM OR TASK] interpretation or relevance does not differ substantially according to respondents' [SEX, AGE, EDUCATION LEVEL, CULTURAL/LINGUISTIC BACKGROUND, etc].	
F.2	Recollection errors do not overly influence assessment of [CONCEPT OF INTEREST].	
F.3	Respondent fatigue or burden does not overly influence assessment of [CONCEPT OF INTEREST].	

F.4	Implementation and/or study design decisions do not overly influence assessment of [CONCEPT OF INTEREST].	
F.5	*Other subcomponent?*	
G	**Scores from the [NAME OF MEASURE] are not overly influenced by measurement error.**	
G.1	Test-retest reliability coefficient indicates scores from the [NAME OF MEASURE] are not overly influenced by variation over time within clinically stable patients.	
G.2	*Other subcomponent?*	
H	**Scores from the [NAME OF MEASURE] correspond to [MEANINGFUL ASPECT OF HEALTH] related to [CONCEPT OF INTEREST].**	
H.1	Correlation coefficients for the relationship between scores on the [NAME OF MEASURE] and [OTHER COA] are as hypothesized *a priori*.	
H.2	Empirical comparisons of scores from the [NAME OF MEASURE] for patient groups known to differ with respect to the [MEANINGFUL ASPECT OF HEALTH] show relationships hypothesized *a priori*.	
H.3	*Other subcomponent?*	
Other?	*Other components needed to justify interpreting scores as measures of the concept of interest?*	

[a]Summary of supporting evidence with reference or link to more details about the evidence.

C. Summary of Qualitative and Quantitative Evidence to Support the Evaluation of the Meaningfulness of COA-based Endpoint Scores

1. Provide a summary of results of efficacy analyses of COA-based endpoint(s) and link to clinical study report(s) (CSR)
2. Support a conclusion of the clinical meaningfulness of COA-based endpoint trial results including, as appropriate, the range of meaningful score differences (MSDs) or meaningful score regions (MSRs) and a summary of qualitative evidence

II. CONTEXT OF USE INFORMATION

This section of the dossier should include information that is specific to the registrational and/or pivotal trial(s) intended for regulatory decision-making.

A. Clinical Trial Population

1. Description of patient population and link to protocol(s)

B. Clinical Trial Design

1. Schedule of COA Assessments
2. Links to Study Schema and the full Schedule of Assessments

If applicable, include the same information for any externally controlled studies.

C. Clinical Trial Objective/Hypothesis Corresponding to COA

D. Endpoint Definition(s) and Positioning

In this section sponsors should, as part of the description of the endpoint definitions and positioning, explicitly note whether multiple endpoints are being used to support inference about a single meaningful aspect of health and specify whether each endpoint is controlled for multiplicity.

E. Desired Claim Related to COA-Based Endpoint

If applicable, provide a link to proposed United States Prescription Information (USPI).

III. MEANINGFUL ASPECT(S) OF HEALTH (MAH)

In this section, provide the following:

1. Conceptual disease model of patients' experience of living with the condition.
2. Definition/Conceptual framework for the MAH.
    a. Diagram of hypothesized (proposed) or final conceptual framework for the MAH showing relationships among patient experiences, items/tasks, domains, and total score. Ensure that the conceptual framework corresponds to the clinical trial endpoints described in the clinical trial protocol and proposed as labeling claims.
    b. Evidence to support the conceptual framework detailing inclusion and omission of patient experiences (including qualitative data from patients, caregivers, clinicians, and/or other relevant parties.)
3. Evidence of importance to patients and/or caregivers, including qualitative data from patients and/or caregivers.
4. Explanation of potential to show effect of intervention on meaningful aspect of health within duration of trial.

IV. DESCRIPTION AND JUSTIFICATION FOR COA

A. Measured Concept of Interest (COI)

1. A description of the COI's relationship to the MAH.
2. If the COI is different from the MAH, a definition/conceptual framework for COI, including:
    a. A diagram of the hypothesized (proposed) or final conceptual framework for the COI showing relationships among patient experiences, items and tasks, domains, and total score. (Ensure that the conceptual framework corresponds to the clinical trial endpoints described in the clinical trial protocol and proposed as labeling claims.)
    b. Evidence to support the conceptual framework for the COI (including qualitative data from patients, caregivers, clinicians, and/or other relevant interested parties.)

B. Description of COA

1. Version of COA (final), type of COA (e.g., PRO, ObsRO)
    a. Exact copy of COA as it was and/or will be administered in the trial (e.g., screenshots of eCOA, photocopy of paper) should be included in Appendix A1 and linked to from here. The screenshots and/or photocopies should be of high resolution and legible.
    b. Any instructions for administration not included in the exact copy of the COA with links to relevant documents (e.g., user manual, training manual).
2. Recall period.
3. Response options and how numeric values are assigned to each response option (if applicable).
4. Mode(s) of administration.
5. Brief description of scoring algorithm(s) of the COA, including any missing data rules.
    a. Total possible score range(s).

C. Evidence-Based Rationale for Interpreting Scores from the COA as Reflecting the COI within the COU

Data to support the evidence-based rationale should be described in sufficient detail and with references to the qualitative and quantitative study protocols, as applicable. Results of analyses, including figures and tables where applicable (e.g., scatterplots for correlations), should be relevant to the discussion. Additional results can be included in Appendix I. All tables and figures should be hyperlinked and included in the List of Tables and List of Figures at the beginning of the evidence dossier. Some

analyses may be relevant to more than one component. In such cases, the analyses and results may be included once and referenced as appropriate.

Evidence should be specific to the planned clinical trial population and indication (i.e., the COU). Sponsors should discuss the submission of evidence from related COUs with the review division before submission.

Component A: The concept of interest, [FILL IN], should be assessed by a [PRO/ObsRO/ClinRO/PerfO] because …

Expanding on this component may not always be needed but would be expected when there are multiple COA types being used to assess a given MAH. If digital health technology (DHT) is used to collect data to assess a given MAH, include the rationale for use of the DHT.

Component B: [NAME OF MEASURE] includes all the important parts of [CONCEPT OF INTEREST]

1. Evidence that the COA captures all of the most clinically important concepts and items, and that items are complete and relevant (appropriate). (Note that all evidence should include a summary demonstrating how study participants are aligned with the target patient population. For example, one can specify detailed demographics and clinical characteristics of study population in a table).
    a. Literature review and documentation of expert input
    b. Qualitative study protocols, interview and discussion guides, data analysis plans, coding dictionaries, summaries of results, and interview transcripts[4] for:
        i. Focus group testing
        ii. Open-ended patient interviews
        iii. Cognitive interviews
        iv. Other methods used to gather patient input.
    c. Origin and derivation of items with chronology of events for item generation, modification, and finalization. (For example, one could provide an item tracking matrix for versions tested with patients showing items retained and items deleted providing evidence of saturation.) Summarize here and include complete materials under Appendix C.
    d. Qualitative study summaries that support validity of item content

[4] Include transcripts in Appendix D when submitted.

i. Link to any quantitative data that complement the qualitative study findings (e.g., item-level floor/ceiling effects).

Component C: [NAME OF MEASURE] is administered as intended by the measure developer.

1. Clear and detailed summaries of any user manuals. (The manual(s) should be included in Appendix B, and if multiple modes of administration were used, all modes should have accompanying user manuals.)
2. Evidence of successful completion of a standardized training program by personnel at all sites.
3. Evidence of intermittent observation of personnel throughout the clinical trial to ensure ongoing adherence to the protocol.
4. If raters are used (e.g., central raters), rater training and certification and credentials; description of process for review, interpretation, and adjudication (if applicable); and scoring of the COA with copies of standard operating procedures (SOPs).

Component D: [RESPONDENTS PROVIDING INFORMATION] understand the instructions and [ITEMS or TASKS] as intended by the measure developer.

D1. Respondents support that they are able, without difficulty, to select an answer that matches their experiences.

1. Summary of the evidence to support the use of the item/task response options or responses
   a. Results of cognitive interviews or other evidence that address whether respondents are able to interpret response options as intended and appropriately choose a score that represents their symptom severity or other patient experience
   b. Brief summary of quantitative analyses results assessing item-level floor and ceiling effects
2. Updated documentation of item tracking (item tracking matrix) as indicated and reflective of final COA

Component E: The method of scoring responses to the [NAME OF MEAURE] is appropriate for assessing [CONCEPT OF INTEREST].

E1. The method of combining responses on items/tasks (i.e., scoring) is appropriate to assess [CONCEPT OF INTEREST].

The approach for combining responses to multiple items/tasks is often expressed as a measurement model that relates responses to particular items/tasks to the score(s) thought to measure the concept of interest.

1. The measurement model
    a. The type of measurement model (i.e., reflective indicator model or composite indicator model) used to generate COA scores
    b. Description and justification for the measurement framework (e.g., classical test theory, item response theory, Rasch)
    c. Rationale and justification for the way in which responses to the multiple items are combined (e.g., as equally weighted average, unequally weighted average, equally weighted sum, or unequally weighted sum) to arrive at a score for the COA
    d. Includes both quantitative evidence and qualitative evidence
    e. Rationale and justification for any score transformation (e.g., log or linear) or normalization (e.g., z-score), if applicable
    f. Scoring manual in Appendix B (if applicable)

E2. The assumption(s) of the psychometric model (e.g., partial-credit model, graded response model) used to generate COA scores are sufficiently met.

1. Description and justification of psychometric model
2. Assumptions of psychometric model and how each assumption was assessed (e.g., model goodness-of-fit, item-level goodness-of-fit)

E3. Empirical assessment of missing data rule shows the method for handling missing [ITEM or TASK] responses in scoring is appropriate for assessing [CONCEPT OF INTEREST].

1. Missing data rule(s), i.e., conditions under which a COA score can still be computed in the presence of missing item/task responses
    a. How missing items or tasks are to be identified and scored as well as the minimum number of items/tasks responses to compute a score
2. Justification for missing data rule(s) (e.g., details and results of a missing data simulation study)
3. Additional considerations for missing item or task responses
    a. Methods implemented in registration/pivotal trial(s) to prevent missing item or task responses
    b. Reasons for missing item or task responses collected during registrational/pivotal trial(s)
4. Link to relevant section in Scoring Manual, if applicable

E4. Scores from the COA are sufficiently sensitive to reflect changes in [CONCEPT OF INTEREST] within the COU.

1. Results of quantitative evaluation of floor and ceiling effects at the COA score level

Component F: Scores from the COA are not overly influenced by processes/concepts that are not part of the concept of interest.

Here one should specify interfering influences on responses to items/tasks and assess the presence and strength of those influences that may influence score interpretation.

F1. Item or Task Interpretation or Relevance Does Not Differ Substantially According to Respondents' Demographic Characteristics (Including Sex, Age, and Education Level) or Cultural/Linguistic Backgrounds.

1. Concerns about differences
    a. Qualitative evidence (e.g., cognitive interviews) used to characterize potential differences across demographic groups
    b. Quantitative analyses (e.g., measurement invariance and/or tests of differential item functioning) to characterize the presence and magnitude of any group differences
2. If there is a need for cultural/linguistic adaptation, include:
    a. Process used to translate and culturally adapt the instrument for populations that will use them in the trial
    b. Description of patient testing, language- or culture-specific concerns, and rationale for decisions made to create new versions.
    c. Copies of translated or adapted versions
    d. Rationale and/or scientific evidence supporting the cultural and linguistic equivalence of the original and adapted instruments
    e. Description and discussion of any issues related to the adaptation
    f. Copies of translated or adapted versions

F2. Recollection Errors Do Not Overly Influence Assessment of the Concept of Interest. (PROs, ObsROs, and ClinROs).

1. Recall period of the COA and evidence to support the use of the selected recall period including:
    a. Qualitive evidence (e.g., respondents' understanding of the recall period used; respondents' explanations of how they arrived at their answers and what period they were thinking about or what strategy they used to recollect; and respondents' opinions about how far in the past they can recall with accuracy)

F3. Respondent fatigue or burden does not overly influence assessment of [CONCEPT OF INTEREST].

1. Respondent Fatigue or burden
    a. Description of the measure length and/or complexity of the target COAs
    b. Evidence to support that the respondents do not feel fatigued and/or burdened (e.g., from cognitive interviews and/or usability testing)
    c. Approaches implemented in registrational/pivotal trial(s) to minimize fatigue and burden, especially when a COA is included as part of a long battery of measures and/or is frequently assessed (should be included in trial protocol(s))

F4. Implementation and/or study design decisions do not overly influence assessment of [CONCEPT OF INTEREST].

Provide evidence on whether implementation and/or study design decisions might influence how well COA scores reflect the concept of interest.

1. Expectation bias
    a. Approaches implemented in registrational/pivotal trial(s) to minimize the influence of biases, including expectation bias (e.g., a detailed summary of the study design (e.g., randomized, controlled, and double-masked))
        i. When masked trials are not feasible, evaluation of the potential for influence on the COA by expectation bias and evidence that it will not unduly affect the study results
    b. Discussion of any unintentional unblinding (e.g., due to adverse event or clinical laboratory values)
2. COA Implementation
    a. Order of COA administration (in particular, relative to all other COAs administered), and link to relevant section(s) of protocol(s) where this is prespecified
    b. Design of data collection system (e.g., electronic diary, paper diary)
        i. Whether respondents are allowed to skip items/tasks
        ii. When respondents are not allowed to skip items/tasks, how response data from incomplete COA assessment are handled, such as when a respondent starts but does not finish the COA at a given assessment timepoint
    c. When using electronic data capture, include:
        i. Nature of the electronic equipment (e.g., operating system, screen size) and whether a respondent-provided or Sponsor-issued device is used

ii. How respondents are reminded to complete planned COA assessment
iii. Reporting windows for an assessment timepoint as well as how these windows are enforced
iv. Description of interaction between the eCOA and the user (e.g., menu options, skip logic, confirmation screens.
v. Description of robustness of the implementation (e.g., usability testing, pilot study, frequency of device backups, service level for provisioning replacement devices, help desk for site staff as well as patients)

d. When applicable, COA completion rates at relevant timepoints
   i. Provide reasons for missing observations at relevant timepoints (reasons for missing observations include, but are not limited to, patient was too ill to complete, site staff error, technical problems, patient non-compliance, loss to follow-up)
   ii. Specifically for oncological products, refer to FDA guidance for industry Submitting Patient-Reported Outcome Data in Cancer Trials (November 2023)
e. Practice effects (if applicable) and strategies implemented in registrational/pivotal trial(s) to minimize the influence of practice effects on trial results

Component G: Scores from the COA are not overly influenced by measurement error.

Sponsor should evaluate the most likely sources of variability in scores within the context of use, and discuss quantitative study(ies) conducted to support that the COA scores are not overly influenced by measurement error.

G1. Test-retest reliability coefficient indicates scores from the COA are not overly influenced by variation over time within clinically stable patients.

1. Describe and provide justification for:
   a. Criterion to identify stable patients
   b. Assessment time points used for analyses
   c. Model used to calculate ICC
   d. Criteria for assessing adequate level of test-retest reliability
2. Summarize the study results in this section
3. If a standalone reliability study was conducted, include study protocol (including patient inclusion/exclusion criteria) in Appendix F and include a full study report

G2. Inter/Intra-rater reliability coefficient indicates scores from the COA are not overly influenced by variation over different COA raters (e.g., clinicians) or within the same COA rater over time.

1. Summarize the study results in this section including comparability of raters as well as patient characteristics between reliability study and pivotal trial(s)
2. Describe in detail the study design and methods (e.g., the timepoint used to support the analyses, statistical models used to calculate the reliability indexes (e.g., intra-class correlation coefficient (ICC)))
3. Submit rater qualifications/credentials and training materials in Appendix B
4. If a standalone reliability study was conducted, include study protocol (including patient inclusion/exclusion criteria) in Appendix F and include a full study report

Component H: Scores from the COA correspond to the meaningful aspect of health related to the COI.

H1. Correlation coefficients for the relationship between scores on the COA and other supportive COAs are as hypothesized a priori.

1. Convergent/Divergent evidence
    a. Specify which reference measures will be used to support convergent and divergent evidence
    b. Include the a priori hypothesized relationships among the concepts measured by any proposed reference measures and the concept(s) measured by the proposed COA
    c. Specify the criteria for describing the strength of a correlation

H2. Empirical comparisons of scores from the COA for patient groups known to differ with respect to the MAH show relationships hypothesized a priori.

1. Known-groups evidence
    a. Specify appropriate reference measures and justify the corresponding cutoff values that represent clinically distinct levels of symptom severity and/or impairment
    b. Provide details of the proposed model and the hypothesis tests that will be performed

V. DESCRIPTION AND INTERPRETATION OF COA-BASED ENDPOINT(S)

A. Computation of Proposed COA-Based Endpoint(s)

Detailed algorithm, including missing data rule(s) for the endpoint (Specify where the endpoint algorithm is described in the statistical analysis plan (SAP) and include link to final SAP)

1. Evidence to support endpoint algorithm including each missing data rule

B. Rationale to Support Endpoint Selection

For items 1-5 below, references can be made to previous sections of the dossier with more details as needed.

1. The MAH that the endpoint is thought to reflect.
2. The clinical trial objective or hypothesis corresponding to the endpoint, ensuring that the objective or hypothesis is specific (e.g., "To assess the efficacy of product X by comparing the patient-reported physical functioning between arms at 24 weeks" rather than "To assess the efficacy of product X by comparing the patient-reported outcomes of product X vs. Y").
3. The role of the endpoint (e.g., primary endpoint, secondary endpoint, other).
4. How the COA-based endpoint will assess clinical benefit for the intended indication.
5. Explanation for why the selected COA is fit-for-purpose in the planned trial. (PFDD Guidance 3 discusses developing an evidence-based rationale for interpreting the COA score as a measure of the concept of interest. That rationale is part of the overall rationale to support the use of the endpoint.)
6. Evidence that the endpoint provides sufficient precision to estimate the effect of treatment when used with the proposed study design, sample size, and analysis plan. (Measurement error associated with the COA used in the construction of the endpoint is one of the factors reflected in the precision of the endpoint.)
7. Support for the importance of the endpoint to patients and/or caregivers from literature review and/or primary data collection. (There are well-established relevant outcomes such as organ failure and death that do not require additional support. If a multi-component endpoint is used, justification for the components included and the algorithm for combining them into the endpoint is needed.)
8. Strengths and limitations of the proposed endpoint.

C. Evaluating the Meaningfulness of COA-Based Endpoint Scores

1. Description and justification for the methodologies used to evaluate the meaningfulness of COA-based endpoint scores
    a. Methodologies used to derive score interpretation metrics (i.e., meaningful score differences (MSDs) or meaningful score regions (MSRs)

b. Methodologies used to apply derived score interpretation metrics (e.g., average horizontal gap, vertical gap)

2. For the chosen score interpretation metric(s), descriptions of the analyses conducted and discussion of how the analyses informed the final score interpretation metrics (e.g., bookmarking, anchor-based analyses, qualitative interviews) (For approaches used to derive Meaningful Score Differences, provide the evidence listed below.)
   a. When anchor-based analyses are used, include:
      i. Evidence that the concept assessed by each anchor variable is sufficiently related to the concept of interest of the COA
      ii. Evidence to support the appropriateness of the response options of each anchor
      iii. Evidence to support the scoring of each anchor variable
      iv. Designation of each anchor measure as primary or secondary anchor that is supported by the hypothesized relationship of the anchor to the COI of the COA
      v. Correlation coefficients assessing the strength of the relationships between the anchor variables and the COA endpoint scores
      vi. For global impression of severity (GIS) anchor variables, a table of cross-tabulation of responses at baseline and at the fixed timepoint
      vii. The target anchor change category, i.e., the amount of change on a GIS scale or the response on a global impression of change (GIC) scale that represents a meaningful benefit to patients (the amount of change should be based on patient/caregiver input)
      viii. Empirical cumulative distribution function (eCDF) curves and probability density function (PDF) curves for the COA endpoint scores by change in anchor variable (for GIS scales) and anchor variable response (for GIC scales)
      ix. Tables with distribution of COA endpoint scores by change in anchor variable response (for GIS scales) and by anchor variable response at the fixed timepoint (for GIC scales)
      x. Table of distribution of COA endpoint scores by baseline GIS response for patients who achieved the target anchor change category
      xi. Table of distribution of COA endpoint scores by baseline GIS response for all patients
   b. Qualitative interviews and/or surveys (when used)
      i. Provide description of how qualitative data are used to inform selection of MSDs

ii. If interviews occurred during concept elicitation, if there were cognitive debriefing or exit interviews, or if exit surveys were conducted, include study documents (e.g., protocol, interviewer guide, transcripts)

iii. Specify when interviews were conducted (e.g., during stand-alone study, prior to start of phase 3 trial, after completing the main portion of the study before unblinding)

c. When bookmark study is used, include the following:

i. MSD recommendations from study participants (e.g., from different rounds)

ii. Validity evidence to support the meeting material used in the study (e.g., vignettes, ordered item booklet, development of cut scores (such as the algorithm and item parameter estimates in deriving the cut scores), cut score labels development and validation)

iii. Study documents that are used in the bookmark study by the facilitator (e.g., facilitator slide deck), by the participants (e.g., vignettes used), by both facilitator and the participants (e.g., intermediate MSD results between rounds that are shown to the participants for discussion)

iv. Any documented information regarding the participants' input (e.g., to show that they understand the process), materials participants are given regarding how to participate in the bookmarking study, and participants' perspectives on the final MSD recommendation

d. When existing literature is used, include:

i. Summary of relevant existing literature and conclusions, including methodology used to determine the MSDs and conclusions (Include copies of all relevant existing literature)

ii. Justifications for how these findings generalize to the COU (e.g., description of how comparable these study samples are to the clinical study sample)

iii. Description of how any differences between patient populations affect the final proposed MSDs

e. When other approaches are used, include:

i. A rationale for the approach, including how it considers the patient voice in determining MSDs

ii. Study methodology and analysis information.

iii. A description and discussion of relevant results, including tables, figures, and other supporting evidence

f. Approach used to finalize MSD range
    i. Discussion of rates of patient misclassification and impact of patient baseline symptom severity
    ii. Final MSD range
g. Approaches used to derive Meaningful Score Regions
    i. When anchor-based analyses are used, include the following (refer to IV.C.3.a.i.1-6 above): box plots of the COA endpoint score by anchor variable response, eCDF curves, and PDF curves of the COA endpoint score by GIS score at relevant time points (e.g., baseline)
    ii. When Qualitative interviews are used, refer to Section IV.C.3.a.ii.1-2 in this document
    iii. When a bookmark approach is used, include the following: (refer to Section IV.C.3 in this document): MSR recommendations from study participants (e.g., from different rounds) and any documented information regarding the participants' input to show that they understand the process, material they are given regarding how to participate in the bookmarking study, and their perspective on the final MSR recommendation
    iv. When existing literature is used, refer to IV.C.3.a.iv.1-4 above
    v. When Other approaches are used, include a rationale for the approach, explaining how it considers the patient voice in determining MSR and a description and discussion of relevant results, including tables, figures, and other supporting evidence
    vi. The approach used to finalize MSR range should include discussion of rates of patient misclassification and the Final MSR range

3. Details and discussion of how derived final range of score interpretation metrics (i.e., MSDs, MSRs) were applied to evaluate the meaningfulness of the COA endpoint scores in clinical trial(s)
    a. Description of method(s) (e.g., average horizontal gap, vertical gap)
    b. eCDF curves of COA endpoint scores by treatment arm annotated to aid interpretation (e.g., sample size, median, reference line(s) for proposed range of thresholds)
        i. When using the average horizontal gap
            a) Plot of the horizontal gap between eCDF curves
            b) If using MSDs, a figure illustrating the observed treatment effect with corresponding 95% confidence interval and MSD values

c) If using MSRs, figure illustrating observed least squares mean with corresponding 95% confidence interval, the COA endpoint scores, and approximate MSRs

ii. When using the vertical gap

a) Plot of the vertical gap between eCDF curves with relevant score interpretation metric values

b) If using MSDs, percentage of patients in each treatment arm that achieved each MSD

c) If using MSRs, table illustrating percentages of patients by treatment arm who achieved meaningful improvement

4. Overall conclusion regarding meaningfulness of COA-based endpoint scores in the context of the trial(s)

VI. APPENDICES

Appendix A: COA Copies/Screenshots

A1: Exact Copies of COAs Discussed in Dossier (including anchors)

A2: Prior Versions of Target COA(s)

Appendix B: COA-Related Manuals and Training Materials

B1: COA User Manual

B2: COA Scoring Manual

B3: Patient Training

B4: Investigator Training

B5: Other Training

Appendix C: Item Tracking Matrix

Appendix D: Transcripts

D1: Focus Groups

D2: Open-Ended Patient Interviews

D3: Cognitive Interviews

D4: Other Transcripts

Appendix E: Psychometric Analysis Plans (PAPs) and Psychometric Analysis Report

Appendix F: Standalone/Observational Study Protocol(s) (if applicable)

Appendix G: Results of Usability Testing

Appendix H: Summary of COA-Related Regulatory History

Appendix I: Additional Results of Analyses

Appendix J: Copies of Literature Cited/Referenced