U.S. FOOD & DRUG ADMINISTRATION

PATIENT-FOCUSED DRUG DEVELOPMENT
GUIDANCE PUBLIC WORKSHOP

Methods to Identify What is Important to Patients & Select, Develop or Modify Fit-for-Purpose Clinical Outcomes Assessments

Workshop Date: October 15-16, 2018

Discussion Document for Patient-Focused Drug Development Public Workshop on Guidance 3:

SELECT, DEVELOP OR MODIFY FIT-FOR-PURPOSE CLINICAL OUTCOME ASSESSMENTS

TABLE OF CONTENTS

TABLE OF FIGURES

TABLE OF TABLES

I. INTRODUCTION

This discussion document is for the workshop for the third in a series of four methodological ***patient-focused drug development (PFDD)*** guidance documents[1] that FDA is developing to describe in a stepwise manner how stakeholders (patients, researchers, medical product developers and others) can collect and submit ***patient experience data***[2] and other relevant information from ***patients*** and ***caregivers*** for medical product[3] development and regulatory decision making.

Guidances 1 and 2 cover, respectively, the selection of patients from whom to collect information, and how to elicit information from these patients in a robust operational manner. Guidance 3 will address how to refine the list of important impacts and ***concepts*** elicited from patients, as described in Guidance 2, to develop potential study ***instruments*** (i.e., ***clinical outcome assessments*** [COA])[4] . It will discuss best practices to ensure that a COA is fit for its intended purpose in medical product development so that the effects seen in clinical trials can be interpreted and communicated as a clear ***clinical benefit*** that is meaningful to patients. Guidance 3 is primarily intended to inform and guide the work conducted by medical product developers (hereon referred to as sponsors) studying a particular investigational medical product for treatment of an identified disease with the intention of seeking medical product approval by FDA, as well as for COA developers. The document (and this discussion document for the workshop) is therefore written with the assumption that a sponsor and/or COA developer will be conducting or directing the work described and that it will be submitted for FDA review.

A. QUESTIONS FDA HAS IDENTIFIED FOR THE OCTOBER WORKSHOP

With this discussion document FDA seeks input from patient stakeholders, researchers, medical product developers, and others on how best to communicate FDA's current thinking on approaches to the selection, development or modification of a COA. Questions for readers to consider:

1. Does the Roadmap Diagram (**Figure 3**) in the Guidance 3 discussion document capture the appropriate elements to strategize for the selection and/or development of a COA for use in clinical trials? If not, what are other factors that should be considered and where should they be positioned in the diagram?

---

[1] The four guidance documents that will be developed correspond to commitments under section I.J.1 associated with PDUFA VI under Title I of the FDA Reauthorization Act of 2017. The projected timeframes for public workshops and guidance publication reflect FDA's published plan aligning the PDUFA VI commitments with some of the guidance requirements under section 3002 of the 21st Century Cures Act. https://www.fda.gov/downloads/forindustry/userfees/prescriptiondruguserfee/ucm563618.pdf

[2] The Glossary defines many of the terms used in this discussion document. Words or phrases found in the Glossary appear in bold italics at first mention.

[3] A drug, biological product, or medical device.

[4] In this document, the term "clinical outcome assessment" is interchangeable with "instrument," "tool," or "measure." A COA is defined as an assessment of a clinical outcome (i.e., an outcome that describes or reflects how an individual feels, functions or survives). Clinical outcomes can be assessed through a report by a clinician, a patient, a non-clinician observer or through a performance-based assessment.

2. Does the decision tree diagram (**Figure 6**) in the Guidance 3 discussion document capture the process to select, develop, or modify a COA sufficiently? If not, what are other factors that should be considered in this process and where should they be positioned in the diagram? Should this diagram replace the "Wheel and Spokes" diagram in the current PRO Guidance (Figure 3 in FDA PRO Guidance)?[5]

3. Important considerations are needed for special populations, such as pediatric, the cognitively impaired, rare diseases, and patients from different language and cultural groups. Does the Guidance 3 discussion document capture all the relevant special populations? What other populations should be identified for this FDA Guidance? Are there any other factors to consider when selecting, developing, and implementing COAs for these populations?

    a. What other factors need to be considered when determining a reasonable minimum age to self-report in a reliable and valid manner?
    b. What other factors need to be considered when determining a reasonable minimum level of cognitive function to self-report?
    c. How to address selection of COAs for people who move between a self-report status and inability to self-report?
    d. What are other factors and/or approaches to consider when using COAs in multinational, multicultural, and/or multiregional studies?
    e. Does the Guidance 3 discussion document appropriately present the important considerations for selection, development, and/or modification of COAs in rare diseases in sufficient detail and in a feasible manner? If not, what are other factors and/or approaches to consider?

4. Does the Guidance 3 discussion document capture the most appropriate and feasible methods to determine within-patient meaningful score changes in COA instruments? Are there any other methods to consider?

5. Are there recommendations for any changes to the definitions we include for the categories of COAs (PRO, ObsRO, ClinRO, PerfO)? Are any additional categories of COAs recommended?

    a. Digital monitoring sensors can be used for clinical outcome assessment (e.g., step counts collected via actigraphy). Please suggest approaches or methods to provide evidence of fitness for purpose (content validity, construct validity, reliability, ability to detect change) for these tools. For example, walking speed rather than step count may be most relevant and meaningful to a particular patient population.

6. FDA strives to maintain flexibility in our evaluation of evidence, taking into account feasibility and practicality. Does the discussion document appropriately describe how

[5] https://www.fda.gov/downloads/drugs/guidances/ucm193282.pdf

FDA will assess whether a COA is fit for purpose?

7. Does the discussion document present information about best practices for COA selection, development, and/or modification in a manner that can reasonably and rigorously be implemented in medical product development?

8. Is the audience described for Guidance 3 appropriate? If not, what are recommended changes?

9. How do the good measurement principles presented in this discussion document apply to PerfOs and ClinROs, and what other evidence is needed?

    a. There is existing literature related to PerfOs and ClinROs (e.g., PerfO White Paper[6] and ISPOR Task Force ClinRO paper[7]). Which principles from existing literature or other sources are important and appropriate for inclusion in FDA guidance?

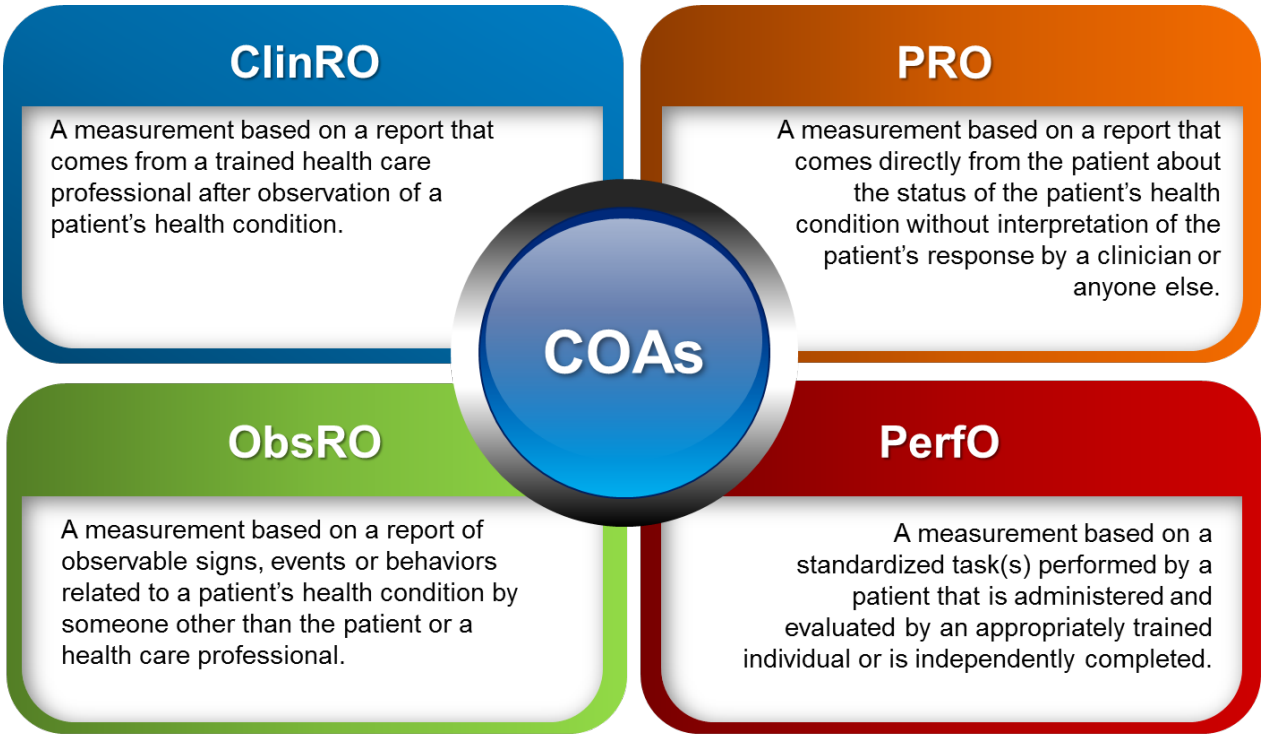
II. OVERVIEW AND SCOPE

This discussion document explains the general principles related to COAs in clinical trials. There are different types of COAs: ***patient-reported outcome*** (PRO), ***clinician-reported outcome*** (ClinRO), ***observer-reported outcome*** (ObsRO), ***performance outcome*** (PerfO) ***tools***, as well as certain COAs derived from technologies, such as mobile health technologies (e.g., activity monitors, sleep monitors) that do not fall into one of the other types of COAs. Mobile health technologies can be considered a COA depending on the ***intended use*** in a clinical trial. An overview of the COA types is provided in **Figure 1**. Specific information related to ObsRO assessments can be found in **Appendix 5**. Information related to other types of COAs will be addressed in the future.

[6] http://journals.sagepub.com/doi/pdf/10.1177/2168479018772569

[7] Powers III, JH, Patrick DL, Walton MK, et al. Clinician-Reported Outcome (ClinRO) Assessments of Treatment Benefit: Report of the ISPOR Clinical Outcome Assessment Emerging Good Practices Task Force. Value Health 2017; 20(1):2-14.

Figure 1. Overview of COA types



*There are certain types of COAs derived from mobile health technologies (e.g., activity monitors, sleep monitors) that do not fall into one of the other types of COAs.

III. BACKGROUND

FDA considers the use of ***patient input*** an important part of medical product development that can foster innovation and the availability of safe and effective medical products. Patient input can be included in not only the selection of ***clinical outcomes*** but also to ensure the appropriateness of instruments used to collect trial data. Patient input plays a critical part in medical product development by helping to ensure investigations of the effect of treatments assess outcomes that are meaningful to patients.

In instances where patient input cannot be obtained or reported reliably (e.g., young children, individuals with cognitive problems), other stakeholder input (e.g., a clinician or other trained health care professional and/or primary caregiver(s)) can provide important information regarding what is most valuable to assess in patients. As a result, information on clinical benefit or ***risk*** from the patients' perspective can be included in labeling[8] or communicated in a way that

[8] Labeling, as used in this document, refers to the information about an FDA-approved medical product intended for the health care provider to use in treating patients. See 21 CFR 201.56 and 201.57 for regulations pertaining to prescription drug (including biological drug) labeling. Section 201.56 specifically describes the need for labeling that is not false or misleading.

is accurate and not misleading. Finally, patient input can help inform ***benefit-risk assessment*** for regulatory decision making.

IV. CLINICAL OUTCOME ASSESSMENTS IN MEDICAL PRODUCT DEVELOPMENT

How can COA data impact regulatory decision making? An important aspect of medical product development is the demonstration of clinical benefit and how that ***benefit*** is measured. COAs are often primary, co-primary, or pre-specified secondary ***endpoints*** in registration trials used to support medical product approval and ***labeling claims*** or other communications regarding clinical benefit. Clinical benefit is defined as a positive clinically meaningful effect of an intervention on how an individual feels, functions, or survives. FDA uses COAs to determine whether a medical product has been shown to provide clinical benefit to patients. When clinical benefit is demonstrated, a description of that benefit can be provided in labeling or communication in terms of the concept or outcome measured (i.e., the aspect of an individual's clinical, biological, physical, functional state, or experience that the assessment is intended to capture). Tolerability of risk and safety can also be measured by COAs.

Sponsors should determine early in medical product development whether they plan to use COAs in their clinical trials and plan for early interactions with FDA to obtain feedback about their COA measurement strategy from the relevant FDA review division.

How can patient input inform selection/development of COAs? Understanding what is most important to patients can help to develop and/or select tailored COAs to adequately collect meaningful patient experience data. Patient input[9] can identify unmet medical needs and important clinical outcomes to be studied in clinical trials, including COAs. It can also provide clarity regarding disease characteristics (e.g., progression, severity, and chronicity) of the patient population to be studied. It is important to understand the disease and whether the medical product is expected to lead to improvement or to delayed deterioration in the patient's state. This information provides FDA with the opportunity to review the instrument to ensure that it is ***fit-for-purpose***.[10]

Are there approaches and methods to consider in the selection and/or development of COAs? There are different approaches and methods to develop and select COAs. Ultimately, Guidance 3 will provide a ***patient-focused*** outcome measurement approach to COA selection/development for clinical trials (hereon, referred to as the roadmap) (See **Section V**).

This roadmap approach is described in three parts:

1. Understanding the disease or condition (See **Section VA**)
2. Conceptualization of clinical benefit (See **Section VB**)

[9] In such cases in which patients are unable to report their experience; a clinician or other trained health care professional and/or primary caregiver(s), may report on patient experience if it is observable (e.g., signs of disease or condition, functioning, etc.).

[10] A conclusion that the level of validation associated with a tool is sufficient to support its context of use (e.g., population).

3. Selecting, developing, or modifying the COA (See **Section VC**)

Both qualitative and quantitative methods should be used to develop COAs. Data obtained from these methods are useful in documenting the patient experience as it provides the opportunity to capture the patient voice by allowing patients to describe their experience in their own words. Clinician/caregiver input can also confirm what aspects of the patient experience are important from their perspective. The information elicited from these approaches can give both medical product developers and FDA ideas about what are important patient-focused outcomes to measure with COAs in clinical trials.

How does FDA determine whether a COA is fit-for-purpose? To determine the adequacy of a COA, FDA focuses on whether the COA is fit-for-purpose. Some general principles to determine whether the COA is fit-for-purpose include the following:

- The COA is appropriate for its intended use (e.g., study design, patient population)
- The COA validly and reliably measures concepts that are clinically relevant and important to patients
- The COA data can be communicated in a way that is accurate, interpretable, and not misleading (i.e., well-defined)[11]

The qualities or ***measurement properties*** of a COA reviewed when determining if it is fit-for-purpose are as follows:

- ***Content validity*** (See **Section VIB**)
- ***Reliability*** (see **Section VIC.1**)
- ***Construct validity*** (see **Section VIC.2**)
- ***Ability to detect change*** (see **Section VIC.3**)

What are pathways to obtain regulatory advice on COAs?

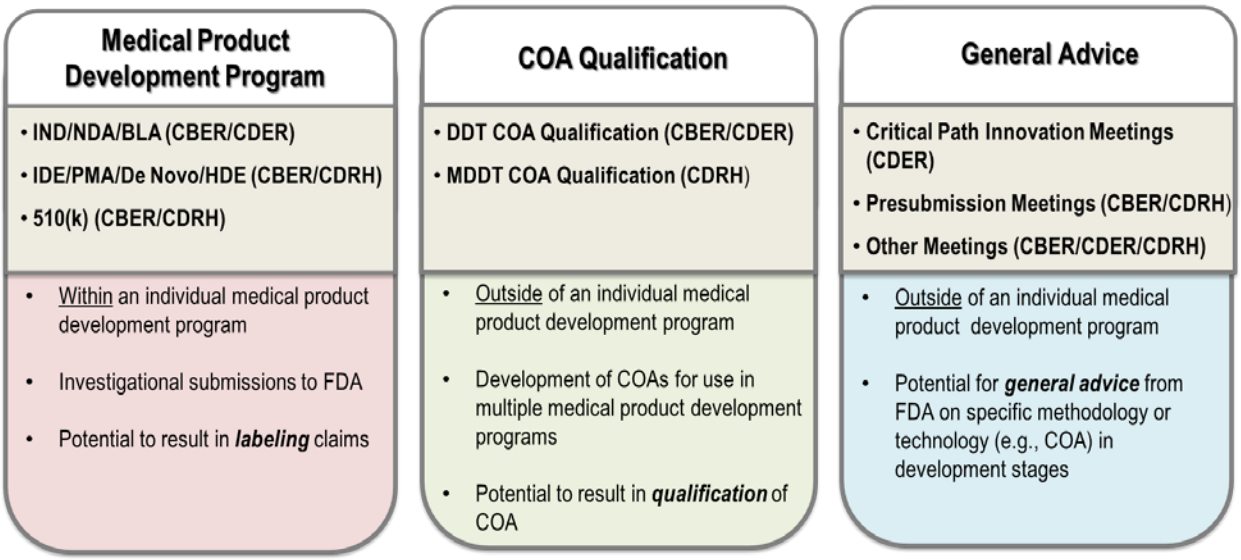
There are different pathways[12] in which sponsors may obtain advice on COAs. These pathways are illustrated in **Figure 2**.[13]

[11] If the COA is appropriately applied in medical product development.

[12] For additional information on pathways, refer to 2014 Guidance for Industry and FDA Staff—*Qualification Process for Drug Development Tools;* 2015 Guidance for Industry—*Critical Path Innovation Meetings;* 2017 Guidance for Industry, Tool Developers, and FDA Staff—*Qualification of Medical Device Development Tools*

[13] CBER may collaborate with CDER/CDRH on some pathways (e.g., COA Qualification, and Critical Path Innovation Meetings)

Figure 2. Pathways for Regulatory Advice on COAs



BLA = Biologics Licensing Application; **COA =** Clinical Outcome Assessment; **DDT =** Drug Development Tool; **HDE =** Humanitarian Device Exemptions; **IDE =** Investigational Device Exemption; **IND =** Investigational New Drug; **MDDT =** Medical Device Development Tool; **NDA =** New Drug Application; **PMA =** Pre-Market Approval

What information should be submitted to FDA for review and advice on a COA for individual medical product development programs?

- Description of intended use and endpoint
- Copy of COA(s) (i.e., current draft version, or an exact copy of the final version as it will be administered in the clinical trial)
- ***Conceptual framework*** of the COA (See **Section IVA**)
- Evidence to support content validity
- Evidence to support other measurement properties
- Scoring information for COA, including scoring interpretation

See also FDA guidance for industry regarding formal meetings with FDA (FDA, 2017a).

How can FDA use COA data beyond labeling claims?

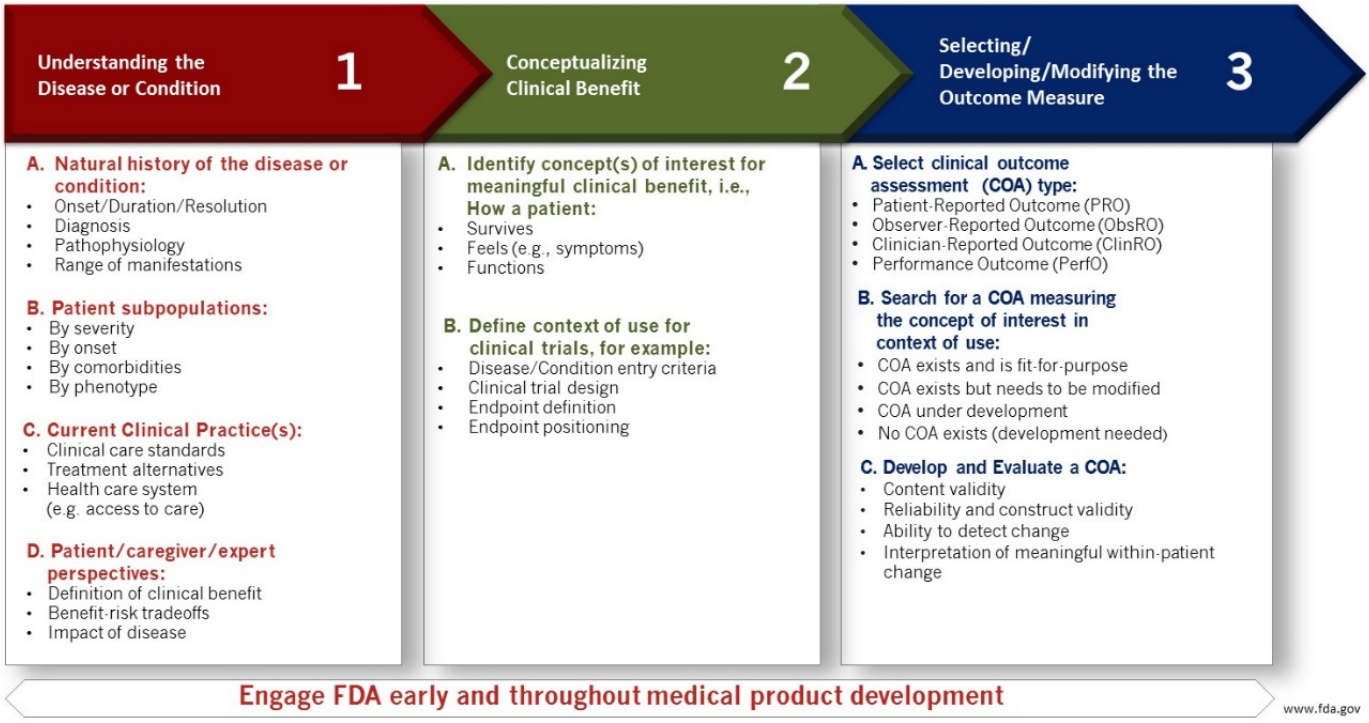
FDA generally reviews COA data as part of the totality of evidence to inform benefit-risk assessment, whether or not labeling claims are granted. Therefore, no single outcome assessment is sufficient on its own to provide the whole picture of the impact of disease and treatment on patients.

V. ROADMAP TO COA SELECTION/DEVELOPMENT FOR CLINICAL TRIALS

In approaching selection or development of a COA, it is important to have an adequate understanding of the disease under investigation and conceptualization of clinical benefit from

the targeted ***treatment effect***. Figure 3 outlines the general approach to select and/or develop/modify COAs for clinical trials.

Figure 3. Roadmap to COA Selection/Development for Clinical Trials



Note: This roadmap can also be used to conceptualize tolerability or risk.

The following subsections provide recommendations on what to consider when selecting or developing/modifying a COA.

A. Understanding the Disease or Condition

While disease understanding is critical to medical product discovery and development research, it is also critical to COA selection and development. Understanding the disease or condition encompasses knowledge of (a) disease natural history, (b) characteristics of patient subpopulations, (c) current clinical practice and therapeutic landscape and (d) patients' and caregivers' perspectives and values. Examples of each of these key elements are highlighted in **Table 1**.

Table 1. Considerations on how to use disease information for development of COA measurement strategy

Natural History of Disease

• **Information on the natural history of disease can be used to:**
  ○ Understand the clinical course of disease including:
  ○ ***Onset of disease***
  ○ ***Duration of disease***
  ○ ***Clinical presentation*** *(e.g., core signs/symptoms and disease impacts)*
  ○ ***Disease behavior*** *(e.g., waxing and waning signs/symptoms)*
  ○ ***Disease trajectory*** *(e.g., progressive, relapsing and remitting, or acute)*
  ○ ***Disease adaptations***
  ○ ***Disease subgroups*** *(e.g.,symptom heterogeneity, phenotypes, genotypes, etc.)*

Patient Subpopulations

• **Information on patient subpopulations can be used to:**
  ○ Identify patients at different stages of a disease with features that might be more measurable using a COA (e.g., asymptomatic versus symptomatic)
  ○ Consider any expected variations in experiences of patients across different subpopulations when selecting or developing COAs.

Current Clinical Practice
• **Information on how the disease is currently treated can influence clinical trial entry criteria, design, and outcome measurement**
Patient/Caregiver/Expert Input & Other Data Sources
• **Information from multiple streams (patients/caregivers /experts, literature, and other data sources) can provide comprehensive insight on aspects of the disease (e.g., symptom burden, disease impacts on daily functioning) and inform selection of a COA.**

B. Conceptualizing Clinical Benefit

Table 2 provides a list of factors to consider when conceptualizing clinical benefit (i.e., how an individual feels, functions, or survives).

Table 2. Considerations for conceptualizing clinical benefit.

Concepts of Interest
• **COAs should include clinically important concepts that define the disease in the targeted population and/or the impacts of the disease.** • **There are multiple variables that can help inform the concepts of interest:** ○ ***Patient input*** *including most common and bothersome aspects of the disease,* ○ ***Disease natural history*** ○ ***Aspect of the condition the treatment can modify*** *during the courses of a clinical trial* ○ ***Targeted labeling***
Context of Use
• **The context of use for the clinical trial should be clearly defined in order to select or develop an appropriate COA.** • **There are multiple variables that can help define the context of use, including but not limited to the following:** ○ ***Disease definition*** *(e.g., disease subtype, disease severity, history of previous treatment)* ○ ***Target population*** *(e.g., demographics, culture and language)* ○ ***Clinical practice and trial setting*** *(e.g., inpatient, outpatient, controlled/uncontrolled trial)* ○ ***Endpoint positioning*** *(e.g., primary, co-primary, secondary or exploratory)*

1. Concepts of Interest

To be able to select or develop an appropriate COA, the trial outcome concepts must be known or hypothesized based on scientific evidence. A concept is the aspect of an individual's clinical, biological, physical, or functional state, or experience that the assessment is intended to capture (or reflect). In clinical trials, a COA can be used to measure the effect of a medical product on one or more concepts. The selection of concepts appropriate for a given trial program should be informed by consultation with patients and/or caregivers, clinical, trial design, and measurement experts as well as literature review.

Example

Scenario: A sponsor is planning to develop a treatment for a neurological disease affecting the central nervous system and include PRO and ObsRO assessments in their clinical trials. After talking to numerous patients with this disease and their caregivers, patients' general complaints were related to headaches, nausea, and vomiting. Additionally, caregivers reported that patients experienced seizures and drowsiness.

What could be some potential concepts that the sponsor can include in their COAs?

- **Headaches** (PRO)
- **Nausea** (PRO)
- **Vomiting** (PRO, ObsRO)
- **Seizures** (ObsRO)
- **Drowsiness** (ObsRO)

Sponsors should factor in the relevance and importance of concepts to the ***target population***, whether core disease-related concepts (e.g., ***signs*** and ***symptoms***) or disease impacts, and how it will inform clinical benefit. FDA recommends measuring, at the minimum, core disease-related concepts. When measuring disease impacts, FDA recommends targeting disease impacts that result from the core disease-related concepts.

Examples

Scenario 1: Patients with nocturia

What could be some potential core disease-related concepts to measure?

- **Frequent urination after going to sleep for the night**

What could be some potential impacts of core disease-related concepts to measure?

- **Daytime functioning (e.g., tiredness, concentration)**
- **Sleep disturbance**

Scenario 2: Patients with Hemophilia A

What could be some potential core disease-related concepts to measure?

- Frequency of bleeding episodes
- Pain (acute/chronic)
- Joint damage

What could be some potential impacts of core disease-related concepts to measure?

- Physical functioning (e.g. impact of pain and joint damage on ability to perform daily activities)

Concepts related to treatment safety, tolerability, or burden may also be measured by COAs, if those concepts represent symptoms or signs that can be reported by patients, caregivers, or clinicians. When assessing treatment safety, tolerability, or burden with a PRO, sponsors should select symptomatic adverse events and other topics in an unbiased manner. FDA recommends that that sponsor provide a strong rationale to support the selection of symptomatic adverse events and other topics that will be assessed, including support from nonclinical and clinical data where available. Symptomatic adverse events should be captured separately from disease-related signs, symptoms, and impacts, where possible.

2. Context of Use

The ***context of use*** is a statement that fully and clearly describes the way the COA is to be used and the medical product development-related purpose of its use**.** FDA recommends that sponsors consider the potential design and logistics of a clinical trial when selecting or developing a COA.

Factors to consider when establishing clinical trial objectives may include, but are not limited to:

- Trial phase (exploratory or confirmatory trials)
- Expected clinical benefit and risks of the medical product
- Targeted labeling claim(s) or communication
- Comparator/control
- Dose frequency and duration
- Route of administration
- Patient population
- Disease or condition
- Endpoint positioning
- Endpoint definition
- Timing of assessment(s)
- Analysis plan
- Missing data imputation algorithm (including missing data plan and how much the COA ***score*** can handle missing ***items*** [i.e., questions or ***tasks*** included in the COA])
- Methods for interpretation of trial results

Examples

Scenario 1: The same sponsor planning to develop a treatment for a neurological disease has decided to target their product to treat a specific phenotype (referred to as phenotype A) of this disease in adult patients (≥18 years). The sponsor plans to use a neurologic-specific COA in an exploratory clinical study.

What is the context of use?
The context of use for the planned neurologic-specific COA is an exploratory study in adults (≥18 years) with a diagnosis of phenotype A.

Scenario 2: A sponsor is planning to develop a treatment for a hereditary immunological (HI) disease. There are currently treatments approved for two subtypes of this disease; the sponsor would like to develop a treatment for a third subtype (referred to as HI Type 3) and would therefore like to develop a tool to measure symptoms associated with this third subtype. The sponsor plans to conduct an initial exploratory study.

What is the context of use?
The context of use for the planned symptom measurement tool is an exploratory study in adult and adolescent patients (≥ 12 years) with a diagnosis of HI Type 3.

The context of use also takes into account how the COA will be used as a trial endpoint. Sponsors should consider how the COA will be incorporated into the planned trial endpoint(s) and statistical analysis plan. An ***endpoint*** is a precisely defined variable (e.g., COA score) intended to reflect an outcome of interest that is statistically analyzed to address a particular research question. A precise definition of an endpoint(s) typically specifies the type of assessment(s) made, the timing of those assessments, the assessment(s) used, and possibly other information, as applicable, such as how multiple assessments within an individual are to be combined.[14]

It is important to note that the COA is not the endpoint. The COA is the instrument that is used to evaluate the intended outcome.

Example

Scenario: A sponsor is planning to use an Itch tool to assess the itch intensity of adolescent patients (≥12 years) with atopic dermatitis in a 12-month clinical trial. The Itch tool is being administered daily to patients in an electronic mode (i.e., electronic tablet).

What is the concept of interest?
- Itch intensity

What is the COA?
- Itch tool (PRO)

What is the variable?
- Itch tool score

What could be a potential endpoint?
- Change from baseline to Week 12 in the weekly mean of the daily itching score

[14] Definition of an endpoint was retrieved from the BEST (Biomarkers, EndpointS, and other Tools) Resource Glossary Website: http://www.ncbi.nlm.nih.gov/books/NBK338448/

Factors to consider when developing an endpoint from a COA:

- Targeted labeling claim
- Context of use
- Concept of interest
- Clinical benefit/risk

More details on development and selection of endpoints will be discussed in Guidance 4. **Figures 4** and **5** show examples of how COAs may be positioned in an endpoint hierarchy. In **Figure 4**, COAs are used as supportive endpoints with a physiologic measure as the primary endpoint intended to support an indication for treatment of a disease. In this case, the clinical trial would need to succeed on the physiologic endpoint before success could be attained on the supportive endpoints. In **Figure 5**, a COA is the primary clinical trial endpoint intended to support an indication for the treatment of symptoms associated with Disease Y and the physical performance and limitation measures would be the supportive endpoints.

Figure 4. Example of endpoint positioning (Treatment of Disease X)

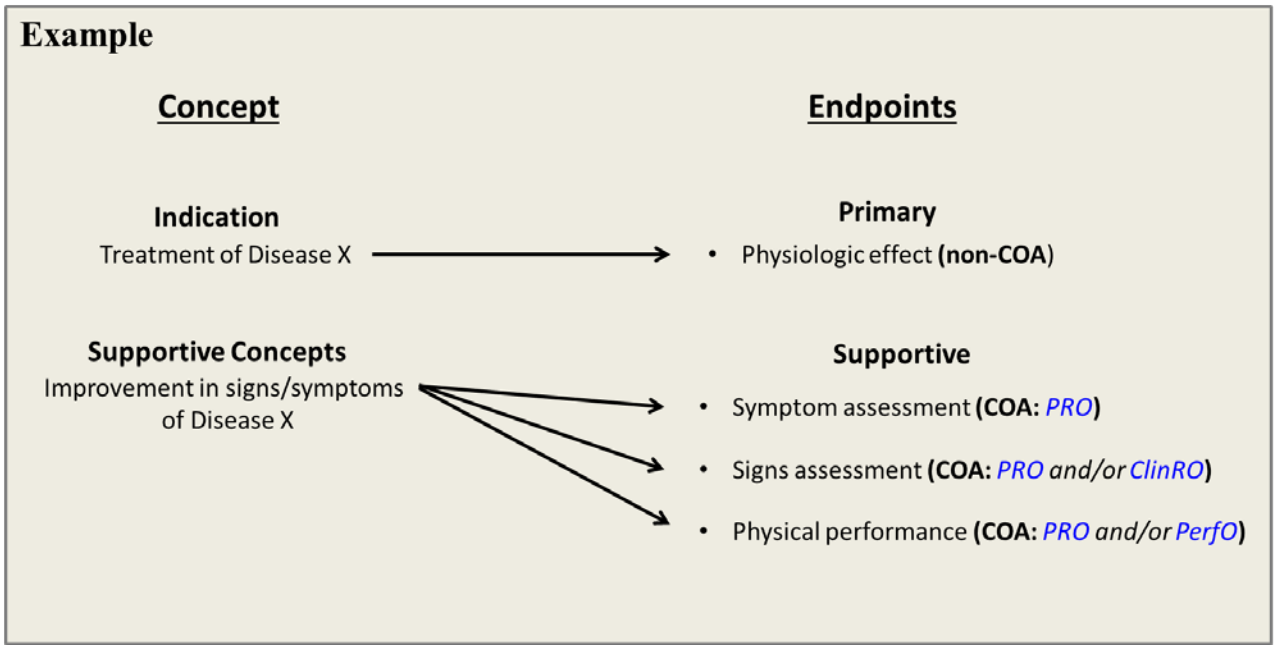
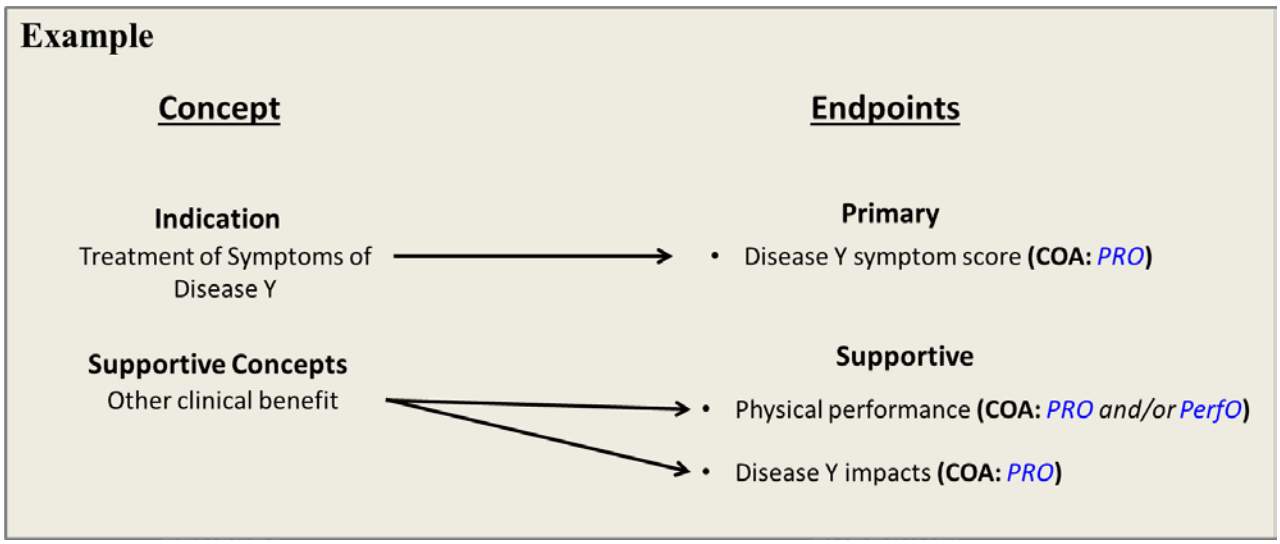


Figure 5. Example of endpoint positioning (Treatment of Symptoms associated with Disease Y)



The adequacy of a COA depends on its role and relationships with planned clinical trial endpoints. For example, regarding endpoint positioning, a high degree of certainty and ***validation*** is particularly important for a COA to be used in the context of a primary or co-primary endpoint.

FDA recommends that sponsors carefully consider the order of COAs in the endpoint hierarchy. Sponsors should provide a proposed endpoint hierarchy for discussion with the FDA early in medical product development, with the understanding that it may evolve.

Additionally, sponsors should specify potential labeling claims and facilitate communication with the FDA about the specific clinical trials designed to assess the planned concepts (FDA, 2007).

C. Selecting/Developing a COA

The process of selecting or developing a COA for use in a medical product development program depends on having adequately characterized the disease or condition, defined the target context of use, and conceptualized a concept of interest that represents clinical benefit (see **Sections IVA** and **IVB** above).

Sponsors should determine early in medical product development whether they plan to use COAs in their clinical trials, as well as engage the FDA in a discussion about their COA measurement strategy.

Table 3 provides a list of factors to consider when selecting or developing a COA.

Table 3. Considerations when selecting or developing a COA

COA Type
• Choice of a COA type is dependent on the targeted concepts, context of use (including patient population), and planned trial endpoint(s). • The observability of the concept is a key determinant in selection of the COA type. Unobservable concepts are generally feelings and sensations (i.e., symptoms), whereas observable concepts could be signs, events, behaviors, or verbal expressions by the patient.
Existing COAs
• Consider resources to leverage existing literature and data, and select instruments to use "as is" or for modification or adaptation. • If there are no existing COAs for the planned context of use, a new COA can be developed.
COA Measurement Properties
• The adequacy of any COA, whether existing, modified, or newly developed, as a measure to support medical product labeling claims depends on whether its characteristics, conceptual framework, content validity, and other measurement properties are fit-for-purpose. • Evaluate and document the development history of the COA instrument.

We encourage instrument developers to make their instruments and related development history available and accessible publicly.

1. Selection of COA Type

For symptomatic conditions or conditions associated with functional impairment, PRO assessments are generally used as they provide direct evidence of how patients feel and function. However, when patients cannot provide self-report, reports based on observation of signs, events, or behaviors that are reflective of how the patient feels or functions are often useful (e.g., ClinRO, ObsRO). In the case that clinical judgment is required to interpret an observation, a ClinRO assessment should be used. Proxy-reported outcome measures (i.e., reports by someone who is not the patient responding as if that person were the patient) are discouraged for measuring concepts that are only known by the patients (e.g., symptoms). For additional information on proxy reports refer to **Appendix 5**. PerfO measures may be used to assess patient functioning (e.g., physical, cognitive, or perceptual/sensory function) in a standardized way with one or a series of standardized tasks.

Examples

Scenario 1: A sponsor is interested in developing a product to relieve pain for earaches in adolescents and adults.

What would be the most appropriate assessment for this context of use?
Because only patients are able to directly report on their level of pain (i.e., pain intensity) a PRO assessment would be the most appropriate tool to assess pain intensity (*unobservable concept*).

Scenario 2: This sponsor is now interested in studying their product in infants.

What would be the most appropriate assessment for this context of use?
Because infants cannot self-report, use of an ObsRO would be appropriate. Although observers cannot validly report an infant's pain intensity, they can report infant behavior thought to be caused by pain, such as crying or tugging on their ear (*observable concept*). An ObsRO or ClinRO assessment could be a tool of choice to measure these concepts.

1. Evaluation and Documentation of COA Development History

FDA will review the documentation of COA development and testing throughout tool and medical product development, beginning early in development (e.g., discovery, invention), and in conjunction with pivotal trial results to determine whether COA communications (e.g., labeling claims, promotional materials, etc.) are substantiated. See **Section VI** for the instrument characteristics and measurement properties evaluated by FDA.

When feasible, FDA recommends that a COA's measurement properties be established before enrollment begins for confirmatory clinical trials. It is at the sponsor's risk to proceed with using a COA in pivotal trials without evaluating its measurement properties (i.e., content validity, reliability, construct validity, and ability to detect meaningful change). Typically, in early clinical trials, a number of COAs may be piloted for exploratory purposes. Exploratory studies (e.g., in early medical product development) are an opportune time to examine a COA's measurement properties and performance prior to initiating confirmatory clinical trials; stand-alone non-interventional studies are another option. The goal of pilot testing COAs is to select and/or refine a COA to be carried forward into registration trials to establish product effectiveness.

2. Search Strategy for COA

The sponsor should consider Columns 1 to 3 in the Roadmap (**Figure 3**) when developing their COA measurement strategy. The key consideration when developing a COA measurement strategy, including selecting, developing or modifying COAs within medical product development, is to start early. The process used to develop or modify a COA is applicable to both pre- and post-market approval(s).

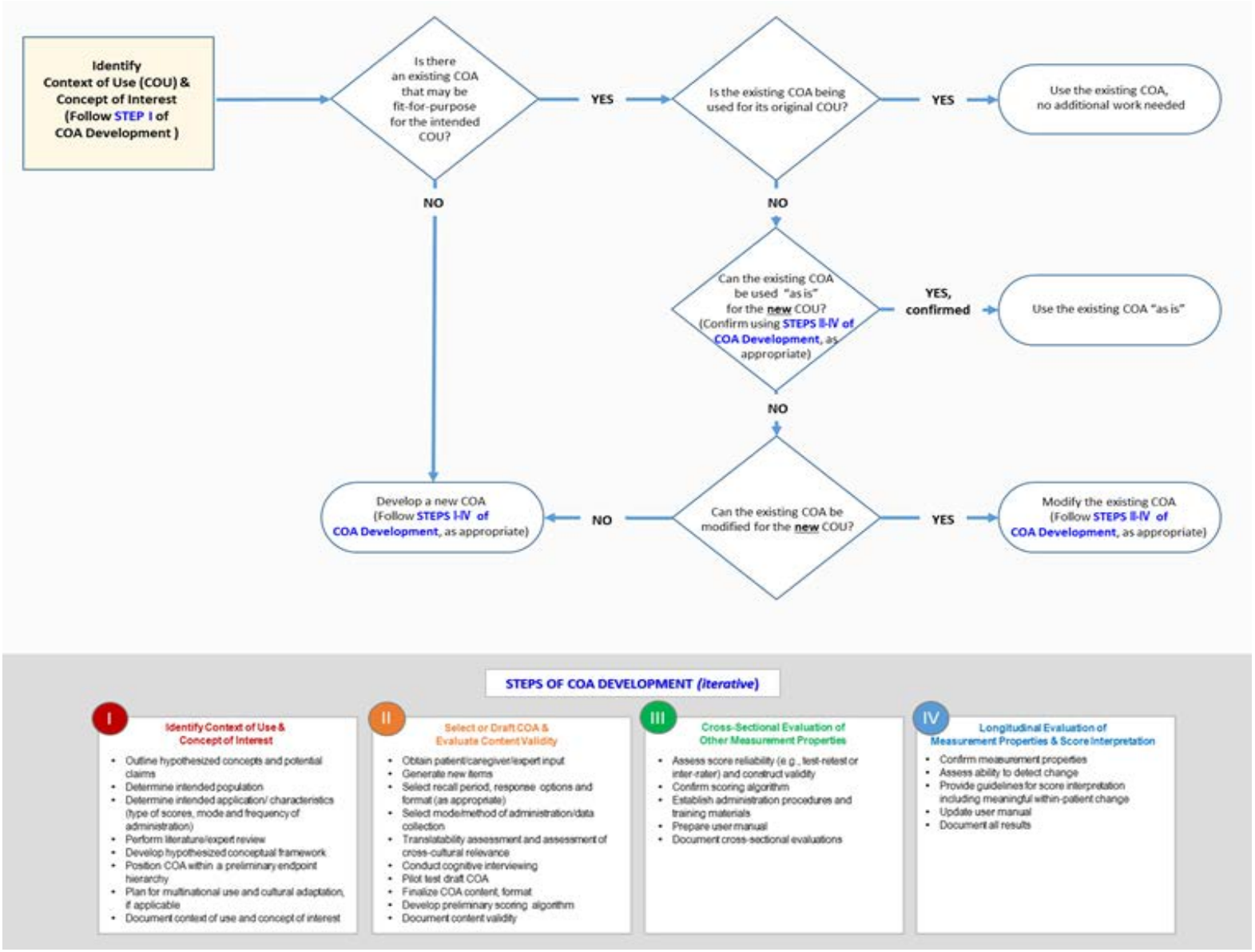
The following factors should be considered when planning to use a COA in support of a labeling claim or in other aspects of regulatory decision making:

- **Availability of existing instruments**. Sponsors may leverage and build upon existing instruments, literature, and data to fit the specific needs of the research question(s). FDA encourages collaboration among sponsors and instrument developers/researchers. It is important to note that some instruments used widely in clinical practice might not be fit-for-purpose for regulatory trials as they may not be designed in a way that would make it likely to be sensitive in detecting treatment effects and discriminating between treatment and placebo arms' scores (e.g., well-defined concepts; clear recall period; distinct and non-overlapping response options representing clinically meaningful gradations; standardized user manual/training materials).

- **Adequacy of COA.** Sponsors should determine the adequacy of the existing COA to measure the concepts of interest (e.g., instrument is fit-for-purpose for the context of the medical product development program; instrument has adequate measurement properties, etc.).

Figure 6 outlines the process of how to determine whether to use an existing instrument, modify an instrument, or develop a new instrument. This figure also summarizes the iterative process used in developing and/or modifying a COA for use in clinical trials for medical products. FDA review of the developmental process documentation is discussed in more detail in **Section VI.**

Figure 6. Process to select, develop or modify a COA



i. Selecting or modifying an existing instrument and determining whether it is fit-for-purpose

As stated previously, before embarking on developing a new instrument, it is important to determine whether an existing instrument can be used "as is" or modified.

An existing COA may be used in the following ways:

- "As is" for the intended population and context of use in which it was initially developed;
- "As is" for a new context of use (e.g., population); or
- Modified for a new context of use.

In all cases, FDA will evaluate the measurement properties to determine whether the instrument is fit-for-purpose.

When modifying an existing instrument, all the steps that are necessary in developing a new instrument may not be applicable. The type of evidence needed to support modifications to an instrument will depend on the type of changes. Sponsors should provide the following evidence for a modified instrument for FDA review:

- Evidence that the modified instrument's instructions, concepts (e.g., core signs/symptoms/functioning), and items are relevant, meaningful, appropriate, and comprehensive relative to its intended measurement concept, intended use, and to the targeted patient population.

- Evidence to confirm the modified instrument's adequacy, which may include:
    - Published literature or previous ***qualitative research***
    - Additional qualitative research may be recommended if the instrument will be used in a significantly different patient population (e.g., a different disease or age group) and sufficient evidence is not available to support content relevance to the target population
    - Additional analyses may be recommended to evaluate the instrument's measurement properties within the new population.

  (This additional research may minimize the risk that the instrument may not perform adequately in a clinical trial.)

The following are some examples of instrument changes that may alter the way respondents respond to the same set of items:

- Changing the timing of, or procedures for, instrument administration within the clinic visit
- Changing the application to a different setting, population, or condition
- Changing the order of items, item wording, response options, or ***recall period*** or adding to or deleting portions of an instrument
- Changing the type of instructions or the placement of instructions within the instrument
- Changing an instrument from paper to electronic format

ii. Developing a new COA

If an existing instrument cannot be used "as is" or modified, then a new COA may need to be developed.

VI. EVALUATION OF A CLINICAL OUTCOME ASSESSMENT

Although there are evidentiary standards that are used to determine whether a COA is adequate for use in clinical trials, FDA maintains flexibility in our evaluation of evidence, taking into account feasibility and practicality. The goal is to ensure that the COA is fit-for-purpose.

Characteristics of COAs that are reviewed by the FDA within the medical development program include, but are not limited to, the following:

- Conceptual framework of the instrument
  - Concepts being measured
- Evidence of content validity
  - Medical condition for intended use
  - Population for intended use
  - Mode of data collection (e.g., electronic)
  - Administration type (e.g., self-administration)
  - Number of items
  - Response options
  - Recall period
  - Scoring, including weighting of items or ***domains***
  - Formatting (e.g., bold text, underlined text, font size, eCOA screen presentation etc.)
  - Respondent burden
  - Translation or cultural adaptation availability
- Evidence of Other Measurement Properties

Appendix 1 lists the type of COA information sponsors should provide to the FDA to facilitate instrument and endpoint review. It is preferable that this information is submitted early in the IND process for feedback prior to initiation of pivotal trials. Requests for FDA input should be addressed to the review division responsible for the medical product in question.

Evaluation of whether a COA is fit-for-purpose for assessing safety/tolerability depends on the concept of measurement and context of use. For example, the content of a COA assessing safety/tolerability will need to be appropriate and relevant for the medical product under investigation. In addition, if the COA assessing safety/tolerability will be used in the context of supporting comparative safety claims, the methods for analysis and evaluation of the COA will be similar to those used to support efficacy claims.

A. Conceptual Framework

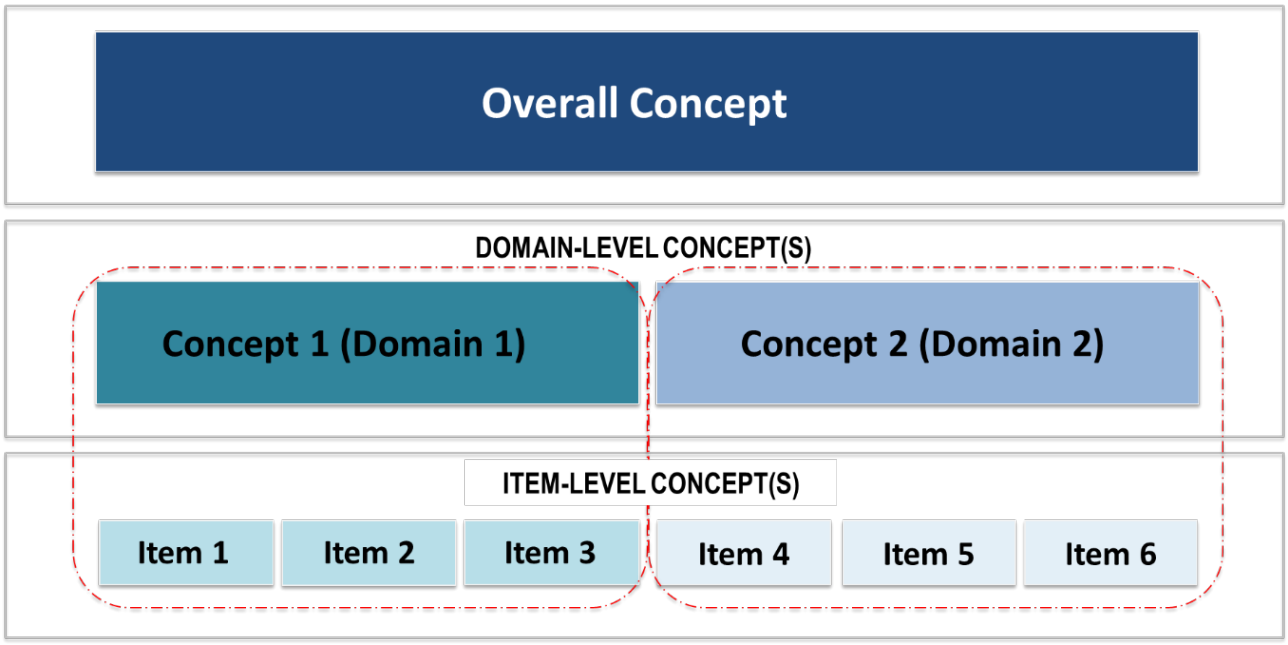
The conceptual framework explicitly defines the concepts measured by the instrument in a diagram that presents a description of the relationships between items (i.e., questions or tasks

included in the COA), domains (sub-concepts), and concepts measured and the scores produced by a COA. Throughout medical product development, FDA can review how individual items are related, how items are related to a domain, and how multiple domains may be related to each other. The conceptual framework informs the Agency regarding the hypothesized scoring of the COA and whether there will be one total (overall) score or separate domain scores (see **Section VIB.11**). FDA may request item-level or domain-level analyses.

The conceptual framework of a COA may evolve and should be confirmed over the course of its development as a sponsor gathers empiric evidence to support item selection and scores. When used in a clinical trial, the COA's conceptual framework should again be confirmed by the observed relationships among items and domains.

The diagram in **Figure 7** depicts a generic example of a conceptual framework of a COA where Domain 1, Domain 2, and Overall Concept each represent related but separate concepts. Items in this diagram are aggregated into domains. The final framework is derived and confirmed by measurement property testing.

Figure 7. Diagram of the Conceptual Framework of a COA



B. Evidence of Content Validity

Content validity is the extent to which the COA measures the concept of interest including evidence that the items and domains are appropriate and comprehensive relative to its intended measurement concept(s), population, and use. The adequacy of a COA's content validity has direct impact on evaluation of the accuracy of a medical product's labeling claims based on that COA. Content validity should be supported by evidence obtained from qualitative studies (e.g., one-on-one interviews, focus groups, or consensus panels, etc.), quantitative studies (e.g., descriptive statistics and other measurement properties), and/or published literature. The content

validity of an instrument should be established prior to evaluating its other measurement properties.

Sample size for qualitative and quantitative studies for establishing content validity should be discussed with FDA. FDA cannot provide specific recommendations regarding the number of individual patient interviews or focus groups for establishing content validity. The sample size for these studies depends on the completeness of the information obtained from analysis of the transcripts. For more complex concepts, a greater the number of patients may be needed in qualitative studies to adequately understand that concept and how it varies across the target population. Generally, the number of patients is not as critical as interview quality and patient diversity included in the sample in relation to intended clinical trial population characteristics.

Evidence of other types of validity (e.g., ***construct validity***) or reliability (e.g., consistent scores) will not overcome problems with content validity because instrument adequacy is evaluated based on whether it appropriately measures the concept it is supposed to measure.

Examples of information that should be submitted to establish content validity include the following:

- Literature review and/or publications
- Documentation of expert input
- Qualitative study protocols and interview guides for focus group or patient/caregiver interviews
- Chronology of events for item generation, modification, and finalization (***item tracking matrix***; evidence of ***concept saturation***)
- Qualitative study summary with evidence to support item relevance, item stems and response options, and recall period
- Qualitative support for meaningful change
- Quantitative study summary with evidence to support item retention and scoring
- Transcripts (if available)

1. Intended population

Using documentation of the process described in **Figure 6** and of the measurement properties (see **Section VIC**), FDA, when conducting its review of the submitted data, will compare the target population studied in the instrument development process to the population enrolled in the clinical trial to determine whether the instrument is applicable for that population.

Specific measurement considerations posed by specific patient populations, such as pediatric, cognitively impaired, or seriously ill patients are discussed in **Section IX**.

Without adequate documentation of relevant stakeholder input (patient/caregiver/clinician), a COA's content validity is likely to be questioned.

2. Concept Elicitation

Input should be elicited from relevant stakeholders (patients/caregivers/clinicians) to inform COA development to identify the core concepts to target for potential inclusion in the COA. In the development of a COA, the relevant stakeholders should be queried about important aspects of the disease or condition through one-on-one interviews and/or focus groups. This process is referred to as ***concept elicitation***.

Concept elicitation should occur in a wide range of patients with the condition of interest and/or other informants such as caregivers and clinicians to represent variations in severity and in demographic characteristics such as age, sex, ethnicity, education, and language groups in accordance with the anticipated clinical trial design to obtain representative input from the underlying target patient population.

FDA will review documentation of relevant stakeholder input to determine that concept saturation has been reached for the core concepts to include in the COA. Concept saturation is reached at the point when no new relevant and important information emerges and collecting additional data will not likely contribute additional relevant core concepts to potentially include in the COA. FDA recognized this may be difficult to achieve in certain populations (e.g., rare diseases).

Refer to the discussion document for the PFDD Workshop on Guidance 2 for additional details on concept elicitation.

3. Item (question or task) and Content Generation

Item generation includes drafting the content of items (i.e., questions or tasks included in the COA), and evaluation of the completeness of item coverage.

Item generation may be based on a combination of literature review and input from important stakeholders (i.e., patients/caregivers/clinicians), for example findings from the concept elicitation interviews. It may be useful to have the target patient population and other stakeholders involved in the initial drafting of the COA content.

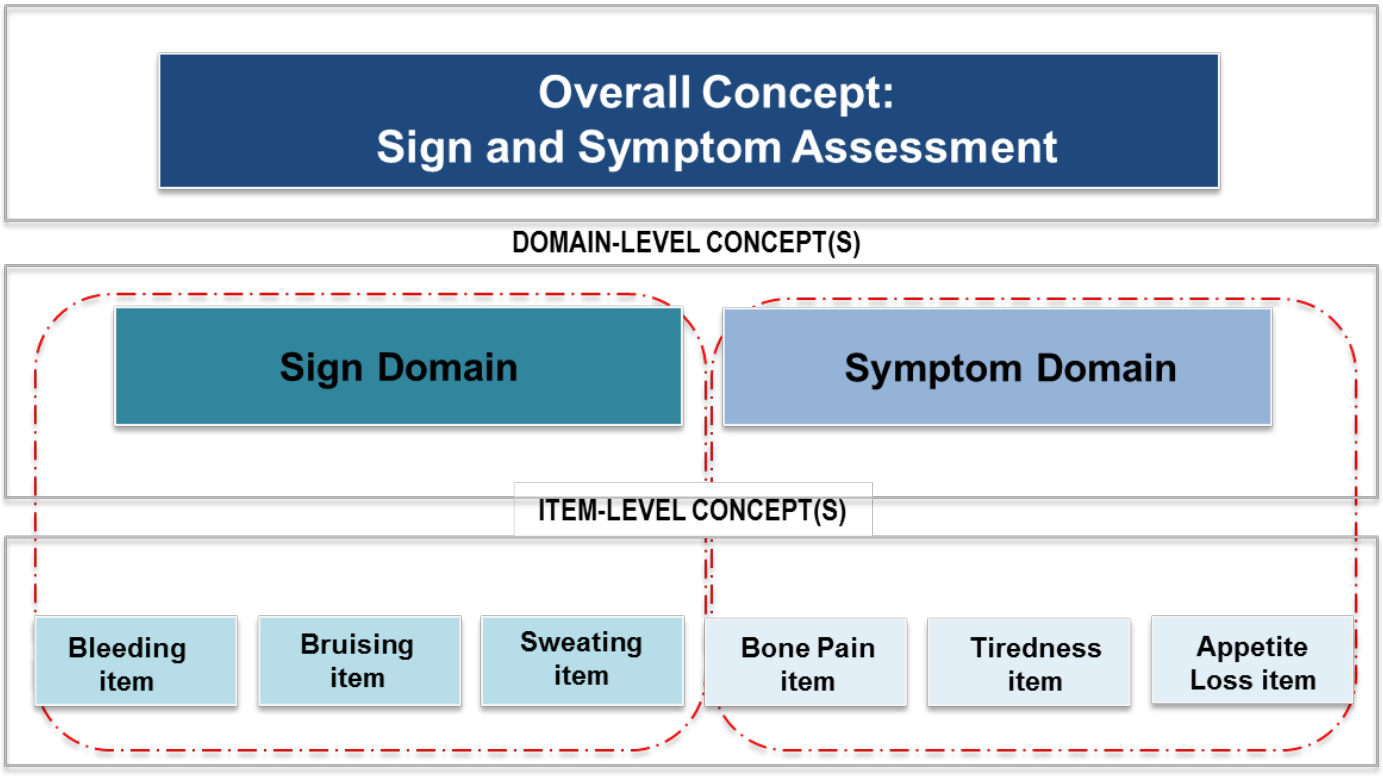
Some key considerations for generating items:

- Design items that are interpreted and understood well by participants (e.g., pilot-test questions)
- Avoid using items that ask two or more items at once (i.e., multi-barreled items)
- Avoid using items that ask patients to respond or perform hypothetically
- Items should be relevant to most of the patients in the clinical trial
- Items should measure the relevant attribute (e.g., intensity, frequency). of the concepts that are most meaningful to patients.

Once the questions or tasks are drafted, they should be tested in ***cognitive interviews*** to ensure interpretation and understanding of the instrument.

Example

What is a well- defined concept?
The concepts are clear and unambiguous. For example, below is a conceptual framework of a hypothetical signs and symptoms instrument that has two domains. The concept of signs and symptoms are explanatory and it is clear that the sign domain is measuring disease signs and the symptom domain is measuring disease symptoms.



762

Example

Scenario: A sponsor plans to use ClinRO and PerfO assessments to measure upper arm mobility in upper limb spasticity. The ClinRO includes items for which the clinician is rating lower limb mobility (i.e., leg functioning) and the PerfO includes activities that assess lower leg movement which is not relevant to upper limb spasticity.

What is the problem with including activities that are not relevant to the target population?
It would be disadvantageous to use a measure with items that include activities irrelevant to the target population. Doing so would miss the opportunity to assess a symptom or impact of importance to patients and may also yield a *bias toward the null*, or a tendency to show no effect of treatment, even if the treatment were effective. In such cases, a negative response (or indication of little to no activity) is not useful.

4. Cognitive Interviews

Understanding of the COA can be evaluated through cognitive interviewing with relevant stakeholders.

Respondent understanding of the COA (initial and drafts) should be tested for the following, where applicable:

- Item wording
- Instructions
- Recall period
- Response options
- Readability
- Concepts included in the conceptual framework are confirmed

Similar to concept elicitation, the COA should be tested in a representative population of relevant stakeholders.

Cognitive interviews should be carried out as an iterative process in which important stakeholders provide feedback and the content is revised and another round(s) of cognitive interviews is conducted until no further changes are necessary. After content validity based on qualitative research has been documented, the COA is ready to undergo an assessment of the other measurement properties.

FDA will review documentation of relevant stakeholder input to determine whether patients/caregivers/clinicians interpret and understand how to complete the instrument as intended.

5. Data Collection Mode and Type of COA Administration

Types of COA administration can include self-administration, interviewer-administration, clinician-administration, and/or instructor-administration. Data collection modes can include paper-based, electronic-based, and telephone-based. FDA intends to review the comparability of data obtained when using multiple data collection modes or types of administration within a single clinical trial to determine whether the treatment effect varies by modes or types used. For modes of data collection that do not include a date and time stamp (e.g., paper), FDA would include as part of its regulatory review, the clinical trial protocol to determine what steps were taken to ensure that patients make entries according to the clinical trial design and not, for example, just before a clinic visit when their responses will be collected.

6. Language Translation and Cultural Adaptation

Because many development programs are multinational, application of COAs to multiple cultures or languages is common in clinical trials. Language translation and cultural adaptation of the COA (including instructions, items/domains, and response options) for multinational studies is strongly recommended. A translatability assessment should be considered early in instrument development to avoid any problematic issues (e.g., irrelevant or inappropriate items,

different content and/or meaning of questions), as translation and cultural adaptation of outcome assessments can affect efficacy findings. It is important to ensure that efficacy assessments are standardized across sites.

Translation and cultural adaptation includes conducting qualitative work on the COA within all languages and cultures in which the trial will be conducted. FDA refers sponsors to the ISPOR principles for the translation and cultural validation process (Wild et al., 2005).

Regardless of whether the instrument was developed concurrently in multiple cultures or languages or whether a fully developed instrument was adapted or translated to new cultures or languages, FDA recommends that sponsors provide evidence that the content validity and other measurement properties are adequately similar between all versions used in the clinical trial. FDA would include as part of its review the process used to translate and culturally adapt the instrument for populations that will use them in the trial.

7. Recall Period (if applicable)

Sponsors should provide support for the rationale and appropriateness of the recall period for a COA. The use of a specified recall period (e.g., past 24 hours or past 7 days) helps standardize reporting by instructing the respondent to recall and report over a defined period. The recall period should be long enough to capture the event or experience of interest, but not so long that the respondent is unable to adequately recall the information, because this can lead to measurement error and potentially a limitation to the responsiveness or sensitivity of the treatment effect.

Some key considerations for selection of a recall period:

- Specify a period of time (e.g., past 24 hours) for items in a COA, where appropriate.
- Consider the respondent's ability to accurately recall the information requested within the period of time specified.
- Select a recall period based on the instrument's purpose and intended use; the variability, duration, frequency, and intensity of the concept measured; the disease or condition's characteristics; and the study treatment.
- For fluctuating signs/symptoms/impacts (e.g., episodic condition), an event log (capturing events as they occur) or a 24-hour recall period may be more appropriate.
- A longer recall period (7-days or more) may be appropriate where day-to-day fluctuations in signs/symptoms/impacts are not expected to occur (e.g., chronic conditions) or when measuring salient events that occur relatively infrequently.
- If a detailed recall of experience over a period of time is necessary, select appropriate methods and techniques to enhance the validity and reliability of data (e.g., asking patients to respond based on their worst (or best) experience over the recall period or make use of a diary for data collection).

Examples

Scenario: A sponsor plans to develop a new drug for the treatment of major depressive disorder. The sponsor intends for this new drug to have a faster onset of treatment effect than existing treatments to show improvement of major depressive disorder symptoms in a shorter period of time. The sponsor plans to develop a PRO instrument to assess core signs of major depressive disorder but is unsure what recall period to use.

What are some considerations for selecting a recall period for an instrument to be used with rapid-acting treatments?

The sponsor should consider the following when selecting a recall period:

- Timing of treatment effect
- A momentary ("right now") recall may be more relevant to capture meaningful information about clinical benefit and minimizes recall error. A lengthy recall period (e.g., 7 days) may miss capturing some important information related to the patient's experience.
- The recall period of the PRO instrument should correspond to the frequency of assessments

Scenario: A sponsor is developing a COA measurement strategy to evaluate physical functioning in patients with a debilitating motor neuron disease. The sponsor plans to use both PerfO and ObsRO tools. The selected ObsRO tool has a daily recall period for the observer to report on activities completed by the patient.

What are some of the challenge(s) using a daily recall period in this scenario and potential ways to overcome some of these limitations?

While items with shorter recall periods are generally preferable, in this scenario, the observer may have difficulty responding to items asking about activities that the patient may do infrequently (e.g., walking a long distance, climbing several flights of stairs). For these types of items, a longer recall period (e.g., over the past week) may be necessary to correspond with the frequency of occurrence.

Scenario: Asthma is a variable disease in which symptoms are episodic and may fluctuate within a day, as well as from day-to-day. A new asthma PRO instrument measure has been developed to assess asthma symptoms over a 24-hour period using an end-of-day diary. The sponsor plans to administer the PRO instrument daily in their clinical trials to capture the patient's experience.

What are some considerations when reviewing the appropriateness of the proposed 24-hour recall period and timing of diary administration for a disease with symptoms that may fluctuate throughout the day, such as asthma?

- Timing of when symptoms may be worse during the 24-hour period (morning versus

night)
- Variability in symptom occurrence during the 24-hour period
- Patients' ability to recall their symptom experience accurately
- Exposure to triggers
- Impact of symptoms during the day
- Comparing appropriateness of daily assessment versus twice-daily assessment

8. Response Options

Sponsors should consider the appropriateness of the type of ***response scale*** based on their study purposes (e.g., PerfO, ObsRO, PRO, ClinRO). Response scales can be tested within the cognitive interviews including potentially card sorting exercises. Card sorting is a research method for uncovering how people understand and classify information (e.g., sorting response options and/or pictures or photographs in a way that makes sense to them (e.g., order of increasing severity). This would test whether the respondent can order the response options as intended.

Example

Scenario: A sponsor has developed a treatment for a dermatological condition most commonly seen in pediatrics, which presents with an unbearable itch and skin scaling. The sponsor plans to develop both ClinRO and PRO tools. The sponsor plans to use an accompanying photoguide with a verbal rating scale with the ClinRO tool. For the PRO tool, the sponsor plans to use a pictorial scale to assess itch intensity in young children.

What are some ways to test the response options for these assessments?
One potential approach to evaluate whether the response options are appropriate for either assessment is through a card sorting exercise.

Using an example with a photoguide for a ClinRO assessment, clinicians would be given shuffled photographs of different severity levels so they could sort the photographs into a classification system (i.e., sort photographs into an order of increasing severity), as well as provide input on whether the photographs are clinically appropriate and representative of the range of different scaling severity levels.

With the use of a pictorial scale for the PRO assessment, children would be provided with pictures representing the different response options for the concept being measured and sort and order them into categories that make sense to them (e.g., ordering the pictures (faces and/or cartoon person) in what would illustrate not itchy at all to very itchy.

Appendix 2 describes some of the various types of item response options that are typically seen in COAs (i.e., PRO, ObsRO, and ClinRO instruments).

Key considerations for item response options:

- Response options should be understood by respondents according to what is intended and should be grammatically and logically consistent with the item stem.
- Each response option should be distinct and non-overlapping and should represent a clinically meaningful gradation of disease (e.g., patients may not distinguish between *intense* and *severe* if both are offered as response choices to describe their pain).
- Wording used in responses is clear and appropriate (e.g., a response option that may be unclear is anchoring a ***scale*** using the term *normal*, which assumes that patients understand what is normal for the general population).
- Response options are appropriate for the intended population that will complete the assessment. For example, patients with visual impairment may find a VAS difficult to complete.
- Instructions to the respondent (i.e., patient, clinician, caregiver) for completing items and selecting responses for the items are adequate.
- The number of response options is justified empirically (e.g., using qualitative and/or ***quantitative research***, initial pilot testing, or existing literature).
- Responses for an item are appropriately ordered.
- Responses for items avoid potential ***ceiling*** or ***floor effects*** (e.g., it may be necessary to introduce more responses to capture worsening or improvement so that fewer patients respond at the response continuum top or bottom).
- Responses do not bias the direction of responses and are balanced (e.g., bias exists if possible responses are weighted towards one end).
- Use similar types of response options for items under the same domain (i.e., not mismatching response option types in a domain) to make it easier to interpret the data.
- In some instances, it may be appropriate to add a "not applicable," "not attempted," or "unable to do" option

9. COA Format, Instructions, and Training

It is important that the instrument format used in the clinical trial be consistent with the format that is used during the COA development process. Format refers to the exact questionnaire, diary, or interview script appearance used to collect the COA data. Format is specific to the type of COA administration and the data collection mode. Sponsors should provide the exact copy of the COA that will be used in their trial. FDA will include in its regulatory review the specific format used in the clinical trial including the order and numbering of items, the presentation of response options in single response or grid formats, the grouping of items, patterns for skipping items, and all instructions to interviewers or patients.

Training, including the process of implementation of the COA in a clinical trial can be incorporated into the user manual. FDA recommends that sponsors submit a COA user manual specifying how to incorporate the COA into a clinical trial in a way that minimizes administrator burden, patient/caregiver burden, missing data, and poor data quality.

10. Respondent and Administrator Burden

Undue physical, emotional, or cognitive strain on patients generally decreases the quality and completeness of COA data.

Factors that can contribute to respondent or administrator burden include the following:

- Length of questionnaire, interview, or task
- Difficulty of questionnaire or task (e.g., physical performance or cognitive testing)
- Formatting
- Font size too small to read easily
- New instructions for each item
- Requirement that patients consult records to complete responses
- Privacy of the setting in which the COA is completed (e.g., not providing a private space for patients to complete questionnaires containing sensitive information about their sexual performance or substance abuse history)
- Inadequate time to complete questionnaires, interviews, or tasks
- Inadequate time to administer questionnaires, interviews, or tasks
- ***Literacy*** level too high for population
- Items or tasks that patients are unwilling to complete
- Perception by patients that the interviewer wants or expects a particular response
- Need for physical help in responding for self-report (e.g., turning pages, holding a pen, assistance with a telephone, computer keyboard, or electronic device)

The degree of respondent burden that is tolerable for instruments in clinical trials depends on the frequency and timing of COAs, trial duration, and on respondent's cognition, illness severity, or treatment toxicity.

Indications of excessive respondent burden, through use of inappropriate items or response options, or other factors, include increases in the level of missing data and refusal rates.

11. Scoring of Items and Domains

For each item, numerical scores generally should be assigned to each answer category based on the most appropriate scale of measurement for the item (e.g., nominal, ordinal, interval, or ratio scales).

A ***scoring algorithm*** is a set of pre-specified rules to assign numerical value or values to quantify the responses to the COA. A scoring algorithm may create a single score from a single item or multiple items (e.g., domain score). This algorithm should incorporate the measurement scale of the items as discussed above.

Sponsors should propose a scoring algorithm for the proposed COA (s) and also submit an updated scoring algorithm(s) as COA development progresses. FDA will review the following:

- Rationale for the proposed scoring algorithm
- Evidence that the summary score is appropriate

    - Using qualitative research or defined statistical techniques, sponsors should justify the method chosen to combine items to create a score or to combine domain scores to create a general score.
    - Over-weighting may be a concern when the number of items varies per measured concept without a rationale, the values associated with the response options vary by item or the number of response options varies.
- Rationale for the weights used in the scoring algorithm
- Rationale and interpretability for the proposed score transformations (if applicable)
- Details on how missing COA data will be handled

The conceptual framework for COAs intended to measure a multi-domain concept will be complex because identifying all of the appropriate domains and items of the multi-domain concept can be difficult. Multi-domain COAs can be used to support claims about a multi-domain concept if the COA has been developed to measure the important and relevant domains of the multi-domain concept contained in the claim. However, the complex nature of multi-domain COA often raises significant questions about how to interpret and report results in a way that is not misleading. For example, if improvement in a score for a multi-domain concept (e.g., symptoms associated with a certain condition) is driven by a single responsive item (e.g., pain intensity improvement) whereas other important items (e.g., other symptoms) did not show a response, a general claim about the multi-domain concept (e.g., improvements in symptoms associated with the condition) cannot be supported. However, that single responsive item or domain may support a claim specific to that item or domain.

Capturing the treatment's effects on the core signs, symptoms, and impacts using separate items is encouraged because it would provide detailed information regarding the treatment's effects on each sign, symptom, and impact. If appropriate, these separate items could be combined into a summary score.

C. Evidence of Other Measurement Properties: Reliability, Construct Validity and Ability to Detect Change

Once the COA's content validity has been established, FDA considers in its review the following additional COA measurement properties: reliability, construct validity, and ability to detect change. (Note: data related to the instrument's other measurement properties will not be reviewed by FDA until content validity of the instrument has first been established). Establishing adequate measurement properties of the COA will not only further support the its content validity, it will also help reduce the noise in the instrument and may make it more sensitive to detect treatment benefit.

Ideally, exploratory studies are an opportune time to evaluate measurement properties of a COA because the study design and patient population will be similar to the confirmatory trials. Generally, it is more likely to observe patient changes in interventional exploratory studies in comparison to non-interventional exploratory studies (e.g., studies to evaluate COA measurement properties) because of treatment intervention. Data of within-patient change is necessary to evaluate the COA's ability to detect change and to establish a clinically meaningful within-patient change threshold.

Appendix 3 lists some of the measurement properties of COAs that are reviewed by FDA.

Additional details of these measurement properties are described in the following subsections.

1. Reliability

Because clinical trials measure change over time, the adequacy of a COA for use in a clinical trial depends on its reliability or ability to yield consistent, reproducible estimates of true treatment effect. It also should be noted that reliability is a necessary but not sufficient condition to establish evidence of validity.

The reliability indices to be used to demonstrate reproducibility of score depend on the type of COAs. For most COAs, the indices of test-retest reliability are usually sufficient. Test-retest reliability indicates whether the score is reproduced for the same patient across at least two time points whose condition has not changed.

- For PRO and PerfO measures completed by the patients themselves, the indices of test-retest reliability are needed to demonstrate reliability. (Note, in some trial settings it is not feasible to evaluate test-retest reliability (e.g., acute disease conditions, rapid-acting treatments). Discuss with FDA to confirm if your measure's planned use may fall in this category.)
- In the case of ClinRO, ObsRO, and interviewer-administered PRO or PerfO measures, the persons other than the patients are administering or completing the assessments and therefore could be generally regarded as raters. For these COAs, depending on the intended use of the COAs and the study design, assessing intra-rater reliability and inter-rater reliability may be necessary to demonstrate reproducibility of the scores. For example:
    - For COAs or study design where the same rater will rate several patients, it may be necessary to examine the intra-rater reliability. A COA demonstrates adequate intra-rater reliability when there is high agreement among COA ratings by the same rater on multiple patients of the same disease condition.
    - In the case where multiple raters are used to rate the same patient, it may be necessary to examine the inter-rater reliability. A COA demonstrates adequate inter-rater reliability when there is high agreement on COA ratings among multiple raters for the same patient at the same time point.
- For PerfO measures that utilize devices to capture and/or record the data, the ability of the devices to perform reliably and consistently also needs to be documented.

Indices of internal consistency, the extent to which items comprising a scale measure the same construct (e.g., Cronbach's Alpha coefficient), may not constitute sufficient evidence of reliability in the absence of reliability indices to assess score reproducibility.

2. Construct Validity

Construct validity of a COA is determined by evidence that relationships among items, domains, and concepts conform to *a priori* hypotheses concerning logical relationships that should exist

with other related measures or characteristics of patients and patient groups (e.g., a COA intended to measure physical function should have a positive association with another existing physical function measure).

FDA reviews the construct validity of a COA to determine whether the documented relationships between results gathered using the current instrument and results gathered using other related measures are consistent with *a priori* hypotheses concerning those relationships (i.e., discriminant and convergent validity). An example of assessing convergent validity would be to examine the associations between a patient global impression of symptoms severity and the endpoint score of a multi-item symptom measure. FDA also will review evidence that the COA can differentiate among clinically distinct groups hypothesized *a priori* (i.e., known groups analysis).

There is a special type of convergent validity, called the ***criterion validity***, where evidence of validity is established by quantifying the relationship between the scores of a COA and scores of a known gold standard measure of the same concept. If a criterion measure is proposed, sponsors should provide rationale and support that the criterion is an accepted gold standard measure (i.e., relevant, valid, and reliable).

3. Ability to Detect Change

FDA reviews a COA's ability to detect change using data that compares change in COA scores to change in other similar measures that indicate that the patient's state has changed with respect to the concept of interest. A review of the ability to detect change includes evidence that the instrument is sensitive to gains and losses in the measurement concept and to change across the entire range expected for the target patient population.

When patient experience of a concept changes, the value(s) for the COA measuring that concept also should change. If there is clear evidence that patient experience relative to the concept has changed, but the value(s) of the COA do not change accordingly, then either the ability to detect change is inadequate or the COA's content and/or construct validity should be questioned. Conversely, if there is evidence that value(s) of the COA are affected by changes that are not specific to the concept of interest, the COA's content and/or construct validity may be questioned.

The ability of a COA to detect change may influence the calculation of sample size for evaluating the effectiveness of treatment. In general, an inability of a COA to detect change tends to support the null hypothesis of no treatment effect.

D. Interpretation of Meaningful Change

FDA is interested in what constitutes a meaningful within-patient change in the concepts assessed by the COAs. Statistical significance can sometimes be achieved for small group-level mean differences; however, statistical significance alone does not indicate whether an individual patient has experienced a meaningful clinical benefit. Additionally, to holistically determine what is a meaningful change, both benefit and risk (i.e., improvement and

deterioration) may need to be accounted for. This document is not directly addressing this integration of benefit and risk, but the methods described can be used to help interpret benefit or risk. As such, special consideration should be given by the sponsor to assess how meaningful the observed differences are likely to be. To aid in the interpretation of the COA endpoint(s) results, sponsors should propose an appropriate threshold(s) (e.g., a range of score change) that would constitute a clinically meaningful within-patient change in scores in the target patient population for FDA review.

In addition, if the selected threshold(s) are based on transformed scores (e.g., linear transformation of a 0-4 raw score scale to a 0-100 score scale), it is important for the sponsors to consider score interpretability of the improvement threshold(s) for both transformed scores and raw scores, i.e. whether the selected threshold(s) based on transformed scores also constitute a clinically meaningful within-patient change for the raw scores. Depending on the proposed score transformation, selected improvement threshold(s) based on transformed scores may reflect less than one category change on the raw score scale, which is not useful for the evaluation and interpretation of clinically meaningful change.

Meaningful Within-Patient Change vs. Between-Group Mean Differences

Individual within-patient change is different than between-group mean difference or treatment effect. From a regulatory standpoint, FDA is more interested in what constitutes a meaningful within-patient change in scores from the ***patient perspective*** (i.e., individual patient level). The between-group mean difference is the difference between the average score change between two study arms that is commonly used to evaluate treatment difference, but it does not address the individual within-patient change that is used to evaluate whether a meaningful score change is observed. A treatment effect is different than a meaningful within-patient change. The terms minimally clinically important difference (MCID) and minimum important difference (MID) do not define meaningful within-patient change if derived from group-level data. Additionally, the minimum change may not be sufficiently to serve as a basis regulatory decisions.

1. Anchor-based Methods to Establish Meaningful Within-Patient Change

FDA recommends the use of anchor-based methods supplemented with both empirical cumulative distribution function (eCDF) and probability density function (PDF; often estimated using kernel density estimation) curves to establish a threshold(s), or a range of thresholds, that would constitute a meaningful within-patient change score of the target COA for the target patient population. Anchor-based methods utilize the associations between the concept of interest assessed by the target COA and the concept measured by independent anchoring measure(s), often other COAs. The anchor measure(s) are used as external criteria to define patients who have experienced a meaningful change in their condition. The meaningful change scores of the COA measure can then be derived from the group of patients who are identified as having experienced meaningful change based on the anchor measure(s). Sponsors should provide evidence for what constitutes a meaningful change on the anchor scale. **Table 4** lists some considerations for anchor measures.

Table 4. Considerations for Anchor Measures

Considerations for Anchor Measure(s):

- Selected anchors should be plainly understood in context, easier to interpret than the COA itself, and sufficiently correlated to the targeted COA.
- Multiple anchors should be explored to provide an accumulation of evidence to help interpret meaningful within-patient score change which can also be a range.
- The following anchors are recommended to generate appropriate threshold(s) that represent a meaningful within-patient change in the target patient population:
    - Static, current-state global impression of severity scale (e.g., patient global impression of severity or PGIS)
    - Global impression of change scale (e.g., patient global impression of change or PGIC)
    - Well-established clinical outcomes (if relevant)
- A static, current state global impression of severity scale is recommended at minimum, when appropriate, since these scales are less likely to be subject to recall error than global impression of change scales; they can also be used to assess change from baseline.

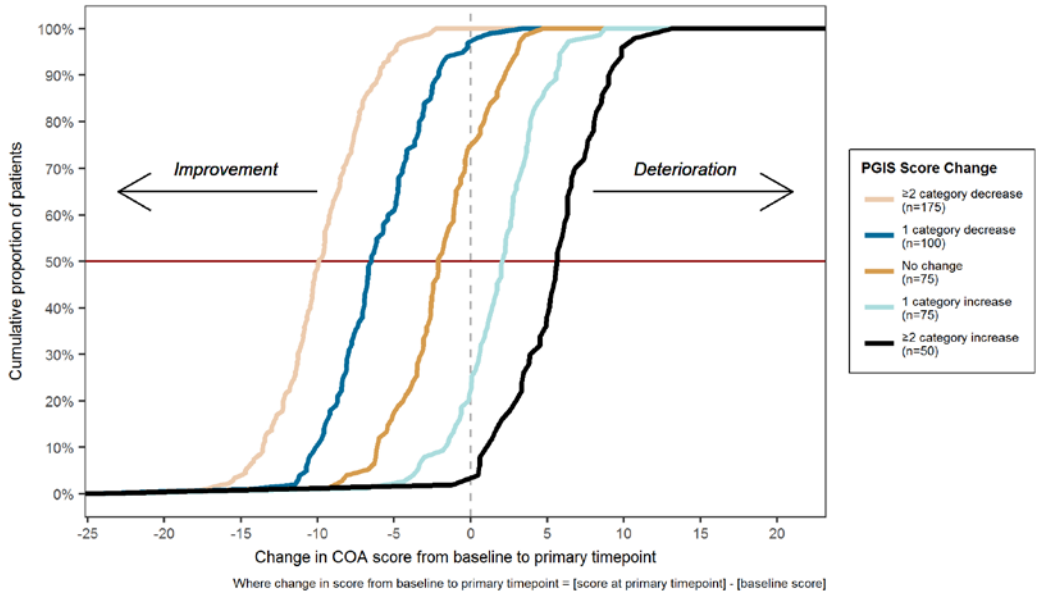
Refer to **Appendix 4** for example copies of generic PGIS and PGIC scales.

2. Using Empirical Cumulative Distribution Function (eCDF) to Supplement Anchor-based Methods

The eCDFs display a continuous view of the score change (both positive and negative) in the COA endpoint score from baseline to the proposed time point on the X-axis (horizontal axis), with the Y-axis (vertical axis) representing the cumulative proportion of patients experiencing up to that level of score change. The eCDF curve should be plotted for each distinct anchor category as defined and identified by the anchor measure(s) (e.g., much worse, worse, no change, improved, much improved). The meaningful within-patient threshold of the target COA should be explored by the eCDF of the anchor category where the patients are defined and judged (by the anchor measure) as having experienced meaningful change in their condition.

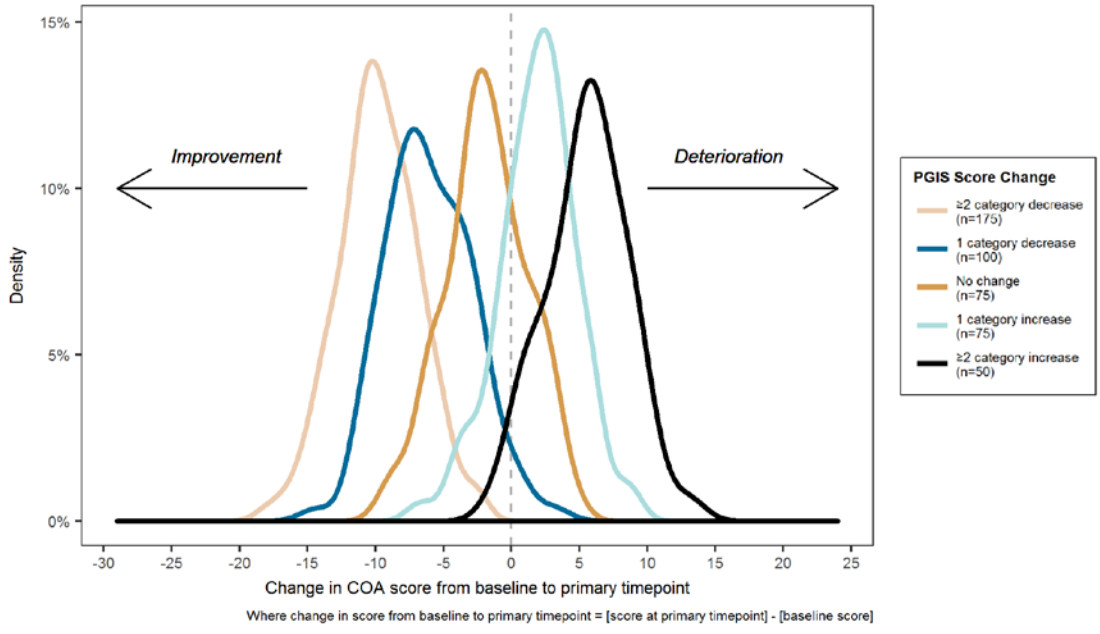
As a reference, **Figure 8** provides an example of a eCDF curve. Note that the median change is indicated by the red line in this example.

Figure 8. Example #1 of Empirical Cumulative Distribution Function (eCDF) Curves of Change in COA Score from Baseline to Primary Time Point by Change in Patient Global Impression of Severity (PGIS) score



The PDF curves are useful in aiding the interpretation of eCDF curves. Compared to eCDF curves, PDF curves may provide an easier overview of the shape, dispersion, and skewness of the distribution of the change from baseline in the endpoint of interest across various anchor categories. **Figure 9** provides an example of a PDF curve.

Figure 9. Example #1 of Density Function (PDF; often estimated using kernel density estimation) Curves of Change in COA Score from Baseline to Primary Time Point by Change in PGIS Score



PDF curves can be especially informative for diagnosis purpose when there is not clear consistent separation between the eCDF curves. In **Figure 10**, there is not clear separation between the 1-category decrease and the "No change" curves. Examination of **Figure 11** suggests that the variance differs across the PGIS change categories.

Figure 10. Example #2 of Empirical Cumulative Distribution Function (eCDF) Curves of Change in COA Score from Baseline to Primary Time Point by Change in Patient Global Impression of Severity (PGIS) score

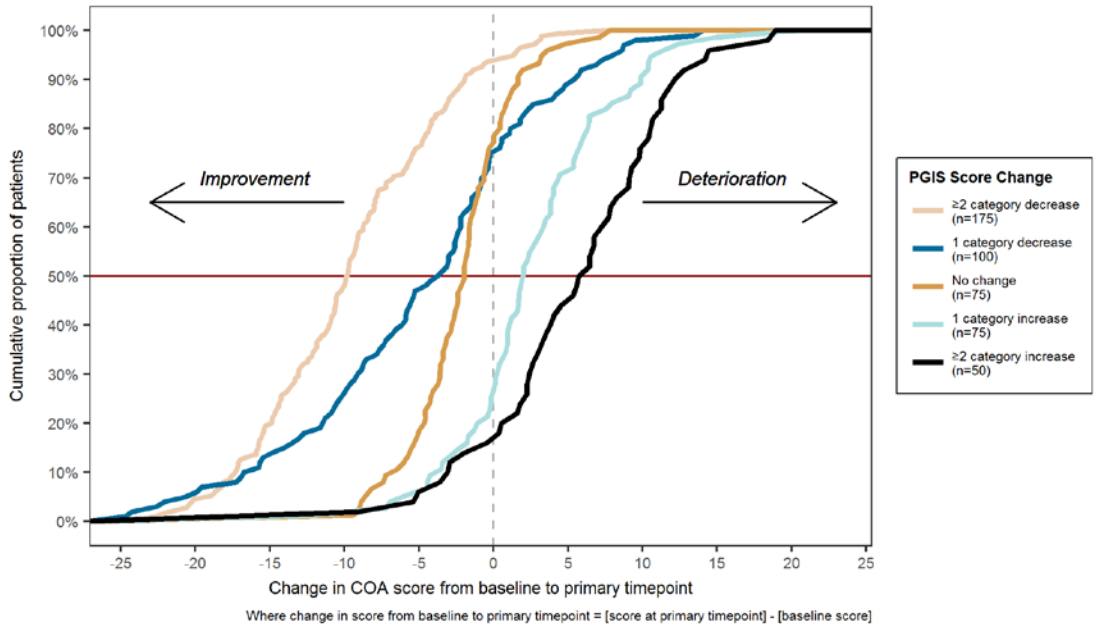
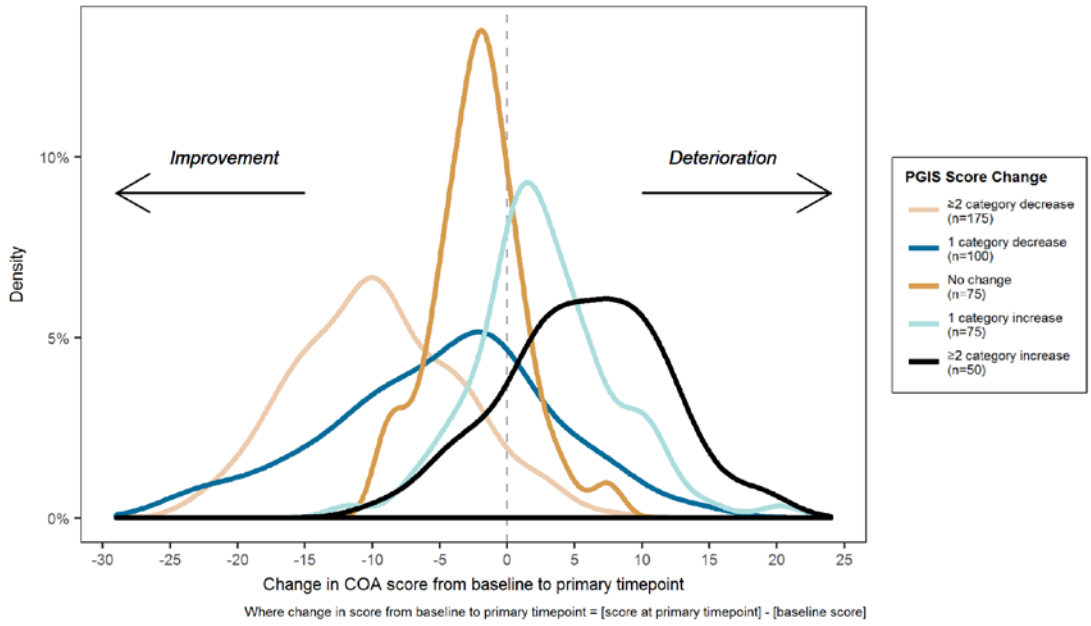


Figure 11. Example #2 of Density Function (PDF; often estimated using kernel density estimation) Curves of Change in COA Score from Baseline to Primary Time Point by Change in PGIS Score



3. Other Methods

Other methods may be explored to complement the anchor-based methods or when anchor-based methods are not feasible (i.e., when no adequate anchor measure(s) are available). For example, patients can be queried via cognitive interviews, exit interviews, or surveys to help inform the improvement threshold.

Distribution-based methods (e.g., effect sizes, certain proportions of the standard deviation and/or standard error of measurement) do not directly take into account the patient voice and as such cannot be the primary evidence for within-patient clinical meaningfulness. Distribution-based methods can provide information about measurement variability.

VII. CLINICAL TRIAL DESIGN CONSIDERATIONS

In general, clinical trial planning should proceed sequentially starting with clear trial objectives, including the specification of well-defined endpoints (i.e., "what is to be estimated" to address a specific scientific question of interest), and an appropriate analysis plan including exploration of the robustness of the inference through sensitivity analyses. Sponsors should refer to the ICH E9 guideline for additional details regarding a framework to align planning, design, conduct, analysis (including missing data) and interpretation.[15]

[15] ICH E9 (R1) addendum on estimands and sensitivity analysis in clinical trials to the guideline on statistical principles for clinical trials.

A. General Protocol Considerations for COA Endpoints

Clinical trials using COAs should be designed so that:

- COA data collection corresponds with, and is completed at, times specified in the clinical trial protocol and in correspondence with the clinical trial design.
- COAs intended to support meaningful outcomes to patients (i.e., labeling claims or other communications) are fit-for-purpose, sensitive to detect clinically meaningful change; content and scoring information should be clearly delineated in the clinical trial protocol.
- COAs intended to support approval and/or labeling claims are appropriately positioned in the endpoint testing hierarchy.
- COA measurement is obtained before or at the time of patient withdrawal from the clinical trial.
- Plans for COA measurement after discontinuation from treatment should be driven by the research questions.

1. Endpoint definition(s)

An endpoint is a precisely defined variable intended to reflect an outcome of interest that is statistically analyzed to address a particular research question. A precise definition of an endpoint (i.e., endpoint definition) typically specifies the type of assessments made, the timing of those assessments, the assessment tools used, and possibly other details, as applicable, such as how multiple assessments within an individual are to be combined. Within the protocol, the specific clinical benefit measurement concept(s) should be assessed by fit-for-purpose COA(s) and should be incorporated into a corresponding clinical trial objective or hypothesis and reflected in the endpoint definition and positioning in the testing hierarchy.

B. General Protocol Considerations for Blinding/Masking

1. Blinding (Masking)

The protocol should specify who will evaluate the COA endpoints, outcomes, or measurements in relation to the subjects (e.g., the investigator or an independent evaluator/rater) as well as who the intended reporter of patient information will be (e.g., clinicians, patients or caregivers) and to what extent blinding (masking) will be maintained among the investigators, evaluators/raters and reporters (e.g., clinicians, patients or caregivers). Note that if masking is not possible (e.g., open-label study) and a COA is being proposed as a primary or key secondary endpoint in the endpoint hierarchy, the study design may limit interpretation of data from the COA. Patients' and/or clinicians' knowledge of treatment assignment may lead to systematic overestimation or underestimation of the treatment effect, the magnitude of which is unknown. Use of a control (either concurrent or natural history, as appropriate) is a necessary element of an adequate and well-controlled trial as described in FDA regulations. However, we acknowledge that in some cases, a double-blind, randomized controlled study may not be feasible in the context of the disease, condition, and/or medical product type. In cases where unmasking may occur, this limitation will need to be overcome by demonstrating a substantial clinically meaningful effect in the setting of strict adherence to a well-conducted clinical trial. Note that the size of the effect

as well as the association between the COA and other clinically meaningful measures collected in the trial are used when interpreting clinical trial results.

C. Frequency of Assessments for COA Endpoints

A study assessment schedule should be included in the clinical trial protocol, outlining the timing of assessments for each clinical endpoint listed in the endpoint hierarchy. The following should be considered when determining the most appropriate frequency of assessments for COA endpoints:

- The COA schedule of assessments should correspond directly with the natural course of the disease or condition, research questions to be addressed, trial duration, and be administered within the expected timeframe for observing changes in the concept(s) of interest. The timing should also take into account the COA recall period.

- The timing of COA administration should align with the administration of other pre-specified endpoints (i.e., primary and secondary) and proposed ***data analysis plan***.

- COAs should be administered at baseline. If the trial includes a run-in period during which the effect on the COA might be expected to change (e.g., medication washout, patient behavior modification), this should be taken into account when considering the timing of assessments. Note that some diseases, conditions, or clinical trial designs may necessitate more than one baseline assessment and several COA assessments during treatment.
- The timing of anchor scale administration should align with the administration of the corresponding COA (e.g., patient global impression - severity (PGIS) with PRO timing; clinician global impression of severity with ClinRO timing).

D. Clinical Trial Duration for COA Endpoints

When designing a clinical trial with the intended use of a COA(s) as endpoint measure(s), the required length of the primary assessment period and follow-up period should be determined based on the natural history of the disease or condition and the expected timeframe within which the intervention is expected to demonstrate a positive effect on the outcome(s) of interest. Determination of the clinical trial duration should be driven both by the disease course and endpoint objectives outlined in the clinical trial protocol. It is important to consider whether the clinical trial's duration is of adequate length to assess a durable outcome in the disease or condition being studied and support the proposed labeling claim(s). Generally, the duration of a COA assessment period should be the same duration as indicated for other measures of effectiveness in the clinical trial protocol.

E. Design Considerations for Multiple Endpoints (Including COA Endpoints)

Since most diseases have more than one relevant clinical outcome, trials can be designed to examine the effect of a medical product on more than one endpoint (i.e., multiple endpoints). Additional details regarding regulatory considerations for a multiple endpoint approach have

been published by FDA (FDA, 2017b).

F. Use of Electronic Mode of Administration

There are two main data collection modes of capturing COA data in clinical trials – paper and electronic. Examples of potential advantages to implementing electronic data capture are as follows:

- **No need for manual secondary data entry** of raw paper data into an electronic database for data analysis; manual data entry can potentially introduce human error.
- **Direct transmission** into an electronic database reduces risk to data integrity.
- **Alarm or reminder capabilities** can be set at regular intervals, or incoming phone calls when using Interactive Voice Response (IVRS), to minimize the risk of missing data and to increase the potential for greater patient compliance.
- **Time and date stamp capabilities** ensure patient compliance - providing verification of data completion at the appropriate times according to the clinical trial design. This helps eliminate the occurrence of the "parking lot" phenomenon, where a patient might fill out all of the daily paper diary entries spanning weeks of data in one sitting, immediately before handing them into the investigator.
- **Real-time data recording and transmission** (e.g., recording of signs and symptoms experienced with each bowel movement; patient logs of pain and or rescue medication use) facilitates site data monitoring, allowing site staff to know which patients are out of compliance (e.g., with COA assessment, medication use, etc.) and follow-up with those patients in a timely manner.
- **Remote data capture** allows for reduced frequency and duration of in-person clinic visits, thereby reducing both site and patient burden.

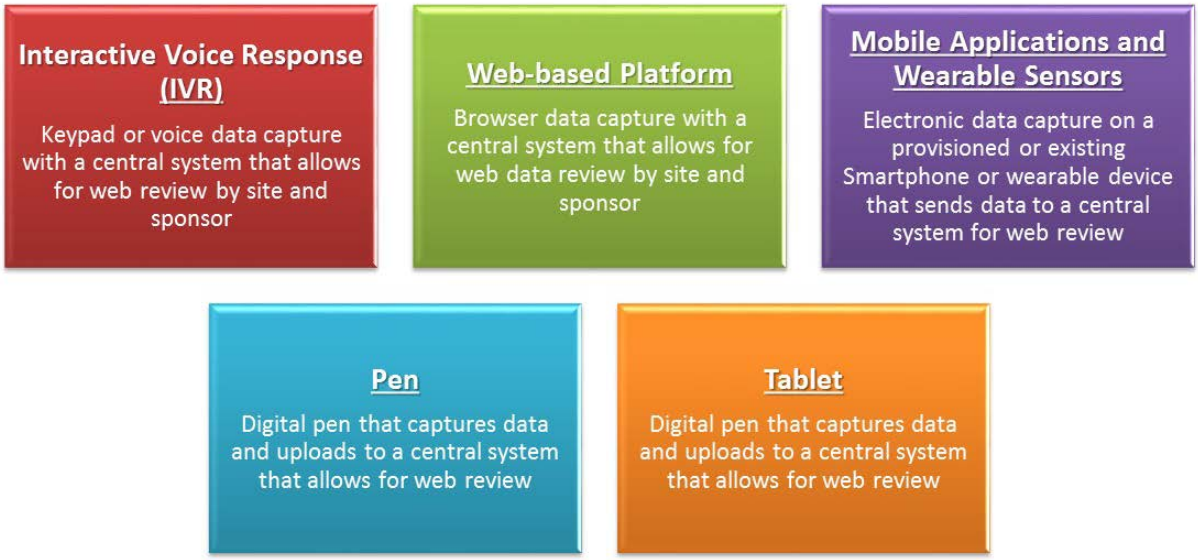
Recommendations for electronic (e) modes of administration for COA (eCOA) have been set forth by the FDA and should be considered when determining the suitability of each subtype for implementation in the context of a clinical trial (FDA, 2013).

1. eCOA Selection

Figure 12 shows five main eCOA subtypes:[16]

[16] http://www.appliedclinicaltrialsonline.com/comparing-five-methods-collect-patient-driven-edata

Figure 12. eCOA Subtypes



When data is captured electronically, FDA recommends that sponsors consider using an electronic device that enables programming of daily reminders using an alarm function, in addition to external alarm methods (e.g., email, phone call, and/or text alerts), when feasible. Automated reminders and alarms tend to minimize missing data and allow for automatic recording of other important information (e.g., timestamps for data input) (FDA, 2013). For sponsors who proceed with electronic data capture, FDA recommends that they have a back-up plan (e.g., web-, phone-, or paper-based) implemented in case there are any malfunctions with the electronic devices although we caution against mixing data collection modes (e.g., paper and electronic) when data will be pooled for analysis.

Additionally, although there is increasing interest in having subjects bring their own devices (e.g., smartphone, tablet) (with back-up device option for those without their own devices), FDA recommends that sponsors reduce variations in instrument format and functionality from one device to another by using a single platform throughout a clinical trial rather than mixing platforms over the course of the study. If a sponsor chooses to proceed with having subjects bring their own devices, they should present a detailed plan for Agency review and comment to ensure that the instrument will function as intended across device platforms.

2. Paper-electronic Migration and Equivalence

When considering the migration of a paper COA to an electronic format, the following design elements might change:

- Design decisions (e.g., multiple items on a page versus one item per screen)
- Skip patterns and/or adaptive design
- Introduction of automated compliance reminder alarms
- Potential for forced response (i.e., not allowing respondents to skip items in order to complete the COA)

FDA evaluates paper to electronic format COA migrations in the context of whether the sponsor will need to compare or pool COA data from mixed data collection modes in a single clinical trial. Note that mode equivalence testing is not necessarily required in all cases, for regulatory purposes. The magnitude of changes to paper questionnaire content and the extent to which those changes alter the meaning or interpretation of the instrument's items and/or response options determines whether an equivalency study will be recommended (Coons et al., 2009). When switching from paper to electronic data collection modes, sponsors should develop separate device-related respondent instructions and training materials for submission to FDA for review and comment.

Additional considerations regarding eCOA migration and equivalency testing can be found in the International Society for Pharmacoeconomics and Outcomes Research (ISPOR) Task Force paper on Evidence Needed to Support Measurement Equivalence between Electronic and Paper-Based Patient-Reported Outcome (PRO) Measures.[23] When equivalence testing is recommended, a small nonrandomized study may be adequate to compare the distribution of responses between versions of a questionnaire with different formats. If the COA will be used in a significantly different patient population (e.g., a different disease or age group), FDA may recommend conducting qualitative studies to confirm content validity in the new population. A small randomized study to ascertain the measurement properties in the new population may minimize the risk that the instrument will not perform adequately in a clinical trial.

Compared to paper COAs, additional documentation for eCOAs may be important for FDA to review, such as design features like skip patterns and forced response.

Example

Scenario: A sponsor is developing an eCOA and the goal is for respondents to answer each question in the eCOA.

What are some ways to minimize inaccurate data for forced responses?
Include the following response options to allow skipping of items when completing the eCOA, if necessary:

- "I prefer not to answer this item"
- "Not applicable"

With regard to missing data, it is helpful if sponsors describe their plans for addressing potential missing data in analyses. Even if sponsors implement a forced response, respondents can simply turn off their device which would result in missing data.

Sponsors should include for FDA review any device ***usability testing*** analyses and results, as well as electronic screen shots of the instrument, patient, investigator, and site training materials and documentation related to migrating or reformatting an existing instrument from paper to electronic format, when data from mixed data collection modes will need to be compared or pooled.

3. Device Validation

eCOAs should undergo a rigorous validation process prior to implementation in clinical trials to ensure device and program functionality and performance stability within the clinical trial context. Essential components of the validation process are outlined in the International Society for Pharmacoeconomics and Outcomes Research (ISPOR) Task Force paper on the Validation of Electronic Systems to Collect Patient-Reported Outcome (PRO) Data (Zbrozek et al., 2013).

Sponsors are encouraged to perform usability testing with patient cognitive interviews during the validation process, after ***user acceptance testing (UAT)***, to assess device functionality, questionnaire comprehension, and ease of use in the patient population. This step helps to minimize the risk of having poor quality data due to patients' misunderstanding or incomplete understanding of how to use the device and to ensure the device is usable in the patient population.

4. Data-Related Regulatory Considerations

Key eCOA data regulatory considerations are outlined below:

- Sponsors and investigators should ensure that FDA regulatory requirements are met for record keeping, maintenance, and access.[17]
- Source data control should be maintained by the clinical investigator(s) (FDA, 2013)
- eCOA data should be compliant with all FDA regulatory guidelines as well as the International Conference on Harmonization's (ICH) Guideline for Good Clinical Practice.
- Per FDA guidance on requirements for regulatory submissions in electronic format (see bulleted list below) sponsors should submit a study data tabulation model (SDTM) and analysis data sets including raw score data as well as transformed score data if raw score transformation is performed.

The following are key guidelines that should be considered when considering regulatory eCOA data considerations:

  - Guidance for Industry: Computerized Systems Used in Clinical Investigations (FDA, 2007)
  - Guidance for Industry: Electronic Source Data in Clinical Investigations (FDA, 2013)
  - Study Data Technical Conformance Guide: Technical Specifications Document *(incorporated by reference into the Guidance for Industry: Providing Regulatory Submissions in Electronic Format – Standardized Study Data)* (FDA, 2018)
  - 21 CFR Part 11 "Electronic Records; Electronic Signatures"
  - 21 CFR Part 312 (INDs)
  - 21 CFR Part 812 (Devices)
  - ICH Guideline for Good Clinical Practice E6 (R2)

[17] 21 CFR Part 11 "Electronic Records; Electronic Signatures"

VIII. DATA ANALYSES

The statistical analysis considerations for COA endpoints are similar to the statistical considerations for any other endpoint(s) used in medical product development. However, there are some special considerations for COA endpoints as well. The most important of these considerations are discussed in the following sections.

A. General Statistical Considerations

Every protocol should describe the principal data analysis features in the statistical section with a detailed elaboration of the analysis in a statistical analysis plan (SAP). FDA intends to determine the adequacy of clinical trial data to support claims or communication in light of the pre-specified method for endpoint analysis.

A multi-domain COA may successfully support a labeling claim or communication based on one or a subset of the domains measured if an *a priori* analysis plan pre-specifies the domains that will be targeted as endpoints and the method of analysis that will adjust for the multiplicity of tests for the specific claim. The use of domain subsets as clinical trial endpoints presupposes that the COA was adequately developed and validated to measure the subset of domains independently from the other domains. A complex, multi-domain claim cannot be substantiated by instruments that do not adequately measure the individual components of the domain adequately.

B. Multi-Component Endpoints

As described in Section VI. Considerations of Clinical Trial Design, there are often multiple endpoints that would be of clinical interest. Clear distinctions should be made for two types of circumstances when multiple endpoints are encountered: (1) multi-component endpoints and (2) composite endpoints. Definitions and additional details regarding statistical considerations for a multiple endpoint approach were published by the FDA (FDA, 2017b). For a COA with multiple domains, a within-patient combination of all the domain (i.e., component) scores to calculate a single overall rating or status determined according to specific rules creates a *multi-component endpoint*. Multi-component endpoints have a few advantages (e.g., they may reduce multiplicity problems and/or provide some gains in efficiency if different components are generally concordant). The multi-component endpoint needs to be clinically relevant and interpretable.

C. Patient-Level Missing COA Data

Even with the best planning, patient-level COA data may be missing at the end of the clinical trial. Sponsors should provide patients adequate education on the purpose of collecting the COA data to encourage patient compliance with completing COAs and help prevent and reduce the frequency of potential missing data in the first place. Missing data should be distinguished from data that do not exist or data that are not considered meaningful due to an intercurrent event. The protocol and the SAP should address plans for how the statistical analyses will handle missing COA data when evaluating clinical benefit and when considering patient success or patient response.

1. Missing Items within Domains

At a specific patient visit, a domain measurement may be missing some, but not all, items. One approach to handling this type of missing data is to define rules that specify the number of items that can be missing and still consider the domain as adequately measured. Generally, this approach can only be valid under the assumption that all items are equally important and interchangeable. Careful consideration should be given when items on a COA exhibit a hierarchy of clinical importance. Rules for handling missing data should be specific to each COA and usually should be determined during the instrument development process. The SAP should specify plans to assess the impact of missing data (i.e., a missing data simulation study) and all rules for handling missing data. For example, the SAP can specify the proportion of items that can be missing before a domain is treated as missing.

2. Missing Entire Domains or Entire Measurements

Sponsors should clearly define missing data and propose statistical methods that properly account for missing data with respect to a particular estimand. How to handle the missing data for a COA endpoint and any related supportive endpoints should be addressed in the protocol and the SAP. In addition, the sensitivity analyses of the COA endpoints should be prospectively proposed in the protocol and the SAP. These analyses investigate assumptions used in the statistical model for the main analytic approach, with the objective of verifying that inferences based on a particular estimand should be robust to limitations in the data and deviations from the assumptions.

IX. SPECIAL PATIENT POPULATION CONSIDERATIONS

A. Rare Disease Patient Populations

Endpoint selection for a clinical trial involves many factors that are often not well understood for rare diseases, such as:

- Disease etiology and natural history;
- Patient age
- Phenotypic heterogeneity and subsets;
- Rate and variability of symptom/sign occurrence; and
- Disease manifestations most meaningful to patients that might be a target for the medical product.

For many rare diseases, well-characterized efficacy endpoints developed in the disease population are not available. Rare diseases often need more sensitive outcome measures to quantify disease. COAs that may be fit-for-purpose for use across multiple therapeutic areas (e.g., some physical function scales) may not be applicable to some rare diseases. As such, sponsors should consider the testing and thoughtful application of existing COAs developed in other patient populations that appropriately measure the concept of interest for use in rare disease trials. Note that using COAs that have been developed in other patient populations "off the

shelf" without careful thought and evidence to support their suitability in the rare disease patient population can lead to unsuccessful trials and/or difficulty with data interpretation. Developing new endpoints and COAs in rare diseases is often complicated by additional challenges, including but not limited to:

- Small patient populations for inclusion in studies
- Cognitive and/or linguistic developmental differences
- Willingness, ability, and motivation to self-report by age ***subgroups***
- Availability of few disease experts
- Wide geographic dispersion of patients

Where possible, sponsors should conduct well-designed natural history studies independently or through partnerships with patient organizations and/or utilize existing natural history and/or patient registry data. When conducting natural history studies, sponsors should consider including fit-for-purpose COAs that would inform clinical benefit assessment in future rare disease trials in that patient population.

When heterogeneity in disease symptoms, signs, and impacts exists, sponsors should consider defining a clinical trial endpoint based on measurement concepts that are most important to a broad range of patients, while including less common or less important concepts lower in the endpoint testing hierarchy. Note that the reliability, validity, sensitivity, or interpretability of an endpoint may be different across patient subpopulations (e.g., early-stage or slowly-progressing phenotypes vs. severe, late-stage, or rapidly-progressing phenotypes). Therefore, sponsors should consider including COAs appropriate for assessment in a diverse range of patients (e.g., heterogeneity of clinical manifestations) who may benefit from the target treatment, whenever possible.

Sponsors should engage early with therapeutic area and COA experts who understand the nuances of disease progression and COAs in rare diseases, as well as with FDA, to get input on COA selection, modification, or development and implementation processes that will generate reliable, valid, and interpretable data.

The small number of affected patients often necessitates multinational clinical development programs and thus sponsors need to consider the impact of language, culture, and customs on the interpretability and relevance of COAs. Likewise, many rare diseases affect children necessitating the development of age-appropriate endpoints and assessment tools (see **Section IX.B**).

In rare diseases with very small sample sizes, traditional COA development and validation may not be feasible; therefore, FDA is flexible and open to other approaches (e.g., combined concept elicitation and cognitive interview studies) and discussing various approaches with sponsors (Benjamin et al., 2017). Given the challenges of COA development and measurement in rare diseases, FDA encourages pre-competitive collaboration among FDA, patient groups, medical product developers, COA developers, and other stakeholders with the goal of publicly-available COAs for use across multiple medical product development programs. FDA has published more information on medical product development in rare diseases (FDA, 2015).

B. Pediatric Patient Populations

Medical product development in pediatric populations is a high priority for FDA. The use of age-appropriate COAs to support clinical trial endpoints is an important consideration when planning a clinical investigation in pediatric patients. Pediatric patients, and in many cases parents or guardians, input is important to identify clinically relevant and meaningful concepts that are important to patients and ensure these concepts are assessed using fit-for-purpose COAs in the target patient population.

In pediatric trials covering different age ranges, sponsors should consider the following unique characteristics when developing and implementing COAs in pediatric trials:

- Cognitive and linguistic development differences
- Ability to recall their experiences and reliably and validly self-report
- Willingness to self-report or perform a particular task
- Ability and motivation to complete study assessments according to instructions
- The complexity of the measurement concept and the assessment methods used (e.g., appropriateness of the assessment's recall period, requiring children to average their experiences over a specific length of time, consider use of fewer response options, etc.)
- Potential differences in disease manifestations by age groups

As pediatric-specific COAs do not necessarily exist in many therapeutic areas, it can be tempting to adapt a COA from one context of use for use in another (e.g., adults to children). However, inappropriate adaptation of an instrument can lead to problems with the instrument's content validity resulting in uninterpretable clinical trial results that may be unsuitable for regulatory use. In some diseases, signs, symptoms, and functioning may differ across the age span necessitating use of different COAs. Additionally, use of an assessment developed in adults that asks about concepts that may not be applicable to children in their stage of life (e.g. missed work, dating, etc.) would be inappropriate. Any adaptation of a COA should involve the target population and documentation of the adaptation process.

It is important to consider whether a certain type of COA can be validly and reliably completed by young children or those with cognitive impairment. PRO instruments developed for pediatric self-administration should be completed by the child independently without any assistance from the parent or caregiver in order to avoid parental influence on the child's responses. Mode of self-administration should be considered. For example, it may be easier for children to respond to COAs using a touch screen on an electronic device rather than using a personal computer's mouse or keyboard. In addition, COAs that ask simple questions and include few items and few response options are preferable for use with pediatric patients. Self-administration may not be suitable for use with very young children; interviewer-administration by a trained interviewer may be explored as an option with children who cannot reliably and validly self-report.

In addition, young children may be limited in their cognitive understanding of certain concepts, such as ability to compare numbers/amounts (e.g., more/less, greater/fewer), calendar time (i.e., the meaning of "a week" or "a month"), periods of time (e.g., a 24-hour period), and ability to understand sequences of events (e.g., before/after). In these cases, momentary assessments or

PRO instruments that do not require the child to understand and recall over a 24-hour period may be considered. Additionally, young children may be limited in their understanding of certain response scales used in a PRO instrument. For example, a pain intensity assessment using a numeric rating scale that may be appropriate for adults and adolescents who can self-report may not be well-understood by young children. Therefore, sponsors should explore simpler age-appropriate pain scales (e.g., scales with fewer, simpler response options, pictorial scales, etc.) for use with young children.

In general, the review considerations related to the development of pediatric-specific COAs are similar to those detailed for adults in this discussion document. Considerations for developing an observer-reported outcome (e.g., parent- or caregiver-report) will be discussed in **Appendix 5** to this discussion document.

C. Patients Cognitively Impaired or Unable to Communicate (Non-verbal)

For patients who are cognitively impaired or unable to communicate, FDA suggests using COAs other than a PRO (ObsRO, ClinRO, PerfO), or other novel approaches to assess their functioning. Using PRO assessments may not be appropriate in these patient populations as it may be difficult for patients with cognitive impairment or inability to communicate to complete the assessment consistently to provide accurate self-assessments of their own internal states and/or verbalize their feelings accurately.

Severely cognitively impaired patients may also be limited in their understanding of concepts related to calendar time, similar to cases involving young children. As such, the recall period should be carefully selected for this population.

REFERENCES

Benjamin K, Vernon MK, Patrick DL, Perfetto E, Nestler-Parr S, Burke L; ISPOR Rare Disease Clinical Outcomes Assessment Task Force. Patient-reported outcome and observer-reported outcome assessment in rare disease clinical trials: An ISPOR COA Emerging Good Practices Task Force report. Value Health. 2017;20:838-55.

Coons SJ, Gwaltney CJ, Hays RD, Lundy JJ, Sloan JA, Revicki DA, Lenderking WR, Cella D, Basch E; ISPOR ePRO Task Force. Recommendations on evidence needed to support measurement equivalence between electronic and paper-based patient-reported outcome (PRO) measures: ISPOR ePRO Good Research Practices Task Force report. Value Health. 2009;12(4):419-29. (https://www.ispor.org/workpaper/patient_reported_outcomes/Coons.pdf)

Dworkin RH, Turk DC, Farrar JT, Haythornthwaite JA, Jensen MP, Katz NP, et al. Core outcome measures for chronic pain clinical trials: IMMPACT recommendations. Pain 2005;113:9–19.

FDA. (2007). Computerized Systems Used in Clinical Investigations. Guidance for Industry. Retrieved July 23, 2018, from https://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM070266.pdf

FDA. (2013). Electronic Source Data in Clinical Investigations. *Guidance for Industry*. Retrieved July 23, 2018, from https://www.fda.gov/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM328691

FDA. (2015). Rare Diseases: Common Issues in Drug Development. *Draft Guidance for Industry.* Retrieved July 23, 2018, from https://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM458485.pdf

FDA. (2017a). Formal Meetings between the FDA and Sponsors or Applicants of PDUFA Products. *Draft Guidance for Industry*. Retrieved August 1, 2018, from https://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM590547.pdf

FDA. (2017b). Multiple Endpoints in Clinical Trials. *Draft Guidance for Industry*. Retrieved July 23, 2018, from (https://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM536750.pdf)

FDA. (2018). Study Data Technical Conformance Guide: Technical Specifications Document *(incorporated by reference into the Guidance for Industry: Providing Regulatory*

Submissions in Electronic Format – Standardized Study Data). Retrieved July 23, 2018, from https://www.fda.gov/downloads/ForIndustry/DataStandards/StudyDataStandards/UCM384744.pdf

Hawker GA, Mian S, Kendzerska T, et al. Measures of adult pain: Visual Analog Scale for Pain (VAS Pain), Numeric Rating Scale for Pain (NRS Pain), McGill Pain Questionnaire (MPQ), Short-Form McGill Pain Questionnaire (SF-MPQ), Chronic Pain Grade Scale (CPGS), Short Form-36 Bodily Pain Scale (SF-36 BPS), and Measure of Intermittent and Constant Osteoarthritis Pain (ICOAP). Arthritis Care Res (Hoboken) 2011;63:S240–52.

Wild D, Grove A, Martin M, Eremenco S, McElroy S, Verjee-Lorenz A, Erikson P; ISPOR Task Force for Translation and Cultural Adaptation. Principles of good practice for the translation and cultural adaptation process for patient-reported outcomes (PRO) measures: Report of the ISPOR Task Force for Translation and Cultural Adaptation. Value Health. 2005;8(2):94-104. (https://www.ispor.org/workpaper/research_practices/PROTranslation_Adaptation.pdf)

Zbrozek A, Hebert J, Gogates G, Thorell, R, Dell C, Molsen E, Craig G, Grice K, Kern S, Hines S. ISPOR ePRO Systems Validation Good Research Practices Task Force. Validation of Electronic Systems to Collect Patient-Reported Outcome (PRO) Data—Recommendations for Clinical Trial Teams: Report of the ISPOR ePRO Systems Validation Good Research Practices Task Force. Value Health. 2013;16: 480-89. (https://www.ispor.org/ValueInHealth/ShowValueInHealth.aspx?issue=F0949C6B-A38B-45F6-A90B-6433DD97F05C)