
Patient-Focused Drug Development: Selecting, Developing, or Modifying Fit-for- Purpose Clinical Outcome Assessments

**Guidance for Industry, Food and Drug
Administration Staff, and Other Stakeholders**

DRAFT GUIDANCE

This guidance document is being distributed for comment purposes only.

Comments and suggestions regarding this draft document should be submitted within 90 days of publication in the *Federal Register* of the notice announcing the availability of the draft guidance. Submit electronic comments to <http://www.regulations.gov>. Submit written comments to the Dockets Management Staff (HFA-305), Food and Drug Administration, 5630 Fishers Lane, rm. 1061, Rockville, MD 20852. All comments should be identified with the docket number listed in the notice of availability that publishes in the *Federal Register*.

For questions regarding this draft document, contact (CDER) Office of Communications, Division of Drug Information at druginfo@fda.hhs.gov, 855-543-3784 or 301-796-3400; or (CBER) Office of Communication, Outreach and Development at ocod@fda.hhs.gov, 800-835-4709 or 240-402-8010; or Office of Strategic Partnerships and Technology Innovation, Center for Devices and Radiological Health at cdrh-pro@fda.hhs.gov, 800-638-2041 or 301-796-7100.

**U.S. Department of Health and Human Services
Food and Drug Administration
Center for Drug Evaluation and Research (CDER)
Center for Biologics Evaluation and Research (CBER)
Center for Devices and Radiological Health (CDRH)**

**June 2022
Procedural**

Patient-Focused Drug Development: Selecting, Developing, or Modifying Fit-for-Purpose Clinical Outcome Assessments

Guidance for Industry, Food and Drug Administration Staff, and Other Stakeholders

Additional copies are available from:
Office of Communications, Division of Drug Information
Center for Drug Evaluation and Research
Food and Drug Administration
10001 New Hampshire Ave., Hillandale Bldg., 4th Floor
Silver Spring, MD 20993-0002
Phone: 855-543-3784 or 301-796-3400; Fax: 301-431-6353
Email: druginfo@fda.hhs.gov

<https://www.fda.gov/drugs/guidance-compliance-regulatory-information/guidances-drugs>

and/or

Office of Communication, Outreach and Development
Center for Biologics Evaluation and Research
Food and Drug Administration
10903 New Hampshire Ave., Bldg. 71, Room 3128
Silver Spring, MD 20993-0002
Phone: 800-835-4709 or 240-402-8010
Email: ocod@fda.hhs.gov

<https://www.fda.gov/vaccines-blood-biologics/guidance-compliance-regulatory-information-biologics/biologics-guidances>

and/or

Office of Policy
Center for Devices and Radiological Health
Food and Drug Administration
10903 New Hampshire Ave., Bldg. 66, Room 5431
Silver Spring, MD 20993-0002
Email: CDRH-Guidance@fda.hhs.gov

<https://www.fda.gov/medical-devices/device-advice-comprehensive-regulatory-assistance/guidance-documents-medical-devices-and-radiation-emitting-products>

U.S. Department of Health and Human Services
Food and Drug Administration
Center for Drug Evaluation and Research (CDER)
Center for Biologics Evaluation and Research (CBER)
Center for Devices and Radiological Health (CDRH)

June 2022
Procedural

Contains Nonbinding Recommendations

Draft — Not for Implementation

TABLE OF CONTENTS

I.	INTRODUCTION.....	1
A.	Overview of FDA Guidances on Patient-Focused Drug Development.....	1
B.	Purpose and Scope of the Guidance	3
II.	OVERVIEW OF COAS IN CLINICAL TRIALS.....	5
A.	Types of COAs	5
B.	The Concept of Interest and Context of Use for a COA.....	6
1.	<i>The Concept of Interest.....</i>	<i>6</i>
2.	<i>The Context of Use.....</i>	<i>8</i>
C.	Deciding Whether a COA Is Fit-for-Purpose.....	9
1.	<i>The Concept of Interest and Context of Use Are Clearly Described.....</i>	<i>9</i>
2.	<i>There Is Sufficient Evidence to Support a Clear Rationale for the Proposed Interpretation and Use of the COA.....</i>	<i>9</i>
III.	A ROADMAP TO PATIENT-FOCUSED OUTCOME MEASUREMENT IN CLINICAL TRIALS.....	10
A.	Understanding the Disease or Condition and Conceptualizing Clinical Benefits and Risks 12	
1.	<i>Understanding the Disease or Condition.....</i>	<i>12</i>
2.	<i>Conceptualizing Clinical Benefits and Risks</i>	<i>12</i>
B.	Select/Develop the Outcome Measure	12
1.	<i>Selecting the COA Type</i>	<i>12</i>
2.	<i>Evaluating Existing and Available COAs Measuring the Concept of Interest in the Context of Use</i>	<i>12</i>
3.	<i>Special Considerations for Selecting or Developing COAs for Pediatric Populations.....</i>	<i>15</i>
4.	<i>Using DHTs To Collect COA Data.....</i>	<i>16</i>
5.	<i>COA Accessibility and Universal Design</i>	<i>16</i>
C.	Developing a Conceptual Framework.....	17
IV.	DEVELOPING THE EVIDENCE TO SUPPORT THE CONCLUSION THAT A COA IS APPROPRIATE IN A PARTICULAR CONTEXT OF USE.....	19
A.	The Concept of Interest Should Be Assessed by [COA Type], Because . . .	20
B.	The COA Measure Selected Captures All the Important Aspects of the Concept of Interest.	20
C.	Respondents Understand the Instructions and Items/Tasks of the Measure as Intended by the Measure Developer.....	20
D.	Scores of the COA Are Not Overly Influenced by Processes/Concepts That Are Not Part of the Concept of Interest.	21
1.	<i>Item or Task Interpretations or Relevance Does Not Differ Substantially According to Respondents’ Demographic Characteristics (Including Sex, Age, and Education Level) or Cultural/Linguistic Backgrounds.....</i>	<i>21</i>

Contains Nonbinding Recommendations

Draft — Not for Implementation

2. <i>Recollection Errors Do Not Overly Influence Assessment of the Concept of Interest. [PRO, ObsRO, and ClinRO Measures]</i>	22
3. <i>Respondent Fatigue or Burden Does Not Overly Influence Assessment of the Concept of Interest.</i> 22	
4. <i>The Mode of Assessment Does Not Overly Influence Assessment of the Concept of Interest.</i>	22
5. <i>Expectation Bias Does Not Unduly Influence Assessment of the Concept of Interest.</i>	23
6. <i>Practice Effects Do Not Overly Influence the Assessment of the Concept of Interest. [PerfO Measures]</i>	24
E. The Method of Scoring Responses to the COA Is Appropriate for Assessing the Concept of Interest.	24
1. <i>Responses to an Individual Item or Task</i>	24
2. <i>Rationale for Combining Responses to Multiple Items or Tasks</i>	25
3. <i>Scoring a Unidimensional COA: The Reflective Indicator Model</i>	25
4. <i>Scoring for a COA That Summarizes Across Heterogeneous Health Experiences: The Composite Indicator Model</i>	27
5. <i>Scoring Approaches Based on Computerized Adaptive Testing</i>	27
6. <i>Approach to Missing Item or Task Responses</i>	28
F. Scores From the COA Correspond to the Specific Health Experience(s) the Patient Has Related to the Concept of Interest.	28
G. Scores From the COA Are Sufficiently Sensitive to Reflect Clinically Meaningful Changes Within Patients Over Time in the Concept of Interest Within the Context of Use	30
1. <i>Evaluating Responsiveness to Change</i>	31
2. <i>Evaluating Reliability/Precision</i>	31
H. Differences in the COA Scores Can Be Interpreted and Communicated Clearly in Terms of the Expected Impact on Patients' Experiences.	32
V. ABBREVIATIONS	33
VI. USEFUL REFERENCES FOR SELECTING, MODIFYING, AND DEVELOPING CLINICAL OUTCOME ASSESSMENTS	34
APPENDIX A: PATIENT-REPORTED OUTCOME MEASURES	42
I. INTRODUCTION	42
II. CONCEPTUAL FRAMEWORK EXAMPLE WITH PRO MEASURES	42
APPENDIX B: OBSERVER-REPORTED OUTCOME MEASURES	44
I. INTRODUCTION	44
II. CONCEPTUAL FRAMEWORK EXAMPLE OF AN ObsRO MEASURE	45
APPENDIX C: CLINICIAN-REPORTED OUTCOME MEASURES	47
I. INTRODUCTION	47
II. CONCEPTUAL FRAMEWORK EXAMPLE OF A ClinRO MEASURE	48
APPENDIX D: PERFORMANCE OUTCOME MEASURES	49
I. INTRODUCTION	49
II. CONCEPTUAL FRAMEWORK EXAMPLE OF A PerfO MEASURE	51

Contains Nonbinding Recommendations

Draft — Not for Implementation

**APPENDIX E: EXAMPLE TABLE TO SUMMARIZE RATIONALE AND SUPPORT
FOR A COA 52**

Contains Nonbinding Recommendations

Draft — Not for Implementation

1 **Patient-Focused Drug Development: Selecting, Developing, or**
2 **Modifying Fit-for-Purpose Clinical Outcome Assessments**
Guidance for Industry, Food and Drug Administration Staff, and
Other Stakeholders¹

This draft guidance, when finalized, will represent the current thinking of the Food and Drug Administration (FDA or Agency) on this topic. It does not establish any rights for any person and is not binding on FDA or the public. You can use an alternative approach if it satisfies the requirements of the applicable statutes and regulations. To discuss an alternative approach, contact the FDA staff responsible for this guidance as listed on the title page.

3
4
5 **I. INTRODUCTION**

6
7 **A. Overview of FDA Guidances on Patient-Focused Drug Development**

8
9 This guidance (Guidance 3) is the third in a series of four methodological patient-focused drug
10 development (PFDD) guidance documents² that describe how stakeholders (patients, caregivers,
11 researchers, medical product developers, and others) can collect and submit patient experience
12 data³ and other relevant information from patients and caregivers to be used for medical product⁴
13 development and regulatory decision-making. When finalized, Guidance 3 will represent the
14 current thinking of CDER, CBER, and CDRH on this topic. The topics that each guidance
15 document addresses are described below.
16
17

¹This guidance has been prepared by the Office of New Drugs in the Center for Drug Evaluation and Research, in cooperation with the Center for Biologics Evaluation and Research and the Center for Devices and Radiological Health, at the Food and Drug Administration.

² The four guidance documents that will be developed fulfill FDA commitments under section I.J.1 associated with the sixth authorization of the Prescription Drug User Fee Act (PDUFA VI) under Title I of the FDA Reauthorization Act of 2017 (FDARA). The projected time frames 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 (available at <https://www.fda.gov/downloads/forindustry/userfees/prescriptiondruguserfee/ucm563618.pdf>).

³ "Patient experience data" is defined for purposes of this guidance in Title III, Section 3001 of the 21st Century Cures Act, as amended by section 605 of FDARA, to include data that "(1) are collected by any persons (including patients, family members and caregivers of patients, patient advocacy organizations, disease research foundations, researchers and drug manufacturers); and (2) are intended to provide information about patients' experiences with a disease or condition, including (A) the 'impact (including physical and psychosocial impacts) of such disease or condition or a related therapy or clinical investigation; and (B) patient preferences with respect to treatment of the disease or condition."

⁴ For purposes of this guidance a "medical product" refers to a drug (as defined in section 201 of the Federal Food, Drug, and Cosmetic Act (21 U.S.C. 321)) intended for human use, a device (as defined in such section 201) intended for human use, a biological product (as defined in section 351 of the Public Health Service Act (42 U.S.C. 262)).

Contains Nonbinding Recommendations

Draft — Not for Implementation

- 18 • Methods to collect patient experience data that are accurate and representative of the
19 intended patient population (Guidance 1)⁵
20 • Approaches to identifying what is most important to patients with respect to their
21 experience as it relates to burden of disease/condition and burden of treatment
22 (Guidance 2)
23 • Approaches to selecting, modifying, developing, and validating clinical outcome
24 assessments (COAs) to measure outcomes of importance to patients in clinical trials
25 (Guidance 3)
26 • Methods, standards, and technologies to collect and analyze COA data for regulatory
27 decision-making, including selecting the COA-based endpoint and determining
28 clinically meaningful change in that endpoint (Guidance 4)
29

30 Please refer to **Guidance 1** and other FDA guidances⁶ for additional information on patient
31 experience data.
32

33 In conducting research that involves accessing patient experience data or directly engaging with
34 patients, it is important to carefully consider Federal, State, and local laws and institutional
35 polices for protecting human subjects and reporting adverse events. For additional information
36 about human subjects protection, refer to **section IV.A.2 of Guidance 1**.
37

38 FDA encourages stakeholders to interact early with FDA and obtain feedback from the relevant
39 FDA review division when considering collection of patient experience data related to the
40 burden of disease and treatment.⁷ FDA recommends that stakeholders engage
41 with patients and other appropriate subject matter experts (e.g., qualitative researchers, clinical
42 and disease experts, survey methodologists, statisticians, psychometricians, patient
43 preference researchers) when designing and implementing studies to evaluate the burden of
44 disease and treatment, and perspectives on treatment benefits and risks.
45

46 The contents of this document do not have the force and effect of law and are not meant to bind
47 the public in any way, unless specifically incorporated into a contract. This document is intended
48 only to provide clarity to the public regarding existing requirements under the law. FDA
49 guidance documents, including this guidance, should be viewed only as recommendations, unless
50 specific regulatory or statutory requirements are cited. The use of the word *should* in Agency
51 guidance means that something is suggested or recommended, but not required.”

⁵ See the guidance for industry, FDA staff, and other stakeholders *Patient-Focused Drug Development: Collecting Comprehensive and Representative Input* (June 2020). We update guidances periodically. For the most recent version of a guidance, check the FDA guidance web page at <https://www.fda.gov/RegulatoryInformation/Guidances/default.htm>.

⁶ See FDA guidance for industry *Patient Preference Information—Voluntary Submission, Review in Premarket Approval Applications, Humanitarian Device Exemption Applications, and De Novo Requests, and Inclusion in Decision Summaries and Device Labeling* (August 2016), or subsequent guidances in the PFDD series, when available.

⁷ In addition to the general considerations discussed in this guidance, a study may need to meet specific statutory and regulatory standards governing the collection, processing, retention, and submission of data to the FDA to support regulatory decisions regarding a marketed or proposed medical product. This guidance focuses on more general considerations that apply to many types of studies, and you should consult with the review division and applicable guidance regarding any other applicable requirements.

Contains Nonbinding Recommendations

Draft — Not for Implementation

52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84

B. Purpose and Scope of the Guidance

This document provides guidance that is generally applicable to COAs, including patient-reported outcome (PRO), observer-reported outcome (ObsRO), clinician-reported outcome (ClinRO), and performance-based outcome (PerfO) measures. Appendices A, B, C, and D include additional considerations for each type of COA, respectively, and multiple illustrations of conceptual frameworks.

This guidance is intended to help sponsors use high quality measures⁸ of patients' health in medical product development programs. Ensuring high quality measurement is important for several reasons: measuring what matters to patients; being clear about what was measured; appropriately evaluating the effectiveness, tolerability, and safety of treatments; and avoiding misleading claims. Such findings may help support regulatory decision-making in a variety of contexts. For example, findings measured by a well-defined and reliable COA in an appropriately designed and conducted investigation generally can be used to support a claim in required medical product labeling if the claim is consistent with the findings and the COA's documented measurement capabilities.⁹

The overall structure of this guidance is as follows:

- Overview of COAs in clinical trials, including:
 - Describing the four types of COAs
 - Specifying what a COA assesses (the concept of interest)
 - Specifying the purpose and context of the COA's assessment (the context of use)
 - Determining whether a COA has sufficient evidence to support its context of use, or is fit-for-purpose (BEST (Biomarkers, Endpoints and Other Tools) Resource 2016)
- A general process, referred to as a Roadmap to patient-focused outcome measurement, that sponsors and COA developers may consider as they select, modify, or develop a COA
- A discussion of components of a well-supported rationale to justify the COA's ability to assess the concept of interest for a specified context of use

⁸A measure is a means to capture data (e.g., a questionnaire) plus all the information and documentation that supports its use. Generally, that includes clearly defined methods and instructions for administration or responding; a standard format for data collection; and well-documented methods for scoring, analysis, and interpretation of results in the target patient population.

⁹ The considerations addressed in this guidance may be relevant to a variety of regulatory decisions that require an assessment of benefit or risk, including but not limited to: drug approval decisions under the standards in section 505(d) of the FD&C Act and regulations in 21 CFR 314; device approval decisions under the standards in sections 513(a)(2) and 515(d) and regulations in 21 CFR part 814; device classification decisions under the standards in sections 513(a)(2) and 513(f) and regulations in 21 CFR parts 807 and 860; investigational new drug and investigational device exemption applications under sections 21 CFR parts 312 and 812; REMS and PMR requirements under sections 505-1 and 505(o)(3) and device post-approval requirements under 21 CFR part 814 subpart E; labeling decisions under 21 CFR parts 201, 801, and 809. Necessarily, this guidance does not attempt to capture all of the regulatory standards that might apply to a sponsor's intended plan of study; sponsors should consult the relevant review division(s) as necessary to discuss their study plans and are responsible for satisfying applicable requirements.

Contains Nonbinding Recommendations

Draft — Not for Implementation

85 This guidance is informed by developments in research and applications of COAs to derive
86 clinical trial endpoints that have occurred since the release of the guidance for industry *Patient-*
87 *Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims*
88 (December 2009) (2009 PRO guidance):¹⁰

- 89 • Patients and caregivers have been increasingly integrated as stakeholders in the
90 development and evaluation of medical products.
- 91 • Several best-practice publications have described recommendations for developing and
92 evaluating COAs, as well as analyzing and reporting COA data. Readers are directed to
93 relevant publications throughout this guidance.
- 94 • The growing need for FDA guidance regarding all types of COAs has motivated the
95 broader scope of the PFDD guidance series compared to the 2009 PRO guidance.
- 96 • The framework discussed in this guidance for development of well-constructed measures
97 is based on developing evidence-based rationales. Several publications have described
98 the development of evidence-based rationales (American Educational Research
99 Association et al. 2014; Kane 2013; Weinfurt 2021). This modern validity framework is
100 useful for discussing the broad range of COAs addressed by this guidance and helps to
101 clarify evidence that may be appropriate to support the rationale for using a particular
102 COA.

103
104 This guidance distinguishes an endpoint from the COA, and the score produced by that COA.
105 The COA includes any instructions, administration materials, content, formatting, and scoring
106 rules. A COA score refers to any numeric or rated values generated by a COA through a
107 standardized process. For example, a score could refer to:

- 108 • A response to a specific item (an individual question, statement, or task that is evaluated
109 or performed by the patient to address a particular concept) on a PRO measure
- 110 • A rating assigned by a clinician (as part of a ClinRO measure) or observer (as part of an
111 ObsRO measure) describing a patient's functioning
- 112 • The result from a performance test, such as grip strength measured in kilograms
- 113 • A combination of item responses assumed to measure some *domain* (a sub-concept
114 represented by responses to a subset of items or tasks from a COA that measures a larger
115 concept; such a COA would comprise multiple domains)
- 116 • A combination of scores from multiple domains to reflect some larger concept

117
118 A COA might produce more than one type of score, especially if the COA is designed to
119 measure more than one concept. In contrast to a COA score, an endpoint is a precisely defined
120 variable intended to reflect an outcome of interest that is statistically analyzed to address a
121 particular research question. A complete definition of an endpoint typically specifies the type of
122 assessments made; the timing of those assessments; the assessment tools used; and possibly other
123 details, as applicable, such as how multiple assessments within an individual are to be combined
124 (see Guidance 4, when available, for a discussion of COA-based endpoints).

125
126

¹⁰We update guidances periodically. For the most recent version of a guidance, check the FDA guidance web page at <https://www.fda.gov/regulatory-information/search-fda-guidance-documents>.

Contains Nonbinding Recommendations

Draft — Not for Implementation

127 **II. OVERVIEW OF COAs IN CLINICAL TRIALS**

128

129 **A. Types of COAs**

130

131 A COA is a measure that describes or reflects how a patient feels, functions, or survives. (Note
132 that although clinical events, including death, require a clinician’s judgment and might be
133 considered ClinROs, they are not discussed further in this guidance. The remainder of this
134 guidance focuses on COAs intended to provide insight into how patients feel and/or function.)
135 COA scores can be used to support efficacy, effectiveness, and safety in the context of a clinical
136 trial to determine the clinical benefit(s) and risks(s) of a medical product. There are four types of
137 COAs and choosing which type(s) of COA to use is driven by the concept(s) of interest to be
138 measured and the context in which it will be applied (the context of use). More than one type of
139 COA can be used in a clinical trial to capture the patient experience and the status of the patient’s
140 disease or condition.

141 The following are the four types of COAs:

- 142 • **Patient-reported outcome** (PRO) measures (Appendix A)
 - 143 ○ Reports come directly from the patient
 - 144 ○ Useful for assessment of symptoms (e.g., pain intensity, shortness of breath),
 - 145 functioning, events, or other aspects of health from the patient’s perspective
- 146 • **Observer-reported outcome** (ObsRO) measures (Appendix B)
 - 147 ○ Reports come from someone other than the patient or a health professional (e.g., a
 - 148 parent or caregiver) who has opportunity to observe the patient in everyday life
 - 149 ○ Useful when patients such as young children cannot reliably report for
 - 150 themselves, or to assess observable aspects related to patients’ health (e.g., signs,
 - 151 events, or behaviors)
- 152 • **Clinician-reported outcome** (ClinRO) measures (Appendix C)
 - 153 ○ Reports come from a trained health-care professional using clinical judgment
 - 154 ○ Useful when reports of observable signs, behaviors, clinical events, or other
 - 155 manifestations related to a disease or condition benefit from clinical judgment
- 156 • **Performance outcome** (PerfO) measures (Appendix D)
 - 157 ○ A measurement based on standardized task(s) actively undertaken by a patient
 - 158 according to a set of instructions

159

160 Another type of measure—a proxy-reported outcome measure—is discouraged by FDA. A
161 proxy-reported measure is an assessment in which someone other than the patient reports on
162 patient experiences as if the individual were the patient. FDA acknowledges that there are
163 instances when it is impossible to collect valid and reliable self-report data from the patient. In
164 these instances, it is recommended that an ObsRO measure be used to assess the patient’s
165 behavior rather than a proxy-reported measure to report on the patient’s experience.

166

167 There has been a rapid evolution in digital health technologies (DHTs), which can be used to
168 collect health care-related data from study participants in clinical trials. A DHT is a system that
169 uses computing platforms, connectivity, software, and/or sensors for health care and related uses.
170 This may include use as a measurement tool for COAs in clinical investigations. Refer to the
171 FDA draft guidance for industry *Use of Digital Health Technologies for Remote Data*

Contains Nonbinding Recommendations

Draft — Not for Implementation

172 *Acquisition in Clinical Investigations* (December 2021)¹¹ for more detailed discussion and
173 recommendations on the use of DHTs in clinical investigations.

174
175 Sometimes composite measures are used that combine the scores from several COAs (or several
176 COAs and biomarkers) into a single score. Discussion of these composite measures is beyond the
177 scope of this guidance. For discussion of composite endpoints in CDER and CBER decision-
178 making, see the draft guidance for industry *Multiple Endpoints in Clinical Trials* (January
179 2017).¹²

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

181
182
183 To precisely describe a COA in the context of a clinical study, sponsors should propose to FDA
184 how they intend to interpret scores from a COA (i.e., what they believe the score measures), how
185 scores will be used, and the context in which scores will be used. In other words, the sponsor's
186 proposal should explicitly reference the concept of interest and the context of use, which are
187 discussed below. Each proposal should reference a specific score because a measure may
188 produce multiple scores

1. The Concept of Interest

189
190
191
192 The concept of interest is the aspect of an individual's experience or clinical, biological,
193 physical, or functional state that the assessment is intended to capture (reflect). Depending on the
194 intervention, the intent of treatment may be to improve a symptom(s) or a specific function (e.g.,
195 ambulation); avoid further worsening of a symptom(s) or further loss of a specific function; or
196 prevent the onset of a symptom or a loss of a specific function. Sponsors might also want to
197 assess whether aspects of how patients feel and/or function could be negatively impacted by
198 receipt of the intervention (i.e., harms). All aspects of health that might be meaningfully affected,
199 positively or negatively, by the medical product could be concepts of interest. The identification
200 of concepts of interest appropriate for a given target patient population in CDER and CBER
201 decision-making is described in Guidance 2 of this series, the draft guidance for industry, FDA
202 staff, and other stakeholders *Patient-Focused Drug Development: Methods to Identify What Is*
203 *Important to Patients* (October 2019) (PFDD Guidance 2).¹³ For some diseases/conditions,
204 important concepts of interest might have already been developed and used in studies based on
205 input from patients, caregivers, clinical experts, and other sources. In such cases, sponsors should
206 reference and summarize the prior work done when justifying their choice of concept(s) of
207 interest.

208
209 In a clinical trial, it is important to carefully select concepts that, when measured adequately:

- 210 • Reflect an aspect of health that is important to patients
- 211 • Have the ability to be modified by the investigational treatment
- 212 • Could demonstrate clinically meaningful differences between study arms within the time
- 213 frame of the planned clinical trial

214

¹¹ When final, this guidance will represent FDA's current thinking on this topic for applicable medical products.

¹² When final, this guidance will represent FDA's current thinking on this topic for applicable medical products.

¹³ When final, this guidance will represent FDA's current thinking on this topic for applicable medical products.

Contains Nonbinding Recommendations

Draft — Not for Implementation

215 Often, a single disease or condition is associated with many concepts. For example, a condition
216 that causes chronic pain may also be associated with fatigue and impact on physical and social
217 functioning. To help focus a medical product development program, sponsors should identify the
218 primary manifestations of a disease or condition (i.e., core concepts of a disease or condition).
219 Other important concepts might represent the downstream impact of these core concepts on other
220 aspects of how a patient feels or functions.

221
222 For example, when evaluating a treatment for the management of moderate to severe
223 endometriosis-associated pain, it may be important to assess a core concept such as dyspareunia,
224 defined as pain with intercourse. In addition, to further evaluate clinical benefit, a strategy to
225 assess the impact of moderate to severe endometriosis-associated pain severity on daily activities
226 could also be assessed.

227
228 In addition to selection of the concept(s) of interest, the aspect(s) of the concept(s) of interest that
229 will be assessed should also be considered. Aspects might include presence/absence, frequency,
230 intensity, worst experience, and for concepts of interest reflecting a patient’s functioning the
231 amount of difficulty experienced or level of assistance needed. Patient and/or caregiver input can
232 be used to identify which aspect(s) of a concept is most impactful for patients. This input will
233 help sponsors in selecting or developing a COA that measures what is important to patients.

234
235 A conceptual model can be useful for representing patients’ specific health experiences that
236 result from their disease/condition, the health concepts that describe those specific
237 experiences,¹⁴ and the concept(s) of interest selected for assessment. For example, Figure 1
238 displays a hypothetical, conceptual model underlying activities of daily living (ADLs) as the
239 concept of interest. In the figure, specific health experiences of the patient (Activities 1-15) are
240 conceptualized in terms of five different health concepts—hygiene, continence, dressing,
241 feeding, and mobility. For example, the activities collected under the concept “mobility” might
242 include getting in and out of bed, being able to stand from a sitting position, and walking across a
243 room. The five health concepts together make up a more general health concept known as ADLs,
244 which the sponsor has selected as the concept of interest that will be assessed using a COA.¹⁵
245 Such a conceptual model can be helpful to sponsors and FDA for communicating about the
246 concept to be measured and for determining whether a proposed COA captures the entirety of a
247 concept of interest.

248

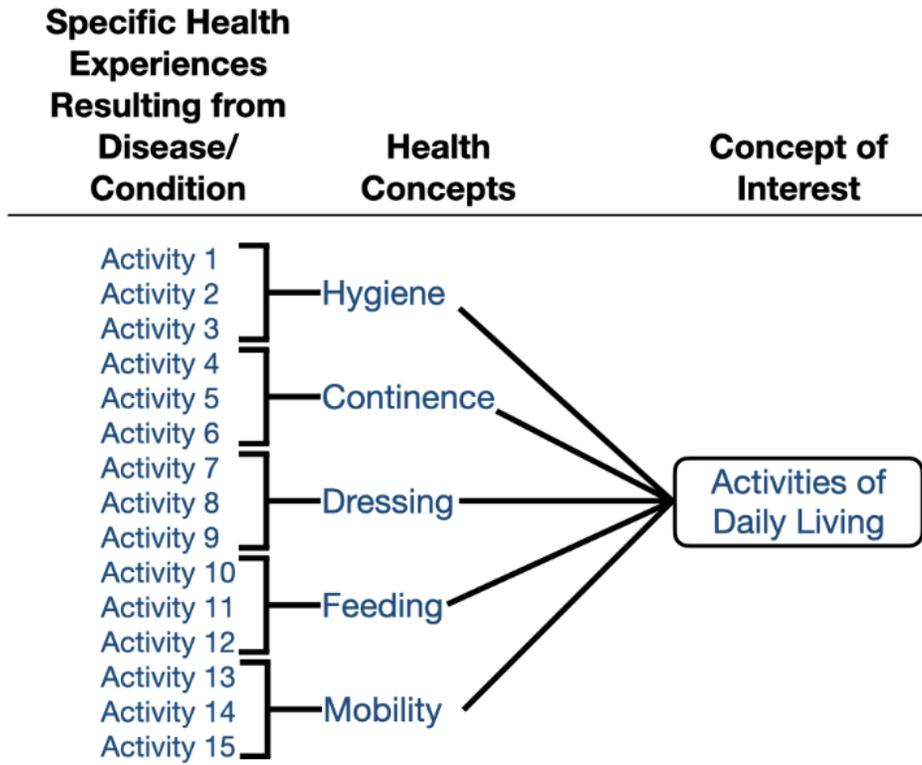
¹⁴ Note that what is referred to as a *health concept* in this guidance is the same as what Walton et al. (2015) refer to as a *meaningful health aspect*. The former term is used to avoid confusion that might arise from multiple uses of *aspect*.

¹⁵ Here ADLs are both a higher-level health concept that includes the lower-level health concepts, such as hygiene, continence, as well as being the health concept chosen as the measured concept of interest.

Contains Nonbinding Recommendations

Draft — Not for Implementation

249 **Figure 1. Hypothetical Conceptual Model for Activities of Daily Living**



250

251

2. *The Context of Use*

252

253

254

255

256

257

258

The context of use should clearly specify the way COA scores will be used as the basis for an endpoint, including the purpose for their use in a medical product development program. The appropriateness of a COA is evaluated within the proposed context of use.

Context of use considerations may include the following:

259

260

261

262

263

264

265

266

267

268

269

270

271

272

- **Use of the COA:** Clinical trial objectives and how the COA will be used to support COA-based endpoints (e.g., computing the mean COA score at 12 weeks)
- **Target Population:** Including a definition of the disease or condition; participant selection criteria for clinical trials (e.g., baseline symptom severity, patient demographics, comorbidities); and expected patient experiences or events during the trial (e.g., that some patients will require assistive devices)
- **Study Context:** The clinical trial design in which the COA is to be used, including the type of comparator group and whether those providing responses or participating in the tasks for the COA (patients, observers, clinicians, trained raters) are masked with respect to treatment assignment and/or study visit)
- **Timing** of when assessment(s) of the COA is conducted
- **COA Implementation:** Including the site for COA collection (e.g., inpatient hospital, outpatient clinic, home); how the COA will be collected (e.g., DHT, paper form); and by whom (e.g., patient, study coordinator, investigator, parent/caregiver.)

Contains Nonbinding Recommendations

Draft — Not for Implementation

273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316

C. Deciding Whether a COA Is Fit-for-Purpose

A COA is considered fit-for-purpose when “the level of validation associated with a medical product development tool is sufficient to support its context of use” (BEST (Biomarkers, Endpoints and Other Tools) Resource, 2016). Whether a COA is fit-for-purpose is determined by the strength of the evidence in support of interpreting the COA scores as reflecting the concept of interest within the context of use. Fit-for-purpose in the regulatory context means the same thing as *valid* within modern validity theory, i.e., validity is “the degree to which evidence and theory support the interpretations of test scores for proposed uses of tests” (American Educational Research Association et al. 2014).

Decisions about whether a COA is fit-for-purpose are based on two considerations:

1. The Concept of Interest and Context of Use Are Clearly Described

Section III.C describes what constitutes a clear statement of the intended interpretation of COA scores as measures of the concept of interest within the context of use. The statement should explicitly specify the concept of interest and the context of use in enough detail to describe clearly how the COA is intended to be used.

2. There Is Sufficient Evidence to Support a Clear Rationale for the Proposed Interpretation and Use of the COA

Regardless of whether sponsors propose to use an existing COA, a modified COA, or a newly developed COA, sponsors should present a well-supported rationale for why the proposed COA should be considered fit-for-purpose. The rationale is a set of reasons supported by evidence.

The rationale may have multiple components (see section IV, Table 1) and each component should be justified by one or more sources of evidence, including for example literature reviews; natural history studies; qualitative studies with patients, caregivers, or other stakeholders; and quantitative studies.

To determine whether sufficient justification has been provided for the rationale, FDA will review each part of the rationale and assess whether an appropriate type and amount of evidence has been presented. The evidence for a particular part of the rationale is weighed relative to the degree of uncertainty about that part. The greater the uncertainty, the greater the need for additional evidence to support that part of the rationale. In addition to the degree of uncertainty about each part of the rationale, FDA considers the context of use, and may consider the broader impact on the target patient population and medical product development of collecting additional evidence (Leptak et al., 2017), when determining whether a COA is fit-for-purpose. Section IV provides guidance about how to develop a clear rationale with supporting evidence.

Contains Nonbinding Recommendations

Draft — Not for Implementation

317 **III.A ROADMAP TO PATIENT-FOCUSED OUTCOME MEASUREMENT IN**
318 **CLINICAL TRIALS**

319
320 This section describes a general Roadmap to patient-focused outcome measurement in clinical
321 trials (see Figure 2). Sponsors and COA developers are not required to use this approach, and it
322 may not fit every development program, but it has worked well for a number of COAs. FDA
323 recommends sponsors seek FDA input as early as possible and throughout medical product
324 development to ensure COAs are appropriate for the intended context of use.

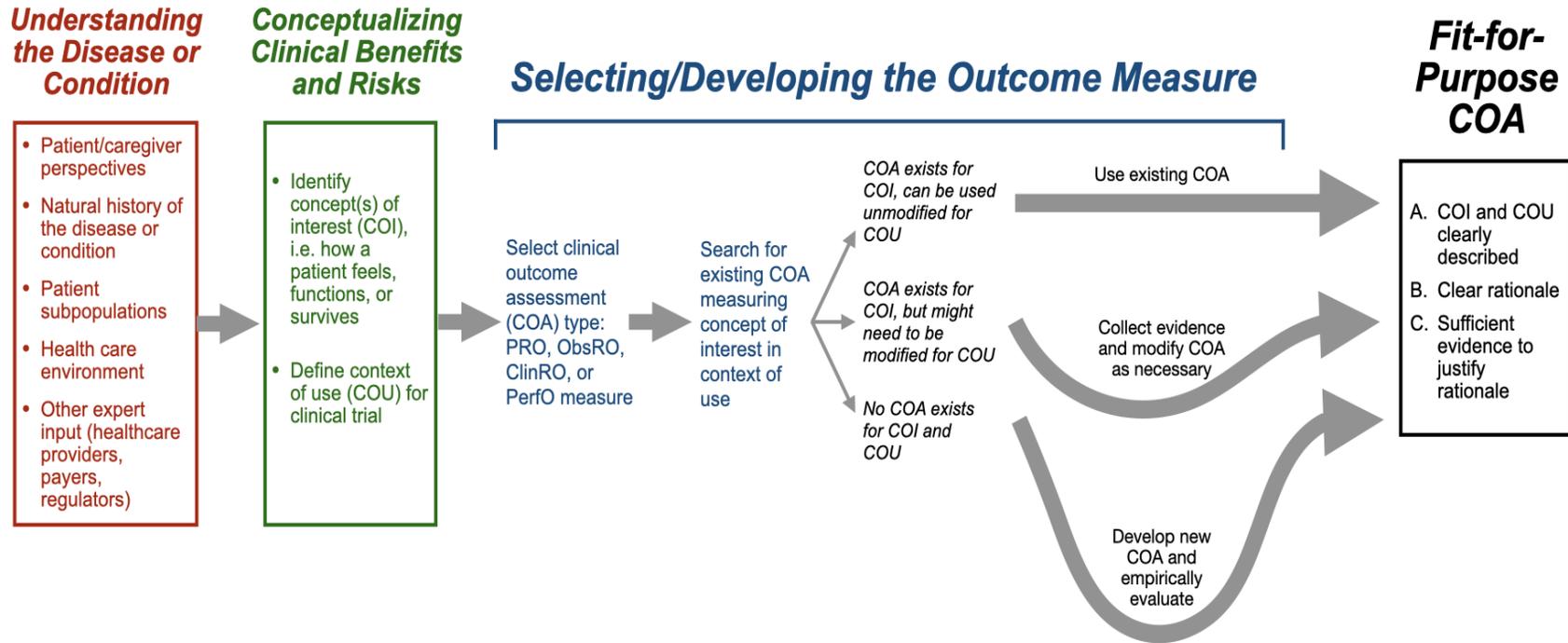
Contains Nonbinding Recommendations

Draft — Not for Implementation

325

326 **Figure 2: Roadmap to Patient-Focused Outcome Measurement in Clinical Trials**

327



328

Contains Nonbinding Recommendations

Draft — Not for Implementation

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

1. Understanding the Disease or Condition

The first step involves considering the manifestations and natural history of the disease or condition; important patient subpopulations; the clinical environment in which patients with the condition seek care; and patient and/or caregiver perspectives on the disease, its impacts, and therapeutic needs and priorities. One important outcome of this step is understanding and summarizing the important signs, symptoms, and health impacts patients with the disease or condition might experience.

2. Conceptualizing Clinical Benefits and Risks

The next step involves considering which aspect(s) of the patient’s experience with the disease/condition and/or its treatment will be targeted by the medical product. This consideration leads to identifying the concept(s) of interest (see section II.B.1) and context of use, including the population of interest, clinical trial design, and the trial objective and endpoints.

A conceptual model can be used to support the first two parts of the Roadmap. When little is known about a patient population and/or their health experiences, a hypothesized conceptual model can be developed based on literature review and/or expert clinical input. Then qualitative research with patients and/or caregivers can be conducted to evaluate and, if necessary, modify the conceptual model (see PFDD Guidance 2 and Patrick et al. 2011a). Note for relatively simple and narrow concepts, such as presence of itch, a simple definition might suffice without a more elaborate conceptual model. However, for more complex health experiences, we recommend a clear and detailed conceptual model for subsequent steps of the Roadmap. A conceptual model comprises one component of a conceptual framework (see section III.C).

B. Select/Develop the Outcome Measure

There are several steps involved in selecting or, if necessary, developing a COA to measure the concept of interest.

1. Selecting the COA Type

Sponsors and measure developers should consider what type of COA is most appropriate for assessing the concept of interest in the context of use. Considerations for selecting a specific type of COA are discussed in section II.A. and in Appendices A-D. Sometimes multiple COA types may be used to measure the concept of interest.

2. Evaluating Existing and Available COAs Measuring the Concept of Interest in the Context of Use

Contains Nonbinding Recommendations

Draft — Not for Implementation

373 FDA recommends conducting a search to identify a COA that measures the concept of interest in
374 the intended context of use and is available for use.¹⁶ Existing COA measures for which there is
375 already experience in the relevant context of use are generally preferred, particularly when
376 measuring well-established concepts (e.g., pain intensity). Sponsors can identify potential
377 measures by searching the scientific literature; repositories of measures, including item banks
378 comprising previously developed and tested items; and other resources [FDA COA Qualification
379 Program, 2021; FDA Medical Device Development Tools (MDDT), 2021]. When searching for
380 existing COAs, the conceptual model for the concept of interest can be used to assess whether an
381 existing measure addresses the full content of the concept of interest.

382

383 There are several possible outcomes of conducting this search.

384

385 a. An Appropriate COA Exists for the Concept of Interest in the Same Context
386 of Use: Use Existing COA

387

388 If a COA exists to assess the concept of interest in the same context of use as intended in the
389 sponsor's trial, the sponsor should assess its sufficiency; provide the rationale for selection of the
390 COA; and summarize the evidence that supports that rationale (such as details on the prior
391 experience with this COA, especially prior studies in which the COA was used, and evidence of
392 how well it performed).

393

394 There are times when an existing COA may not have all the evidence recommended to support
395 its use because the COA is still under development or was developed a long time ago, or for
396 other reasons. For example, some types of studies (such as an assessment of test-retest reliability)
397 may not have been conducted or some documentation may not be available for some steps in the
398 development. Sponsors should summarize all existing information and evidence that supports the
399 rationale for the use of the COA and assess how well the rationale is supported by the available
400 information. In some instances, adequate evidence may be found in the literature or available
401 clinical trial data, while in other instances, it may be necessary to collect additional evidence for
402 the rationale before the COA can be considered fit-for-purpose.

403

404 COAs being used in registries, natural history studies, or observational trials may or may not be
405 fit-for-purpose in other contexts of use. Sponsors should ensure that there is sufficient evidence
406 to support the use of such COAs within the intended context of use in the planned clinical trial.

407

408 b. A COA Exists for the Concept of Interest for a Different Context of Use:
409 Collect Additional Evidence and Modify COA as Necessary

410

411 If a COA exists that assesses the concept of interest but was not developed for the sponsor's
412 context of use (e.g., was not developed for the same target patient population), then the sponsor
413 should evaluate whether the COA can be used in the different context of use and provide
414 supporting evidence or explanations supporting the new context of use. Evidence presented in
415 prior work on the COA may suffice to support the rationale for its use in the new context of use.

¹⁶ FDA encourages the sharing of COAs among sponsors and researchers to promote efficiency and to maximize the returns on the efforts made by patients who cooperated in their development.

Contains Nonbinding Recommendations

Draft — Not for Implementation

416 Alternatively, if the existing evidence leaves too much uncertainty about the appropriateness of
417 use in the new context of use, we recommend the collection of additional evidence.

418
419 A sponsor may also consider modifications intended to improve the COA’s ability to reflect the
420 concept of interest. Modifications could include, but are not limited to, changes to:

- 421 • Instructions/training materials
- 422 • Item or task content (e.g., omitting, adding, or modifying wording of items and/or
423 response options; translating from one language to another; modifying the activity
424 performed for a PerfO)
- 425 • Order of the items or tasks
- 426 • Recall period
- 427 • Format of the measure (e.g., paper or electronic device)
- 428 • Method of scoring, including changes to the scoring algorithm

429
430 The sponsor should carefully consider the impact of the proposed modifications to an existing
431 COA. Any alteration of the COA could potentially constitute the creation of a new measure and
432 result in altering the measure’s scores and/or their interpretation. Some modifications are
433 unlikely to alter the scores or their interpretation (e.g., changing the display on a tablet-based
434 administration from one item per screen to three items per screen), whereas other changes are
435 likely to affect scores and their interpretation (e.g., changing the recall period from 1 day to 7
436 days). In the latter case, the modification may, in effect, create a new measure. The type of
437 evidence (qualitative and/or quantitative) to support modifications of a COA will depend on the
438 type of changes that are proposed and the way in which the new context of use differs from the
439 one for which the COA was originally developed. Sponsors should support their assessment,
440 with appropriate evidence, that the modified measure adequately measures the concept of interest
441 in the new context of use.

442
443 References are available that address considerations for modifying a COA (see Rothman et al.
444 2009 and Papadopoulos et al. 2019).

c. No COA Exists for the Concept of Interest: Develop a New COA and Empirically Evaluate

448
449 It is beyond the scope of this guidance to provide specific recommendations for developing all
450 types of COAs, but helpful references that address measure development are provided at the end
451 of this guidance (e.g., de Vet et al. 2011; Fayers and Machin 2016). There are general principles
452 regarding the development process for any type of new COA:

- 453
454 • Clearly document all steps and data collected in the development process. For COAs
455 involving multiple items, this includes an item tracking matrix that describes the history
456 of the development and modification of all items.
- 457 • Develop and provide convincing evidence to support the rationale for interpreting COA
458 scores as a measure of the concept of interest in the context of use (discussed in section
459 IV). Support for the rationale includes evaluation of relevant measurement properties.

Contains Nonbinding Recommendations

Draft — Not for Implementation

- 460 • Consider and evaluate potential limitations of the proposed COA. For example, could
461 measurement of the concept of interest be affected by processes or concepts not part of
462 the concept of interest (see section IV.D)?
- 463 • Create a user manual for the COA describing how to administer the measure. For most
464 types of COAs, it is important to create training materials (e.g., for investigators,
465 patients, observers, or clinicians) so that assessments are conducted in a consistent way.
- 466 • Document the method of scoring the COA, including how missing items or tasks should
467 be handled. There should be clear justifications for the approach to scoring and
468 addressing missing data.

469
470 When the sponsor is developing or significantly modifying a COA, FDA does not recommend
471 evaluating measurement properties for the first time in a registration¹⁷ trial, because it may be
472 too late to learn that the COA is not performing as it should, potentially jeopardizing the success
473 of a medical product development program. Earlier trials represent an opportune time to evaluate
474 measurement properties of COAs and sponsors are encouraged to include prospectively planned
475 analyses to inform subsequent trials.¹⁸ If this is not a feasible option, FDA recommends
476 conducting a standalone observational study prior to the initiation of a registration trial(s) to aid
477 in the development of a fit-for-purpose COA measure(s). Furthermore, using data from the
478 observational study to evaluate the psychometric properties and performance of a proposed COA
479 measure prior to the registration trial will reduce the risk of using a COA that may not perform as
480 expected, and therefore may not detect a treatment effect.

481
482 Early in the development process, sponsors are encouraged to request a meeting with FDA to
483 discuss plans for newly developed COAs.

484
485 FDA encourages the sharing of COAs among sponsors and researchers to promote efficiency and
486 to maximize the returns on the efforts made by patients who cooperated in its development.

487 488 3. *Special Considerations for Selecting or Developing COAs for Pediatric Populations*

489
490 If the concept of interest can be reliably measured across the age spectrum of the trial patient
491 population, we recommend using one simple version of a COA for patients of all ages in a study.
492 Including multiple versions of a COA for different age groups in the same trial is generally not
493 recommended because it may introduce unwanted measurement variability. However, depending
494 on the concept of interest, at times it may be necessary to use multiple versions of a COA and/or
495 different COA types to measure a concept, because assessment of the target concept may differ
496 substantially across the age and developmental spectrum (e.g., gross motor functioning in infants
497 and adolescents). Using multiple COAs to measure a concept in a trial impacts statistical analysis
498 plans and trial power (see Guidance 4, when available).

499
500 When pediatric self-administered COAs are feasible, the COAs should be completed by the child
501 independently, without any assistance from caregivers, investigators, or anyone else, to avoid

¹⁷ In this guidance, *registration trials* are used to stand for what different groups call pivotal trials, confirmatory trials, and clinical trials for marketing authorization.

¹⁸ Sponsors should also use data from later clinical trials to confirm, to the extent possible, the measurement properties evaluated in earlier phase trials.

Contains Nonbinding Recommendations

Draft — Not for Implementation

502 influencing the child’s responses. Computer-administration, including automated reading of
503 items, using a touch screen, or games, may make it easier for children to self-report. Self-
504 administration and self-report may not be suitable with very young children and therefore might
505 call for alternative approaches, such as interviewer-administration by a trained interviewer and/or
506 different COA types.

507
508 Young children may be limited in their understanding of certain response scales used in a PRO
509 measure (e.g., a 0 to 10 numeric rating scale, more/less comparison, references to periods of
510 time). Simplified age-appropriate response scales (e.g., scales with few and simple response
511 options, broadly culturally acceptable and interpretable pictorial scales) should be considered for
512 use with young children and may be useful for all ages. Supporting evidence for the suitability of
513 a COA for specific pediatric populations should address age-relevant vocabulary, language
514 comprehension, comprehension of the target concept, and relevance of the recall period.

515
516 References are available that discuss measurement in pediatric patient populations (Arbuckle and
517 Abetz-Webb 2013; Bevans et al. 2010; Matza et al. 2013; Papadopoulos et al. 2013). Also, refer
518 to PFDD Guidance 2, section VI (Managing Barriers to Self-Report) for considerations on how
519 to obtain input from pediatric patients.

520

521 *4. Using DHTs To Collect COA Data*

522

523 DHTs can be used to implement a COA, such as collecting responses to items from a PRO
524 measure or assessing the patient’s activity functioning in a PerfO. As in any COA development,
525 the concept of interest and the context of use must be clearly identified. Early in the clinical
526 development program, based on input from patients and/or caregivers, the sponsor should define
527 and provide rationale to justify the use of the DHT for measuring important feature(s) of the
528 concept of interest in the target population. See the DHT draft guidance *Use of Digital Health*
529 *Technologies for Remote Data Acquisition in Clinical Investigations* (December 2021)¹⁹ for
530 more information about specifying the minimum technical (e.g., operating system, storage
531 capacity, sensors) and performance (e.g., accuracy and precision) specifications.

532

533 *5. COA Accessibility and Universal Design*

534

535 A portion or all of the target patient population may benefit from accessibility features and
536 universal design²⁰ considerations. Usability testing is recommended for accessibility features for
537 a selected COA, along with human factors testing (see Guidance for Industry and FDA Staff,
538 *Applying Human Factors and Usability Engineering to Medical Devices*, 2016, for guidance on
539 CDRH decision-making). The following resources should be reviewed to ensure the COA is
540 accessible for patients with impairments (e.g., vision impairment/low vision, hearing
541 impairment/deaf or hard of hearing):

- 542 • The World Wide Web Consortium (W3) has a Web Accessibility Initiative (WAI) with
543 resources and recommendations for making electronically delivered material more

¹⁹ When final, this guidance will represent FDA’s current thinking on this topic.

²⁰In the context of COAs, *universal design* is consideration for the design and composition of a COA so that it can be accessed, understood, and used to the greatest extent possible by all people, inclusive of people with disabilities.

Contains Nonbinding Recommendations

Draft — Not for Implementation

544 accessible to people, see <https://www.w3.org/WAI/> and [https://www.w3.org/TR/low-](https://www.w3.org/TR/low-vision-needs/)
545 [vision-needs/](https://www.w3.org/TR/low-vision-needs/).

546 • Section 508, a U.S. Government website, has resources addressing universal design,
547 including color universal design, creating accessible portable document formats (PDFs),
548 and other topics: <https://www.section508.gov/>.

549
550 Options including assistive technology that may be used by participants, such as screen readers
551 or eye trackers, can allow patients and/or their caregivers to provide reliable reports. Consider
552 which modifications and/or assistive technologies might be useful to assist broad inclusion in
553 COA development, evidence generation, and trials.

554 555 **C. Developing a Conceptual Framework**

556
557 The Roadmap describes a recommended path sponsors can take to arrive at a fit-for-purpose
558 COA. Sponsors can construct an illustration in the form of a conceptual framework²¹ to
559 demonstrate the results of each step along the Roadmap for the selection of COAs in the clinical
560 trial; this framework is particularly helpful to FDA reviewers.

561
562 A conceptual framework summarizes (1) relevant experiences of patients in the target
563 population, (2) specific concepts of interest targeted for assessment, (3) type(s) of COA proposed
564 for each concept of interest, and (4) a representation of how the particular COA is intended to
565 work in order to generate a score reflecting the concept of interest.

566
567 The conceptual framework includes two important representations:

- 568 • The conceptual model, described in section II.B, which depicts the structure of a concept
569 of interest, including the different aspects of the concept and how they relate to patients’
570 experiences.
- 571 • The measurement model, which represents how a COA is intended to work to generate a
572 score(s) that can be interpreted as a measure of the concept of interest in the context of
573 use. How best to represent the measurement model for a specific COA will depend on the
574 type and complexity of the measure, but most measurement models will include the parts
575 of the COA (e.g., items or tasks) and how they are combined to result in a score(s).

576
577 A conceptual framework can be especially helpful when there is more than one concept of
578 interest and COA. Figure 3 illustrates a generic conceptual framework for a clinical trial in which
579 PRO and ObsRO measures are used to assess three related concepts of interest. Viewing the
580 framework from left to right, the patients in the target population have a variety of specific health
581 experiences that may be affected by their disease or condition, including different symptoms
582 (e.g., feeling tired, dizzy, anxious); behaviors (e.g., scratching, waking up at night); and/or
583 activities (e.g., walking up a flight of stairs, talking while walking). Through qualitative studies
584 with patients and clinical expertise, these specific symptoms, behaviors, and/or activities can be

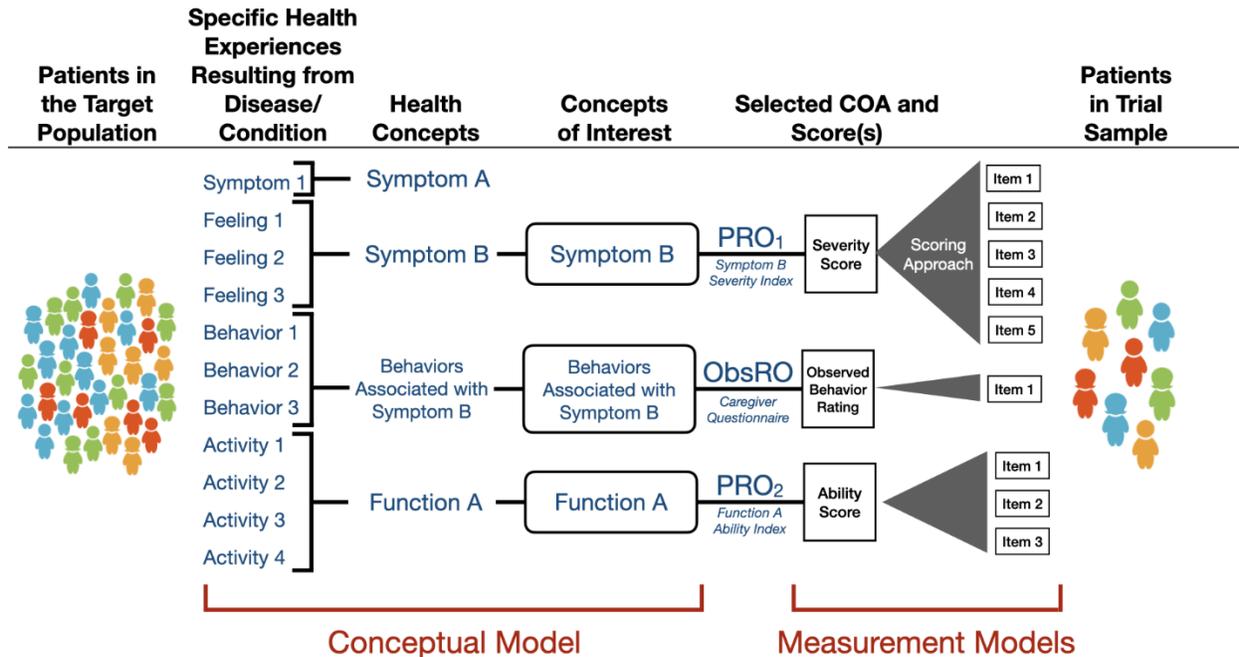
²¹ In the 2009 PRO Guidance, the *conceptual framework* combined a representation of the concept and of the PRO instrument used to measure the concept in a single figure. To accommodate all types of COAs and more complex relationships between health experiences, concepts, and measures, the current guidance’s conceptual framework separates the *conceptual model*, which represents the structure of the concept of interest, from the *measurement model*, which represents how the measure is intended to work in order to measure the concept of interest.

Contains Nonbinding Recommendations

Draft — Not for Implementation

585 documented and summarized under one or more health concepts²² that name the relevant
 586 symptoms, signs, and/or effects on functioning. From among these health concepts, sponsors
 587 select one or more concepts of interest to target for intervention and assessment based on the
 588 importance to patients; the target of the medical product (i.e., mechanism of action, targeted
 589 function); and the feasibility of observing intervention effects within the context of a clinical trial
 590 (e.g., trial duration). Note in Figure 3 that Symptom A is relevant for patients in the target
 591 population but was not chosen by the sponsor as a concept of interest for this trial. A specific
 592 type of COA is then selected to assess each concept of interest, generating a specific score(s)
 593 thought to reflect the concept of interest. Finally, the framework represents the way the measure
 594 is supposed to work to generate a score (i.e., the measurement model; see section IV.E). For
 595 example, a multi-item PRO measure would be represented by the specific items in the measure
 596 and some indication of how the items are combined to arrive at a score. Sponsors can consider
 597 the conceptual framework that best fits their specific development plan.

599 **Figure 3. Illustration of a Generic Conceptual Framework Summarizing Which Patient**
 600 **Experiences Will Be Targeted and How They Will Be Measured**



601
 602 When reading from left to right, the representation provides a high-level view of the thinking
 603 behind the COA strategy—how the experiences of patients in the target population motivate the
 604 selection and measurement of the outcomes of interest. When reading from right to left, the
 605 representation provides an overview of the inference that stakeholders would like to make from
 606 the experiences of the trial participants, expressed as responses to one or more COAs, to the
 607 experiences that would be expected to occur among the larger target population of patients were

²² Note that what is referred to as a *health concept* in this guidance is the same as what Walton et al. (2015) refer to as a *meaningful health aspect*. The former term is used to avoid confusion that might arise from multiple uses of *aspect*.

Contains Nonbinding Recommendations

Draft — Not for Implementation

608 they to receive the investigational medical product(s). More examples of conceptual frameworks
609 pertinent to each type of COA are found in the COA-specific appendices.

610
611 Section III describes a general Roadmap that sponsors might follow to arrive at a fit-for-purpose
612 COA. The next two sections (IV and V) provide more focused guidance on how to construct and
613 support a strong rationale for the proposed interpretation and use of a COA.

614 615 616 **IV. DEVELOPING THE EVIDENCE TO SUPPORT THE CONCLUSION THAT A COA** 617 **IS APPROPRIATE IN A PARTICULAR CONTEXT OF USE**

618
619 Evidence collected in support of the use of a COA should support the rationale that explains how
620 and why the specific COA is expected or intended to work. It is important for FDA to understand
621 each part of a sponsor’s rationale and the evidence being offered in support of each part. This
622 understanding facilitates conversations between FDA and sponsors or measure developers.

623
624 This section describes eight components (see Table 1) that should be considered for inclusion in
625 the rationale and supporting evidence or justification section of submissions to FDA. The
626 discussion below also includes possible sources of evidence to evaluate each component.
627 Different trials and contexts of use might call for different rationale components and/or evidence
628 to support a COA as fit-for-purpose. Note that some types of studies might supply evidence to
629 support more than one component. For example, a qualitative study using cognitive interviews
630 involves asking patients how they understand items from a COA and arrive at their responses
631 (Willis 2005 and Willis 2015) or asking patients and assessors how they interpret instructions for
632 a PerfO. Data from such a study might be used to support components C, D, and F in Table 1.

633
634 **Table 1. Eight Components Comprising an Evidence-Based Rationale for Proposing a**
635 **COA as Fit-for-Purpose**

A	The concept of interest should be assessed by [<i>COA type</i>] because . . .
B	The COA measure selected captures all the important aspects of the concept of interest.
C	Respondents understand the instructions and items/tasks of the measure as intended by the measure developer.
D	Scores of the COA are not overly influenced by processes/concepts that are not part of the concept of interest.
E	The method of scoring responses to the COA is appropriate for assessing the concept of interest.
F	Scores from the COA correspond to the specific health experience(s) the patient has related to the concept of interest.
G	Scores are sufficiently sensitive to reflect clinically meaningful changes within patients over time in the concept of interest within the context of use.
H	Differences in COA scores can be interpreted and communicated clearly in terms of the expected impact on patients’ experiences.

636 *Note:* Listed components are those that are likely but not necessarily needed in the rationale for a specific COA,
637 concept of interest, and context of use. Each rationale can be tailored to the proposed interpretation and use. Each
638 component should be accompanied by comprehensive supporting evidence and justification.

639

Contains Nonbinding Recommendations

Draft — Not for Implementation

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

640
641
642 The sponsor should provide a clear rationale for the choice of type of COA (i.e., PRO, ObsRO,
643 ClinRO, or PerfO) selected to assess the concept of interest. Considerations for selecting the
644 specific type of COA are discussed in Section II.A and Appendices A-D. Note that more than
645 one type of COA might be used to assess different aspects of a concept of interest. For example,
646 a functional outcome could be assessed by a combination of a PRO measure and a PerfO
647 measure for a particular context of use. In such cases, a separate rationale should be provided for
648 each measure.

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

649
650
651
652
653 All important aspects of the concept of interest should be covered by the chosen COA.²³ This
654 includes the specific attribute(s) of interest, such as frequency, intensity, or duration. For narrow
655 and simple concepts that can be assessed with a single item (e.g., asking patients to record how
656 many times they woke up to urinate at night to measure nocturia), it is straightforward to see
657 whether the item content covers the concept of interest. For more complex concepts of interest
658 that include multiple aspects (for example, physical function), all important aspects should be
659 reflected in the content of the COA, or else the concept of interest will only be partly assessed.
660 Similarly, the tasks included in a PerfO should cover all important aspects of the function being
661 evaluated as the concept of interest. The conceptual framework (section III.D) can show how the
662 COA (represented by its measurement model) addresses all important aspects of the concept of
663 interest (represented by its conceptual model).

C. Respondents Understand the Instructions and Items/Tasks of the Measure as Intended by the Measure Developer.

664
665
666
667
668 For PRO, ObsRO, and ClinRO measures, the most straightforward type of support for
669 component C is in the form of cognitive interviews—individual qualitative interviews in which
670 the participants discuss how they understand and respond to each of the components comprising
671 the measure (e.g., their understanding and interpretation of instructions and items in a PRO
672 measure) (Willis 2005, Willis 2015, and Patrick et al. 2011b). For PerfO measures, cognitive
673 interviews with patients regarding task instructions combined with pilot testing tasks can confirm
674 whether patients understand the task they are asked to do, and whether they are able to perform
675 that task.

676
677 We also recommend that measure developers follow good practices in questionnaire design to
678 avoid common pitfalls that could interfere with respondent understanding (e.g., avoiding double-

²³ How well a measure reflects all important aspects of a concept of interest was previously referred to as *content validity* in the 2009 PRO Guidance. The field of measurement, as reflected by the 2014 Standards for Psychological and Educational Testing, has moved from talking about different types of validity to specifying different sources of evidence. Validity is understood as a unitary concept and refers to the “degree to which evidence and theory support the interpretations of test scores for proposed uses of tests” (American Educational Research Association et al. 2014, p. 11), where tests in this case refer to COAs.

Contains Nonbinding Recommendations

Draft — Not for Implementation

679 barreled items, which ask about more than one thing within a single item) (see section IV in
680 PFDD Guidance 2).

681
682 **D. Scores of the COA Are Not Overly Influenced by Processes/Concepts That Are**
683 **Not Part of the Concept of Interest.**

684
685 In a well-designed measure, it is the concept of interest that predominantly affects a patient's
686 responses to items or tasks. Thus, sponsors or measure developers should consider the most
687 likely interfering influences on responses to items or tasks and assess the presence and strength
688 of those influences.

689
690 What follows are examples of some of the most likely sources of interfering influence. When a
691 statement may only be relevant to certain COA types, those types are listed in brackets. Sponsors
692 should consider whether there are additional factors (e.g., differences in use/access to health care
693 related to location, income) not listed here that may influence scores on the COAs being used.

694
695 *1. Item or Task Interpretations or Relevance Does Not Differ Substantially According to*
696 *Respondents' Demographic Characteristics (Including Sex, Age, and Education*
697 *Level) or Cultural/Linguistic Backgrounds.*

698
699 Sponsors and measure developers should consider whether there are any demographic groups for
700 whom items might be interpreted differently or tasks might have different relevance and, if so,
701 evaluate potential differences between groups using qualitative (e.g., cognitive interviews)
702 and/or quantitative methods (e.g., measurement invariance) as appropriate.

703
704 For some trials, COA instruments are used for patients with diverse linguistic and cultural
705 backgrounds. Therefore, it is important to show that such differences are unlikely to influence
706 response to COA items. It is recommended that translation, cultural adaptation assessment, and
707 linguistic validation are conducted early in the COA selection and development process
708 following good practice methodology (Eremenco et al. 2017; McKown et al. 2020; Wild et al.
709 2005). One approach is to describe in detail the process of language translation and/or cultural
710 adaptation (including cognitive interviews) to support the quality of the resulting translation
711 and/or adaptation. A robust process of translation and/or cultural adaptation increases confidence
712 that all trial participants, regardless of their language and/or cultural backgrounds, understand the
713 measure's instructions, items or tasks, and response options similarly.

714
715 For some types of multi-item measures, , one could also present evidence of measurement
716 invariance, including differential item functioning (DIF) (Teresi et al. 2009). Such evidence
717 could demonstrate that item responses provided by respondents from different demographic,
718 linguistic, or cultural backgrounds can be interpreted and scored using the same statistical model.
719 Such studies typically use larger sample sizes (e.g., at least 200 patients per group being
720 compared (Scott et al. 2009)). Before embarking on a large DIF study, sponsors and measure
721 developers might evaluate whether differences for particular items (considering the likely extent
722 of demographic, linguistic, and cultural effects on the item response) will be large enough
723 between groups to substantially change scores in a way that will affect the COA-based trial
724 endpoint.

Contains Nonbinding Recommendations

Draft — Not for Implementation

725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769

2. Recollection Errors Do Not Overly Influence Assessment of the Concept of Interest. [PRO, ObsRO, and ClinRO Measures]

For COAs that involve a recall period (e.g., past 24 hours, past 7 days), sponsors should provide support for the appropriateness of the recall period to be used. FDA recommends a clearly specified recall period to help standardize reporting. The recall period should be shown to be suitable for the intended context of use. Sources of evidence to support a given recall period might include empirical study of the accuracy of different recall periods for the measure and/or literature reviews of recall accuracy for the same or related concept of interests. Note that cognitive interviews can provide justification that a given recall period is inappropriate (e.g., by documenting that respondents generated their response thinking about a shorter period of time than specified by the instrument). But cognitive interviews cannot provide evidence that respondents can recollect with sufficient accuracy. The selected recall period should be short enough to minimize the measurement error and/or potential bias (i.e., systematic inflation or deflation of scores) due to recall error, while also minimizing respondent burden.

3. Respondent Fatigue or Burden Does Not Overly Influence Assessment of the Concept of Interest.

Consider whether COAs may induce respondent fatigue and burden due to measure length, complexity, and/or frequency of assessment. For data collected from patients during clinic visits, the order in which COA data are collected (e.g., before or after blood draws and other data collection) can influence respondent fatigue or burden. Respondents who feel fatigued or overburdened during an assessment might not provide data reflective of the underlying disease or the impact of treatment. Evidence from cognitive interviews and/or usability testing may provide insight as to whether a COA might lead to fatigue and/or burden. Sponsors may wish to explore approaches to reduce burden, such as having patients complete assessments at home the day before a clinic visit. Patient experience of burden might also be addressed by improving patients' motivation through explaining the reasons for and importance of any lengthy, complex, and/or frequent assessments.

4. The Mode of Assessment Does Not Overly Influence Assessment of the Concept of Interest.

There are a variety of modes of administration for COAs, including paper-based forms and electronic data capture using standardized devices (i.e., those used with all participants in a trial), or participants' own mobile devices, computers, or other tools for assessment.

Using a mode of collection different from what was originally used for that COA (e.g., originally used paper, now proposed to use a mobile device) may raise concerns about comparability of assessment to prior experience. Similarly, using different collection modes in the same trial (e.g., different modes for different sites) would raise concerns regarding comparability of assessments in the study. In both cases, part of the COA's rationale for using different modes is that whatever measurement error or bias is created by changing mode of assessment will be too small

Contains Nonbinding Recommendations

Draft — Not for Implementation

770 to affect the assessment of the concept of interest.²⁴ Whether this is reasonable will depend upon
771 the situation and how the adaptation between modes was accomplished.

772
773 Sponsors can increase confidence that collection mode does not meaningfully influence
774 measurements by following best practices for adapting measures to different assessment
775 platforms (Critical Path Institute ePRO Consortium 2014a and 2014b; Byrom et al. 2019;
776 Eremenco et al. 2014). FDA also recommends that sponsors conduct usability testing of the
777 different data collection devices with a small number of respondents.

778
779 If a data collection platform has already demonstrated usability in a group of participants thought
780 to be sufficiently similar to the target population and the content of the measure has already been
781 evaluated using cognitive interviews, it may not be necessary to conduct a new equivalence
782 study, especially if the COA uses typical response scales that have been well studied (Byrom et
783 al. 2019).

784
785 Whether more extensive evidence is needed to support the comparability of scores between
786 assessment modes will depend upon the specifics of each case.

787
788 *5. Expectation Bias Does Not Unduly Influence Assessment of the Concept of Interest.*

789
790 Responses to a COA may be influenced by the respondent's (i.e., patient's, caregiver's, or
791 clinician's) or administrator's (for PerfO measures) expectations of how well the patient should
792 be doing. Such expectations could be based on the patient's assignment to an experimental group
793 in an unmasked trial and/or the duration the patient has been in the clinical trial (e.g., earlier
794 versus later study visits). For ClinRO, ObsRO, and some PerfO measures, expectations might
795 also be based on characteristics of the patient, such as their age or sex. An expectation bias could
796 arise in at least two ways:

- 797
- 798 • For items that use a recall period, respondents may selectively recall those instances
799 when symptoms or functioning were consistent with what the respondent expects. For
800 example, a patient receiving a new medical product that the patient believes is effective
801 provides self-reported assessments of functioning with a 7-day recall at both baseline and
802 follow-up. The patient's expectations of benefit might make it more likely that the patient
803 reports at follow-up based on recollections of more positive instances of functioning
804 rather than negative ones.
 - 805 • Expectations might influence how a respondent or an administrator interprets the
806 meaning of items (Rapkin and Schwartz 2004). For example, consider two patients
807 suffering from rheumatoid arthritis—one 49 years old and the other 82 years old. Relative
808 to the 49-year-old, the older patient might expect that pain and discomfort are normal
809 parts of aging. When asked about the impact of pain on daily functioning using response
810 options of *None, Mild, Moderate, or Severe*, the older patient might interpret the response
811 options differently from the younger patient. Thus, though both patients might have the
812 same degree of pain and functional limitations, the older patient might select *Moderate*
813 while the younger patient selects *Severe*.
- 814

²⁴ In some cases, a new mode of assessment may increase the accuracy or precision of scores.

Contains Nonbinding Recommendations

Draft — Not for Implementation

815 Minimizing the influence of biases, including expectation bias, is very important and can be done
816 by conducting randomized, placebo-controlled, and double-masked trials. Concealing the
817 patients' assignment to study arms will also minimize the influence of patient expectations about
818 whether a treatment will be beneficial.

819
820 *6. Practice Effects Do Not Overly Influence the Assessment of the Concept of Interest.*
821 *[PerfO Measures]*
822

823 For PerfO measures, it is possible that patients' performance on the tasks could improve over
824 time due to practice rather than to real improvements in the concept of interest. Practice effects
825 may be minimized by using different tasks for each assessment whenever possible. Practice
826 effects might also be reduced by administering a PerfO measure less frequently and/or separated
827 by longer periods of time. Patients could also train on the tasks prior to randomization so that the
828 patients' baseline status already reflects the effects of practice. Although randomization can
829 reduce the impact of practice effects, it is still possible within a randomized trial for practice
830 effects to (1) limit the ability of a COA to demonstrate the full magnitude of a treatment effect,
831 and/or (2) differ by treatment arm when the intervention causes changes in cognitive function
832 that facilitates practice effects. Evidence for or against the presence of strong practice effects
833 could be obtained by examining the performance on PerfO measures over time among patients in
834 a natural history or non-affected cohort outside of a trial, or by examining changes over time
835 within a placebo group of a trial.

836
837 **E. The Method of Scoring Responses to the COA Is Appropriate for Assessing the**
838 **Concept of Interest.**
839

840 Every COA provides some way for responses to be recorded or coded as an observed score for a
841 prompt. For example, a PRO measure that assesses current nausea intensity might allow patients
842 to record their responses on a verbal rating scale with four adjectives, producing an observed
843 score between 0 and 3. A walking test might record the distance (or time) a patient walks for a
844 specified time (or distance), producing an observed score in meters (or seconds).

845
846 *1. Responses to an Individual Item or Task*
847

848 For an individual item or task, response options should be non-overlapping and differences
849 among adjacent response categories should reflect true differences in the concept of interest. The
850 wording of the response options should be clear and concrete, and the instructions for making or
851 recording the responses should be clearly understandable. Support for these considerations can
852 come from cognitive interview data, demonstrating that respondents have no difficulty selecting
853 an answer that matches their experience.

854
855 FDA generally does not recommend the use of a visual analog scale (VAS). There are known
856 limitations with its administration (e.g., cannot be administered verbally or over the phone;
857 photocopying or electronic rendering on different monitors or devices lead to different lengths of
858 lines displayed at during a single trial) and interpretability (e.g., higher rates of missing data or
859 incomplete data) (Dworkin et al. 2005 and Hawker et al. 2011).

860

Contains Nonbinding Recommendations

Draft — Not for Implementation

2. Rationale for Combining Responses to Multiple Items or Tasks

If multiple items or tasks are combined to generate a score on a COA, then the rationale for the method of scoring should be described and supported with evidence (Edwards et al. 2017). The approach for combining responses to multiple items/tasks is often expressed as a measurement model that relates responses to particular items/tasks to the score(s) thought to reflect the concept of interest. The rationale and justification for combining items or tasks will depend upon the particular measurement model chosen for the measure. Although there are many possible measurement models that might be appropriate for a COA, two of the more common models are the reflective indicator and composite indicator models.

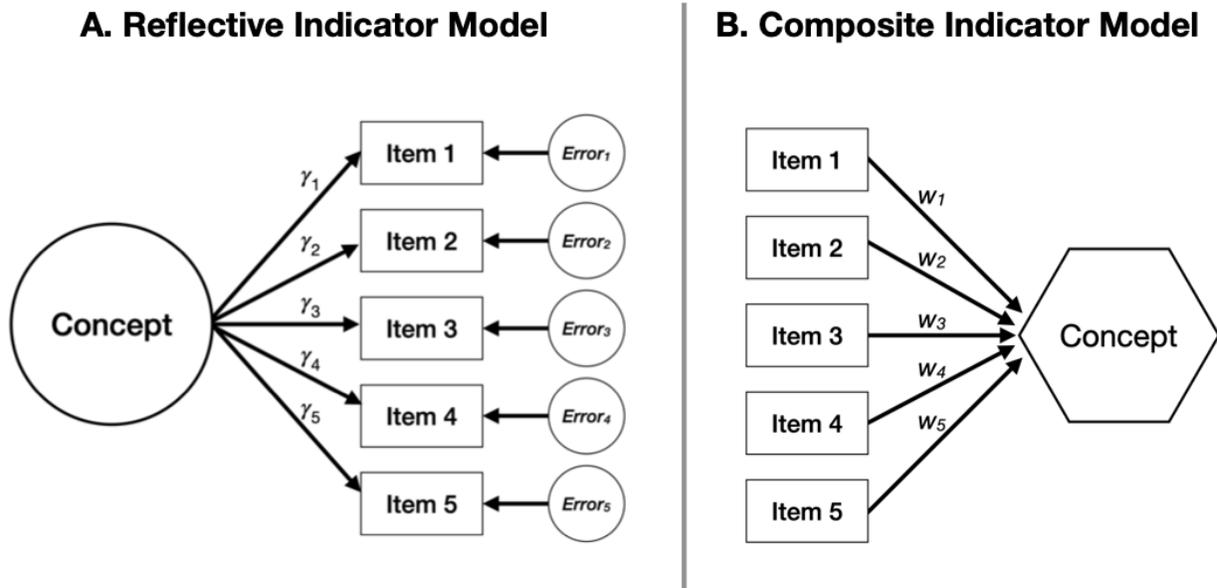
3. Scoring a Unidimensional COA: The Reflective Indicator Model

The justification for combining responses across multiple items for a reflective indicator model is that all the item responses reflect, or are caused by, a single aspect of the patient described by the concept of interest (Fayers and Machin 2016)—an assumption known as *unidimensionality*. For example, a PRO measure might consist of multiple items that ask about lower limb-related mobility. Because the items are all reflections of, or effects of, lower limb-related mobility, the item responses should be consistent with a unidimensional measurement model (see Figure 4A). Statistical evidence including, but not limited to, confirmatory factor analysis can be provided to support the reasonableness of the assumption of unidimensionality. Sponsors or measure developers should also be clear about the psychometric model that is assumed (e.g., Classical Test Theory, Partial Credit Model, Samejima’s Graded Response Model, Rasch Model) and supply statistical evidence in support of model assumptions and fit, as well as relevant model parameters. Note that FDA does not endorse any particular psychometric modeling approach but will review the strength of evidence in support of a model’s use in specific cases.

Contains Nonbinding Recommendations

Draft — Not for Implementation

888 **Figure 4. Representations of Reflective (Panel A) and Composite (Panel B) Indicator**
889 **Models.**



890 Note: In panel A, the concept within a circle is conceptualized as a latent variable; the smaller circles represent
891 measurement error that contributes to the responses of each item; γ denotes the causal effect of the concept on the
892 item response. In panel B, the concept within a hexagon is conceptualized as a composite variable; w indicates the
893 weight (may or may not be equally weighted) used for the item response in computing the calculated composite
894 score that represents the concept.
895

896
897 Some PRO measures based on a reflective indicator model consist of multiple items assessing
898 multiple domains. For such measures, if the multiple domains will be used to assess the
899 concept(s) of interest, a rationale should be given supporting the conceptual distinctiveness of the
900 different domains and psychometric analyses should be provided in support of the assumed
901 dimensionality of the measure (e.g., demonstrating adequate fit of a confirmatory factor analysis
902 model that includes the multiple domains).
903

904 When the assumptions are met, the sample size is large enough, and the model fit is acceptable,
905 item response theory (IRT) models provide an approach to design, evaluation, and the scoring of
906 COAs based on a reflective measurement model. Failure to assess assumptions such as
907 unidimensionality, local independence of items, and measurement invariance may result in
908 inadequate evidence of the properties of an IRT-based COA. When using IRT models to design,
909 evaluate, or score a COA, additional information concerning the items and scale can be provided.
910 In addition to estimated item parameters and corresponding standard errors, the functioning of
911 response categories and DIF can be evaluated. For example, item characteristic curves can be
912 used to examine for signs of redundant response categories for measures developed using IRT
913 for polytomous items. Multiple IRT models and approaches exist. The chosen model or approach
914 should fit with the characteristics of the COA and its items or tasks.
915

Contains Nonbinding Recommendations

Draft — Not for Implementation

4. Scoring for a COA That Summarizes Across Heterogeneous Health Experiences: The Composite Indicator Model

Some measures assess a concept of interest using multiple items that, taken together, define the concept of interest. For example, the concept Basic Activities of Daily Living might be defined by a sponsor or measure developer as the degree to which the patient is able to accomplish everyday tasks that are necessary to live independently. The item content that defines *everyday tasks* might be determined through a consensus process with patients and their caregivers and could result, for example, in items addressing personal hygiene, dressing oneself, toileting, eating, and ambulation. Note that although it is likely that some of the item responses will be associated with one another, it is not necessary, because it is not assumed that all the items are reflective of or caused by a single, underlying thing as was the case for the reflective indicator model. Rather, these items, known as composite indicators, are like the ingredients of what is labeled Basic Activities of Daily Living (see Figure 4B) (Bollen and Bauldry 2011).²⁵

For COAs based on a composite indicator model, sponsors or measure developers should describe and justify the process for selecting the items that make up the measure (e.g., by citing a consensus process with patients and others). A rationale should also be given for the way in which responses to the multiple items are combined to arrive at a score for the COA. For example, one might justify taking the sum of the item responses (which implies they are all weighted equally) based on qualitative or quantitative evidence that patients felt that all the activities described by the items are equally important for daily, independent living.

5. Scoring Approaches Based on Computerized Adaptive Testing

Some COAs make use of computerized adaptive testing (CAT) procedures, whereby the next item administered to a respondent depends upon a running estimate of the respondent's status based on the respondent's responses to prior items. The set of potential items to be administered is known as an item bank. With CAT, it is possible that fewer items will be needed to generate a sufficiently precise score for each patient, making the assessment more efficient and less burdensome to patients. Depending upon the concept of interest being assessed, a CAT may or may not be more efficient than administering the same items to every person.

FDA will consider well-justified approaches. To ensure changes in a patient's scores over time are not due to differences in the items administered, it is critical for sponsors to demonstrate, as for any COA, that (1) the item content aligns with the concept of interest; (2) all of the items are well understood by patients in the target population; (3) the items are well-calibrated in the context of a well-fitting IRT model; and (4) in the context of multinational or multicultural trial populations, all of the items in the item bank have undergone an acceptable process of translation and/or adaptation. FDA recommends not making changes to an item bank mid-trial; however, if

²⁵ It is important to distinguish between a composite indicator measurement model and a composite endpoint. Composite indicators are separate items or tasks, which may or may not be correlated, that are combined to create a new summary variable corresponding to a concept of interest. In contrast, a composite endpoint is a way of constructing an endpoint based on two or more individual clinical outcomes (components). The *composite endpoint* is then defined as the occurrence or realization in a patient of any of the specified components. For more discussion of composite endpoints, see the draft guidance for industry *Multiple Endpoints in Clinical Trials*. When final, this guidance will represent FDA's current thinking on this topic.

Contains Nonbinding Recommendations

Draft — Not for Implementation

956 the item bank undergoes any changes (e.g., maintenance, updating, or addition of new items)
957 during the clinical trial, it is important to demonstrate that the item bank remains well calibrated
958 with respect to the original concept being measured. Sponsors should also describe and justify
959 the stopping rule used for the CAT in terms of the minimum level of measurement precision
960 desired. It is also suggested that stopping rules include considerations of patient burden (e.g., by
961 stopping the CAT after some maximum number of items have been administered). Note that
962 sponsors might consider different CAT stopping rules for different contexts of use.

963
964 As an alternative to full CAT administration, sponsors might also consider a hybrid CAT in
965 which every patient is administered items by the CAT algorithm and a fixed set of items (if not
966 already selected by the algorithm).

967 968 *6. Approach to Missing Item or Task Responses*

969
970 Missing item responses can create problems for interpreting and using scores from a COA with
971 multiple items or tasks. The scoring algorithm should explicitly state the conditions under which
972 a score can still be computed in the presence of missing item/task responses, e.g., specifying the
973 minimum number of items/tasks responses to compute a score and/or how missing items are to
974 be scored. Any rules for handling missing item or task responses should be justified sufficiently
975 (e.g., through a missing data simulation study). A copy of the scoring manual should be provided
976 to FDA so that reviewers can verify and replicate the sponsor's proposals according to the
977 published scoring rules.

978 979 **F. Scores From the COA Correspond to the Specific Health Experience(s) the** 980 **Patient Has Related to the Concept of Interest.**

981
982 Scores produced by a COA should correspond to important aspects of health in the patient's life
983 that the medical product is targeting (Walton et al. 2015). Some measures assess a concept of
984 interest that corresponds directly to the specific health experiences of the patient, such as ADLs
985 (see Figure 1) or patient-reported pain intensity. For such measures, there might be little
986 uncertainty that the scores correspond to the patient's experience. However, other measures
987 might assess a concept of interest that is indirectly related to the specific health experiences that
988 the medical product is targeting.

989
990 For example, an aspect of health that might be important to patients in the target population is
991 lower limb-related function (Walton et al. 2015), which might include specific health
992 experiences like walking from room to room inside a house and hiking outside on an uneven
993 trail. A PRO measure might be used to assess this concept in a relatively direct way by asking
994 the patient about the ease with which they have done a range of activities that require lower
995 limb-related function (corresponding to Activities 1 to 9 in Figure 5). Although measurement
996 error might influence scores on the PRO measure, it is generally thought that those scores are
997 directly related to the lower limb-related activities in the patients' usual lives. However, if there
998 was significant heterogeneity among patients' physical environments and/or wide heterogeneity
999 in the lower limb-related activities that patients undertake, a sponsor might decide instead to
1000 assess patients in a standardized environment via a PerfO measure. Under standardized
1001 conditions, one is no longer directly assessing lower limb-related function outside the test

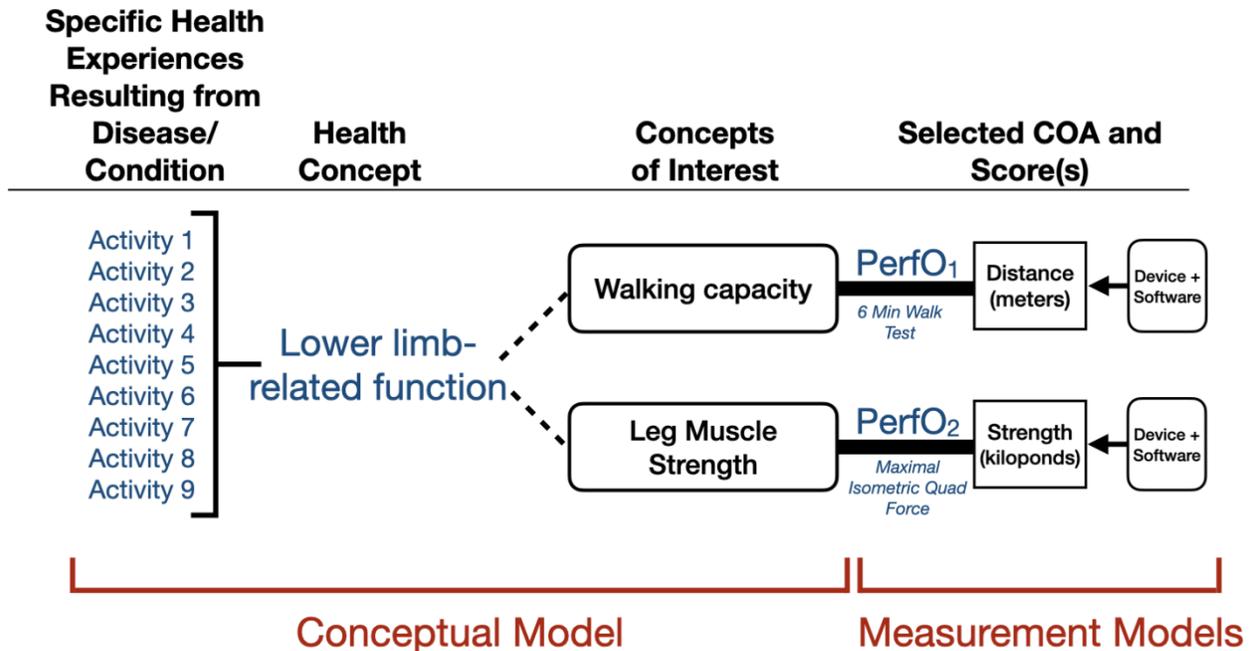
Contains Nonbinding Recommendations

Draft — Not for Implementation

1002 environment. Instead, the concepts of interest being assessed are important subcomponents of
 1003 lower limb-related function that are amenable to standardized assessment, but neither are
 1004 sufficient alone to support an inference about the patient’s overall lower limb-related function.
 1005 Because of this, these measured concepts of interest could be considered more indirect
 1006 reflections of patients’ lower limb-related functioning in their daily lives. In this example, the
 1007 sponsor might decide to measure in the test environment walking capacity and leg muscle
 1008 strength, which are indirectly related to patients’ lower limb-related functioning in their daily
 1009 lives (reflected by the dotted line in the conceptual framework shown in Figure 5). But in the
 1010 rationale for the use of each measure, it would still be important to evaluate how well scores are
 1011 related to the patients’ lower limb-related mobility activities in their usual lives outside of the
 1012 clinical trial context.

1013
1014

1015 **Figure 5. Example Conceptual Framework for Measures of Two Concepts of Interest With**
 1016 **Indirect Relationships to the Patients' Specific Health Experiences** (Note: Dotted lines
 1017 indicate an indirect relationship between the health concept and concept of interest.)



1018
1019
1020

1021 For measures such as these in which the relationship between the scores and the important aspect
 1022 of health is less direct, more uncertainty exists. Thus, sponsors and measure developers might
 1023 seek additional evidence by investigating the relationship between scores on the COA and other
 1024 variables that are expected to be more directly related to the patient’s experience. This is known
 1025 as convergent evidence.²⁶ The other variables could include alternative measures of, or be related
 1026 to, the measured concept of interest using different methods and/or sources (e.g., observer report

²⁶ *Convergent evidence* was referred to as *convergent validity* in the 2009 PRO Guidance.

Contains Nonbinding Recommendations

Draft — Not for Implementation

1027 or performance tests). For the example shown in Figure 5, the sponsor might assess patient-
1028 reported lower limb-related functioning in daily life along with measures of walking capacity
1029 and leg muscle strength in a phase 2 trial. The sponsor might predict a moderate correlation²⁷
1030 between the PRO measure's scores and scores on the two performance measures and could test
1031 this using the phase 2 trial data.

1032
1033 When a sponsor is collecting convergent evidence, FDA notes that the correlation coefficient
1034 cutoffs based on Cohen (1988) may be too low to be considered as a moderate and/or strong
1035 correlation. FDA reminds sponsors that when prespecifying correlation coefficient cutoffs for the
1036 psychometric statistical analysis plan (SAP), it is important to take into consideration the *a priori*
1037 hypothesized relationship(s) among the concept(s) measured by any proposed reference
1038 measure(s) in the convergent evidence study and the concept(s) measured by the proposed COA
1039 measure. When interpreting correlation coefficients, sponsors should consider the size of the
1040 corresponding coefficient of determination and how the distribution of the variables might
1041 impact the magnitude of the correlation (e.g., attenuation due to restriction of range).

1042
1043 Sponsors and measure developers might also conduct empirical comparisons of scores for patient
1044 groups known to differ with respect to the concept of interest (i.e., known groups validity
1045 evidence²⁸). When a sponsor is collecting known-groups evidence, FDA does not recommend
1046 dividing COA scores into groups based on the distribution(s) of reference measure scores (e.g.,
1047 tertiles, quartiles, medians, or quintiles), because the percentile cutoff values are arbitrary and
1048 may vary across samples. Additionally, patient groups created based on the distribution of
1049 reference measure scores may not represent clinically distinct groups. Sponsors should propose
1050 and justify appropriate cutoff values that connote distinct levels of symptom severity and/or
1051 impact severity. In addition, sponsors should provide details of the proposed model and the
1052 hypothesis tests that will be performed.

1053
1054 **G. Scores From the COA Are Sufficiently Sensitive to Reflect Clinically Meaningful**
1055 **Changes Within Patients Over Time in the Concept of Interest Within the**
1056 **Context of Use.**

1057
1058 Though scores on the measure might correspond to the real experiences of patients (see section
1059 IV.F), the assessments might not have enough sensitivity to detect consequential²⁹ changes
1060 within patients over the duration of a clinical trial. Thus, it is important to show evidence that the
1061 scores are sensitive enough to detect such changes. Note that this assumption refers specifically
1062 to the ability to detect change, which reflects the signal-to-noise ratio of the COA's scores.
1063 Sensitivity to change could vary depending upon the target population, as when floor or ceiling
1064 effects limit the ability to observe change.

1065

²⁷ It would be reasonable in this example for a sponsor to expect a moderate, but not large, correlation in this case. In the example, the sponsor chose PerfO measures rather than PRO measures out of concern for the heterogeneity in the patients' environments. That environmental heterogeneity is expected to reduce the magnitude of the relationship between patient-reported and performance tested assessments of lower limb mobility.

²⁸ The extent to which scores differed between groups known to differ on the concept of interest was referred to as *known groups validity* in the 2009 PRO Guidance.

²⁹ The Agency expects to address the concept of clinically meaningful changes in Guidance 4.

Contains Nonbinding Recommendations

Draft — Not for Implementation

1066 There are two general approaches to providing evidence on this point, with one providing more
1067 direct evidence than the other.

1068

1069 1. Evaluating Responsiveness to Change

1070

1071 One strategy for collecting relatively direct evidence for sensitivity to within-person change (i.e.,
1072 responsiveness) is to examine the relationship between changes in the COA's scores and changes
1073 in some other measure(s) of the same or proximal construct, assessed over the same or
1074 comparable time frames, that would be expected to change for the same reason the COA scores
1075 should change (e.g., natural disease progression or response to an intervention). When changes in
1076 the COA scores track closely with changes in the other measure, there is increased confidence
1077 that the COA scores can reflect changes in the concept of interest. For example, a sponsor might
1078 examine how closely changes in a COA intended to measure the weekly headache pain severity
1079 with changes in the number of days with migraine. It is important to specify hypotheses about
1080 the expected direction and magnitude of the correlation(s) between changes in the COA scores
1081 and changes in the other measure(s) (Mokkink et al. 2011).

1082

1083 2. Evaluating Reliability/Precision

1084

1085 Before direct evidence of responsiveness to change is available, sponsors can evaluate a
1086 prerequisite for responsiveness to change—that there is minimal measurement error in COA
1087 scores.

1088

1089 When evaluating reliability, different types of consistency are relevant to various COAs in their
1090 context of use (see Table 2).

1091

1092 **Table 2. Possible Assumptions About Consistency of Scores**

Scores are reasonably consistent ...	Type of Evidence	Potential Relevance for COA Type			
		<i>PRO</i>	<i>ObsRO</i>	<i>ClinRO</i>	<i>PerfO</i>
... over time within clinically stable patients	Test-retest reliability	X	X	X	X
... across different raters	Inter-rater reliability			X	X ^a
... within the same rater for the same patients (when the patients have not clinically changed)	Intra-rater reliability			X	X ^a
... across different but highly related or similar tasks	Evaluation of score differences between related tasks or sets of tasks				X

1093 ^aApplies only if the PerfO measure requires a trained rater as part of the assessment process.

Contains Nonbinding Recommendations

Draft — Not for Implementation

1094
1095 Test-retest reliability should be evaluated in the absence of any systematic intervening effects
1096 other than time. Sponsors should specify one or more criteria to define stable patients. FDA
1097 recommends that, in most cases, intraclass correlation coefficients be calculated using absolute-
1098 agreement, two-way mixed-effects model with the time as a fixed effect (McGraw and Wong
1099 1996; Shrout and Fleiss 1979), as suggested by Shrout and Fleiss (1979) and Qin et al. (2019).
1100 Note that test-retest reliability evidence is only relevant for diseases or conditions in which a
1101 patient’s health status can remain stable for some period of time (e.g., 1 to 2 weeks). In a disease
1102 in which symptoms can vary substantially during a single day, the assumption of consistency of
1103 scores over time may be irrelevant, and so it would not be useful or even possible to collect
1104 evidence of test-retest reliability.

1105
1106 For measures developed using IRT modeling, an alternative estimate of reliability can be
1107 generated based on the information function. The associated standard errors can provide another
1108 method of examining the variability and consistency of scores.

1109
1110 During the development process of a COA, evidence of good reliability might be obtained earlier
1111 in the process (e.g., using a cross-sectional study design). This evidence, along with other
1112 supporting material, might be enough to justify the exploratory use of the COA in prospective
1113 trials (e.g., phase 2).

1114
1115 **H. Differences in the COA Scores Can Be Interpreted and Communicated Clearly**
1116 **in Terms of the Expected Impact on Patients’ Experiences.**

1117
1118 Because findings from clinical trials are used to inform decisions that patients, providers, and/or
1119 payers make, it should be clear what the COA scores reflect and how the magnitude of the
1120 difference(s) relates to patients’ lives. This final component of the rationale includes any
1121 assumptions that might be involved in translating COA score differences into within-patient
1122 changes and why these within-patient changes are considered meaningful in patients’
1123 experiences. The Agency expects to address the concept of clinically meaningful changes and
1124 related justifications in Guidance 4.

1125
1126 For all the potential assumptions of a rationale, the specific versions will depend upon the type of
1127 COA.

Contains Nonbinding Recommendations

Draft — Not for Implementation

1128 **V. ABBREVIATIONS**

1129

1130	ADLs	Activities of Daily Living
1131	CAT	Computerized Adaptive Testing
1132	ClinRO	Clinician-Reported Outcome
1133	COA	Clinical Outcome Assessment
1134	DHTs	Digital Health Technologies
1135	DIF	Differential Item Functioning
1136	IRT	Item Response Theory
1137	ObsRO	Observer-Reported Outcome
1138	PerfO	Performance-Based Outcome
1139	PFDD	Patient-Focused Drug Development
1140	PRO	Patient-Report Outcome
1141	SAP	Statistical Analysis Plan
1142	VAS	Visual Analog Scale

Contains Nonbinding Recommendations

Draft — Not for Implementation

1143 **VI. USEFUL REFERENCES FOR SELECTING, MODIFYING, AND DEVELOPING** 1144 **CLINICAL OUTCOME ASSESSMENTS** 1145

1146 *Please note that the citation of a scientific reference in this guidance does not constitute FDA’s*
1147 *endorsement of approaches or methods presented in that reference for any particular study.*
1148 *Study designs are evaluated on a case by case basis under applicable legal standards.*

1149
1150 American Educational Research Association, American Psychological Association, National
1151 Council on Measurement in Education, 2014, Standards for educational and
1152 psychological testing, Washington, DC: American Educational Research Association.
1153

1154 Arbuckle, R, and L Abetz-Webb, 2013, “Not Just Little Adults”: Qualitative Methods to Support
1155 the Development of Pediatric Patient-Reported Outcomes, *Patient*, 6(3):143-159.
1156

1157 Benjamin, K, MK Vernon, DL Patrick, E Perfetto, S Nestler-Parr, and L Burke, 2017, ISPOR
1158 Rare Disease Clinical Outcomes Assessment Task Force. Patient-Reported Outcome and
1159 observer-Reported Outcome Assessment in Rare disease Clinical Trials: An ISPOR COA
1160 Emerging Good Practices Task Force Report, *Value Health*, 20:(8)38-55.
1161

1162 BEST (Biomarkers, Endpoints and Other Tools) Resource, 2016, available at
1163 <https://www.ncbi.nlm.nih.gov/books/NBK338448/>
1164

1165 Bevans, KB, AW Riley, J Moon, and CB Forrest, 2010, Conceptual and Methodological
1166 Advances in Child-Reported Outcomes Measurement, *Expert Review of*
1167 *Pharmacoeconomics & Outcomes Research*, 10(4):385-396.
1168

1169 Bollen, KA, and S Bauldry, 2011, Three Cs in Measurement Models: Causal Indicators,
1170 Composite Indicators, and Covariates, *Psychological Methods*, 16(3):265–284.
1171

1172 Bushnell, DM, KP McCarrier, EN Bush, L Abraham, C Jamieson, F McDougall, MH Trivedi,
1173 ME Thase, L Carpenter, and SJ Coons, 2019, Symptoms of Major Depressive Disorder
1174 Scale: Performance of a Novel Patient-Reported Symptom Measure, *Value in Health*,
1175 22(8):906-915.
1176

1177 Byrom, B, C Gwaltney, A Slagle, A Gnanasakthy, and W Muehlhausen, 2019, Measurement
1178 Equivalence of Patient-Reported Outcome Measures Migrated to Electronic Formats: A
1179 Review of Evidence and Recommendations for Clinical Trials and Bring Your Own
1180 Device, *Therapeutic Innovation & Regulatory Science*, 53(4):426-430.
1181

1182 Chung, DC, S McCague, ZF Yu, S Thill, J DiStefano-Pappas, J Bennett, D Cross, K Marshall, J
1183 Wellman, and KA High, 2018, Novel Mobility Test to Assess Functional Vision in
1184 Patients With Inherited Retinal Dystrophies, *Clinical & Experimental Ophthalmology*,
1185 46(3):247-259.
1186

1187 Cohen, J. *Statistical Power Analysis for the Behavioral Sciences* (2nd Ed). 1988. Hillsdale, NJ:
1188 Lawrence Erlbaum Associates.

Contains Nonbinding Recommendations

Draft — Not for Implementation

- 1189
1190 Cook, DA, R Brydges, S Ginsburg, and R Hatala, 2015, A Contemporary Approach to Validity
1191 Arguments: A Practical Guide to Kane's Framework, *Medical Education*, 49(6):560–575.
1192
- 1193 Coons, SJ, CJ Gwaltney, RD Hays, JJ Lundy, JA Sloan, DA Revicki, WR Lenderking, D Cella,
1194 and E Basch, 2009, ISPOR ePRO Task Force. Recommendations on Evidence Needed
1195 To Support Measurement Equivalence Between Electronic and Paper-Based Patient-
1196 Reported Outcome (PRO) Measures: ISPOR ePRO Good Research Practices Task Force
1197 report, *Value Health*, 12(4):419-29.
1198
- 1199 Critical Path Institute ePRO Consortium, 2014a, Best Practices for Electronic Implementation of
1200 Patient-Reported Outcome Response Scale Options, accessed August 19, 2019, available
1201 at [https://c-path.org/wp-](https://c-path.org/wp-content/uploads/2014/05/BestPracticesForElectronicImplementationOfPROResponseScaleOptions.pdf)
1202 [content/uploads/2014/05/BestPracticesForElectronicImplementationOfPROResponseScaleOptions.pdf](https://c-path.org/wp-content/uploads/2014/05/BestPracticesForElectronicImplementationOfPROResponseScaleOptions.pdf).
1203
1204
- 1205 Critical Path Institute ePRO Consortium, 2014b, Best Practices for Migrating Existing Patient-
1206 Reported Outcome Instruments to a New Data Collection Mode, accessed August 19,
1207 2019, available at [https://c-path.org/wp-](https://c-path.org/wp-content/uploads/2014/05/BestPracticesForMigratingExistingPROInstrumentstoNewDataCollectionMode.pdf)
1208 [content/uploads/2014/05/BestPracticesForMigratingExistingPROInstrumentstoNewData](https://c-path.org/wp-content/uploads/2014/05/BestPracticesForMigratingExistingPROInstrumentstoNewDataCollectionMode.pdf)
1209 [CollectionMode.pdf](https://c-path.org/wp-content/uploads/2014/05/BestPracticesForMigratingExistingPROInstrumentstoNewDataCollectionMode.pdf).
1210
- 1211 Dayan, SH, J Schlessinger, K Beer, LM Donofrio, DH Jones, S Humphrey, J Carruthers, PF
1212 Lizzul, TM Gross, FC Beddingfield, and C Somogyi, 2018, Efficacy and Safety of ATX-
1213 101 by Treatment Session: Pooled Analysis of Data From the Phase 3 REFINE Trials,
1214 *Aesthetic Surgery Journal*, 38(9):998-1010.
1215
- 1216 de Vet, HCW, CB Terwee, LB Mokkink, and DL Knol, 2011, *Measurement in Medicine*.
1217 Cambridge University Press: Cambridge.
1218
- 1219 Dworkin, RH, DC Turk, JT Farrar, JA Haythornthwaite, MP Jensen, NP Katz, RD Kerns, G
1220 Stucki, RR Allen, N Bellamy, DB Carr, J Chandler, P Cowan, R Dione, BS Galer, S
1221 Hertz, AR Jadad, LD Kramer, DC Manning, S Martin, CG McCormick, MP McDermott,
1222 P McGrath, S Quessy, BA Rappaport, W Robbins, JP Robinson, M Rothman, MA Royal,
1223 L Simon, JW Stauffer, W Stein, J Tollett, J Wernicke, and J Witter, 2005, Core Outcome
1224 Measures for Chronic Pain Clinical Trials: IMMPACT Recommendations, *Pain*, 113:9–
1225 19.
1226
- 1227 Edwards, MC, A Slagle, JD Rubright, and RJ Wirth, 2017, Fit For Purpose And Modern Validity
1228 Theory in Clinical Outcomes Assessment, *Quality of Life Research*, 14(2):1-10.
1229
- 1230 Eremenco, S, SJ Coons, J Paty, K Coyne, A Bennett, and D McEntegart, 2014, PRO Data
1231 Collection in Clinical Trials Using Mixed Modes: Report of the ISPOR PRO Mixed
1232 Modes Good Research Practices Task Force, *Value Health*, 17:501–516.
1233
- 1234 Fayers, PM, and D Machin, 2016, *Quality of Life* (3rd ed.), West Sussex: Wiley Blackwell.

Contains Nonbinding Recommendations

Draft — Not for Implementation

- 1235
1236 FDA COA Qualification Program, 2021, available at [https://www.fda.gov/drugs/drug-](https://www.fda.gov/drugs/drug-development-tool-ddt-qualification-programs/clinical-outcome-assessment-coa-qualification-program)
1237 [development-tool-ddt-qualification-programs/clinical-outcome-assessment-coa-](https://www.fda.gov/drugs/drug-development-tool-ddt-qualification-programs/clinical-outcome-assessment-coa-qualification-program)
1238 [qualification-program](https://www.fda.gov/drugs/drug-development-tool-ddt-qualification-programs/clinical-outcome-assessment-coa-qualification-program).
1239
- 1240 FDA Guidance for Industry and FDA Staff, Applying Human Factors and Usability Engineering
1241 to Medical Devices (February 2016), available at [https://www.fda.gov/regulatory-](https://www.fda.gov/regulatory-information/search-fda-guidance-documents/applying-human-factors-and-usability-engineering-medical-devices)
1242 [information/search-fda-guidance-documents/applying-human-factors-and-usability-](https://www.fda.gov/regulatory-information/search-fda-guidance-documents/applying-human-factors-and-usability-engineering-medical-devices)
1243 [engineering-medical-devices](https://www.fda.gov/regulatory-information/search-fda-guidance-documents/applying-human-factors-and-usability-engineering-medical-devices).
1244
- 1245 FDA Guidance for Industry, Computerized Systems Used in Clinical Investigations (May 2007),
1246 available at
1247 [https://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guid](https://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM070266.pdf)
1248 [ances/UCM070266.pdf](https://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM070266.pdf).
1249
- 1250 FDA Guidance for Industry, Investigators, and Other Stakeholders, Digital Health Technologies
1251 for Remote Data Acquisition in Clinical Investigations (December 2021), available at
1252 [https://www.fda.gov/regulatory-information/search-fda-guidance-documents/digital-](https://www.fda.gov/regulatory-information/search-fda-guidance-documents/digital-health-technologies-remote-data-acquisition-clinical-investigations)
1253 [health-technologies-remote-data-acquisition-clinical-investigations](https://www.fda.gov/regulatory-information/search-fda-guidance-documents/digital-health-technologies-remote-data-acquisition-clinical-investigations)
1254
- 1255 FDA Guidance for Industry, Patient-Reported Outcome Measures: Use in Medical Product
1256 Development to Support Labeling Claims (December 2009), available at
1257 [https://www.fda.gov/regulatory-information/search-fda-guidance-documents/patient-](https://www.fda.gov/regulatory-information/search-fda-guidance-documents/patient-reported-outcome-measures-use-medical-product-development-support-labeling-claims)
1258 [reported-outcome-measures-use-medical-product-development-support-labeling-claims](https://www.fda.gov/regulatory-information/search-fda-guidance-documents/patient-reported-outcome-measures-use-medical-product-development-support-labeling-claims).
1259
- 1260 FDA Guidance for Industry, Electronic Source Data in Clinical Investigations (September 2013),
1261 available at
1262 [https://www.fda.gov/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UC](https://www.fda.gov/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM328691)
1263 [M328691](https://www.fda.gov/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM328691).
1264
- 1265 FDA Report: Complex Issues in Developing Drugs and Biological Products for Rare Diseases
1266 and Accelerating the Development of Therapies for Pediatric Rare Diseases Including
1267 Strategic Plan: Accelerating the Development of Therapies for Pediatric Rare Diseases
1268 (July 2014), available at <https://www.fda.gov/media/89051/download>.
1269
- 1270 FDA Draft Guidance for Industry, Formal Meetings Between the FDA and Sponsors or
1271 Applicants of PDUFA Products (December 2017), available at
1272 [https://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guid](https://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM590547.pdf)
1273 [ances/UCM590547.pdf](https://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM590547.pdf).
1274

Contains Nonbinding Recommendations

Draft — Not for Implementation

- 1275 FDA Draft Guidance for Industry, Multiple Endpoints in Clinical Trials (January 2017),
1276 available at
1277 <https://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM536750.pdf>.
1278
1279
- 1280 FDA Study Data Technical Conformance Guide: Technical Specifications Document
1281 *(incorporated by reference into “Guidance for Industry: Providing Regulatory*
1282 *Submissions in Electronic Format – Standardized Study Data” (June 2021)), available at*
1283 <https://www.fda.gov/downloads/ForIndustry/DataStandards/StudyDataStandards/UCM384744.pdf>.
1284
1285
- 1286 FDA Guidance for Industry, FDA Staff, and Other Stakeholders, Patient-Focused Drug
1287 Development: Methods to Identify What Is Important to Patients (October 2019),
1288 available at <https://www.fda.gov/regulatory-information/search-fda-guidance-documents/patient-focused-drug-development-methods-identify-what-important-patients-guidance-industry-food-and>.
1289
1290
- 1291
- 1292 FDA Draft Guidance for Industry, Rare Diseases: Common Issues in Drug Development
1293 (February 2019), available at <https://www.fda.gov/media/119757/download>.
1294
- 1295 FDA Draft Guidance for Industry, Rare Diseases: Natural History Studies for Drug Development
1296 (March 2019), available at <https://www.fda.gov/regulatory-information/search-fda-guidance-documents/rare-diseases-natural-history-studies-drug-development>.
1297
1298
- 1299 FDA Medical Device Development Tools (MDDT), 2021, available at
1300 <https://www.fda.gov/medical-devices/science-and-research-medical-devices/medical-device-development-tools-mddt>.
1301
1302
- 1303 Hawker, GA, S Mian, T Kendzerska, and M French, 2011, Measures of Adult Pain: Visual
1304 Analog Scale for Pain (VAS Pain), Numeric Rating Scale for Pain (NRS Pain), McGill
1305 Pain Questionnaire (MPQ), Short-Form McGill Pain Questionnaire (SF-MPQ), Chronic
1306 Pain Grade Scale (CPGS), Short Form-36 Bodily Pain Scale (SF-36 BPS), and Measure
1307 of Intermittent and Constant Osteoarthritis Pain (ICOAP), *Arthritis Care & Research*
1308 *(Hoboken)*, 63(11 Suppl):S240–52.
1309
- 1310 Hinkle, DE, W Wiersma, and SG Jurs, 2003, *Applied Statistics for the Behavioral Sciences* (5th
1311 ed.).
1312
- 1313 Kane, MT, 2013, Validating the Interpretations and Uses of Test Scores, *Journal of Educational*
1314 *Measurement*, 50(1):1-73.
1315

Contains Nonbinding Recommendations

Draft — Not for Implementation

- 1316 Kim, J, H Singh, K Ayalew, K Borrer, M Campbell, L Johnson, A Karesh, NA Khin, JR Less, J
1317 Menikoff, L Minasian, SA Mitchell, EJ Papadopoulos, RL Piekarz, KA Prohaska, S
1318 Thompson, R Sridhara, R Pazdur, and PG Kluetz, 2018, Use of PRO Measures to Inform
1319 Tolerability in Oncology Trials: Implications for Clinical Review, IND Safety Reporting,
1320 and Clinical Site Inspections, *Clinical Cancer Research*, 24(8):1780-1784.
1321
- 1322 Leptak, C, JP Menetski, JA Wagner, J Aubrecht, L Brady, M Brumfield, WW Chin, S
1323 Hoffmann, G Kelloff, G Lavezzari, R Ranganathan, JM Sauer, FD Sistare, T Zabka, and
1324 D Wholley, 2017, What Evidence Do We Need for Biomarker Qualification? *Science*
1325 *Translational Medicine*, 9(417).
1326
- 1327 Matza, LS, DL Patrick, AW Riley, JJ Alexander, L Rajmil, AM Pleil, and M Bullinger, 2013,
1328 Pediatric Patient-Reported Outcome Instruments for Research to Support Medical
1329 Product Labeling: Report of the ISPOR PRO Good Research Practices for the
1330 Assessment of Children and Adolescents Task Force, *Value Health*, 16(4):461-479.
1331
- 1332 McCarrier, KP, LS Deal, L Abraham, SI Blum, EN Bush, ML Martin, ME Thase, and SJ Coons,
1333 2016, Patient-Centered Research to Support the Development of the Symptoms of Major
1334 Depressive Disorder Scale: (SMDDS): Initial Qualitative Research, *The Patient*, 9:117-
1335 134.
1336
- 1337 McCleary, KK, 2020, More Than Skin Deep: Understanding the Lived Experience of Eczema
1338 Voice of The Patient Report, accessed October 13, 2021, available at
1339 <https://www.aafa.org/media/2628/more-than-skin-deep-voice-of-the-patient-report.pdf> .
1340
- 1341 McGraw, KO, and SP Wong, 1996, Forming Inferences About Some Intraclass Correlation
1342 Coefficients. *Psychological methods*, 1(1):30.
1343
- 1344 McKown S, Acquadro C, Anfray C, et al.,2020, Good practices for the translation, cultural
1345 adaptation, and linguistic validation of clinician-reported outcome, observer-reported
1346 outcome, and performance outcome measures. *J Patient-reported Outcomes*,4(1):89.
1347
- 1348 Mokkink, LB, CB Terwee, DL Knol, HCW de Vet, 2011, The New COSMIN Guidelines
1349 Regarding Responsiveness, *BMC Medical Research Methodology*, 11:152.
1350
- 1351 Papadopoulos, EJ, DL Patrick, MS Tassinari, AE Mulberg, C Epps, AR Pariser, LB Burke, 2013,
1352 Chapter 42 - Clinical Outcome Assessments for Clinical Trials in Children, in Mulberg et
1353 al., *Pediatric Drug Development: Concepts and Applications*, 2nd edition, pp.545, John
1354 Wiley & Sons, Ltd.
1355

Contains Nonbinding Recommendations

Draft — Not for Implementation

- 1356 Papadopoulos, EJ, EN Bush, S Eremenco, and SJ Coon, 2019, Why Reinvent the Wheel? Use or
1357 Modification of Existing Clinical Outcome Assessment Tools in Medical Product
1358 Development, *Value Health*, 23(2):151-153.
1359
- 1360 Patrick, DL, LB Burke, CJ Gwaltney, NK Leidy, ML Martin, E Molsen, and L Ring, 2011a,
1361 Content Validity—Establishing and Reporting the Evidence in Newly Developed Patient-
1362 Reported Outcomes (PRO) Instruments for Medical Product Evaluation: ISPOR PRO
1363 Good Research Practices Task Force Report: Part 1—Eliciting Concepts for a New PRO
1364 Instrument, *Value Health*, 14:967-977.
1365
- 1366 Patrick, DL, LB Burke, CJ Gwaltney, NK Leidy, ML Martin, E Molsen, and L Ring, 2011b,
1367 Content Validity—Establishing and Reporting the Evidence in Newly Developed Patient-
1368 Reported Outcomes (PRO) Instruments for Medical Product Evaluation: ISPOR PRO
1369 Good Research Practices Task Force Report: Part 2—Assessing Respondent
1370 Understanding, *Value Health*, 14:978-988.
1371
- 1372 Powers, JH III, DL Patrick, and MK Walton, 2017, Clinician-Reported Outcome (ClinRO)
1373 Assessments of Treatment Benefit: Report of the ISPOR Clinical Outcome Assessment
1374 Emerging Good Practices Task Force, *Value Health*, 20(1):2-14.
1375
- 1376 Qin, S, L Nelson, L McLeod, S Eremenco, and SJ Coons, 2019, Assessing Test–Retest
1377 Reliability of Patient-Reported Outcome Measures Using Intraclass Correlation
1378 Coefficients: Recommendations for Selecting and Documenting the Analytical Formula,
1379 *Quality of Life Research*, 28(4):1029-1033.
1380
- 1381 Rapkin, BD and CE Schwartz, 2004, Toward a Theoretical Model of Quality-of-Life Appraisal:
1382 Implications of Findings from Studies of Response Shift, Health and Quality of Life
1383 Outcomes, 2:14.
1384
- 1385 Rothman, M, L Burke, P Erickson, NK Leidy, DL Patrick, and CD Petrie, 2009, Use of
1386 Existing Patient-Reported Outcome (PRO) Instruments and Their Modification: The
1387 ISPOR Good Research Practices for Evaluating and Documenting Content Validity for
1388 the Use of Existing Instruments and Their Modification PRO Task Force Report, *Value*
1389 *Health*, 12(8):1075–1083.
1390
- 1391 Scott, NW, PM Fayers, NK Aaronson, A Bottomley, A de Graeff, M Groenvold, C Gundy, M
1392 Koller, MA Petersen, MAG Sprangers, 2009, A Simulation Study Provided Sample Size
1393 Guidance for Differential Item Functioning (DIF) Studies Using Short Scales, *Journal of*
1394 *Clinical Epidemiology*, 62(3):288-295. doi:10.1016/j.jclinepi.2008.06.003.
1395

Contains Nonbinding Recommendations

Draft — Not for Implementation

- 1396 ShROUT, PE, and JL Fleiss, 1979, Intraclass Correlations: Uses in Assessing Rater Reliability.
1397 Psychological bulletin, 86(2):420.
1398
- 1399 Ständer, S, M Augustin, A Reich, C Blome, T Ebata, NQ Phan, and JC Szepletowski, 2013,
1400 Pruritus Assessment in Clinical Trials: Consensus Recommendations From the
1401 International Forum for the Study of Itch (IFSI) Special Interest Group Scoring Itch in
1402 Clinical Trials, Acta Dermato-Venereologica, 93(5):509–514.
1403
- 1404 Teresi, JA, C Wang, M Kleinman, RN Jones, DJ Weiss, 2021, Differential Item Functioning
1405 Analyses of the Patient-Reported Outcomes Measurement Information System
1406 (PROMIS®) Measures: Methods, Challenges, Advances, and Future Directions.
1407 Psychometrika, published online 2021:1-38. doi:10.1007/s11336-021-09775-0.
1408
- 1409 Tofgard, K, C Gwaltney, J Paty, JP Mattsson, and PN Soni, 2018, Symptoms and Daily Impacts
1410 Associated With Progressive Familial Intrahepatic Cholestasis and Other Pediatric
1411 Cholestatic Liver Diseases: A Qualitative Study With Patients and Caregivers. Poster
1412 presented at the European Society for Paediatric Gastroenterology, Hepatology, and
1413 Nutrition (ESPGHAN) Annual Meeting.
1414
- 1415 Walton, M, J Powers, J Hobart, D Patrick, P Marquis, S Vamvakas, M Isaac, E Molsen, S Cano,
1416 and L Burke, 2015, Clinical Outcome Assessments: Conceptual Foundation—Report of
1417 the ISPOR Clinical Outcomes Assessment – Emerging Good Practices for Outcomes
1418 Research Task Force. Value in Health, 18(6):741-752.
1419
- 1420 Weinfurt, KP, 2021, Constructing Arguments for the Interpretation and Use of Patient-Reported
1421 Outcome Measures in Research: An Application of Modern Validity Theory, Quality of
1422 Life Research, published online February 25.
1423
- 1424 Wild, D, A Grove, M Martin, S Eremenco, S McElroy, A Verjee-Lorenz, and P Erikson, 2005,
1425 ISPOR Task Force for Translation and Cultural Adaptation. Principles of Good Practice
1426 for the Translation and Cultural Adaptation Process for Patient-Reported Outcomes
1427 (PRO) Measures: Report of the ISPOR Task Force for Translation and Cultural
1428 Adaptation, Value Health, 8(2):94-104.
1429
- 1430 Willis, GB, 2005, Cognitive Interviewing: A Tool for Improving Questionnaire Design.
1431 Thousand Oaks, CA: Sage.
1432
- 1433 Willis, GB, 2015, Analysis of The Cognitive Interview in Questionnaire Design. Oxford: Oxford
1434 University Press.
1435

Contains Nonbinding Recommendations

Draft — Not for Implementation

- 1436 Zbrozek, A, J Hebert, G Gogates, R Thorell, C Dell, E Molsen, G Craig, K Grice, S Kern, and S
1437 Hines, 2013, ISPOR ePRO Systems Validation Good Research Practices Task Force.
1438 Validation of Electronic Systems to Collect Patient-Reported Outcome (PRO) Data—
1439 Recommendations for Clinical Trial Teams: Report of the ISPOR ePRO Systems
1440 Validation Good Research Practices Task Force, Value Health, 16:480-89.

Contains Nonbinding Recommendations

Draft — Not for Implementation

1441 **APPENDIX A: PATIENT-REPORTED OUTCOME MEASURES**

1442

1443 **I. INTRODUCTION**

1444

1445 A PRO is a measure based on a report that comes directly from the patient about the status of a
1446 patient's health condition without interpretation of the patient's response by others. A PRO
1447 measure may be the best COA type to assess a concept of interest when the concept of interest is
1448 any of the following and the patient is able to provide reliable self-report:

- 1449 • A feeling or experience known only to the patient, such as pain, itch, shortness of breath
1450 as no one else has direct access to feelings except for the patient
- 1451 • Any type of functioning or activity that is part of the patients' day-to-day life
- 1452 • The patients' satisfaction or dissatisfaction with their treatment and/or functioning
- 1453 • Degree of impact on day-to-day life associated with one or more symptoms

1454

1455 Note that a PRO measure cannot be completed by a proxy reporter, i.e., someone reporting on
1456 behalf of the patient (see Appendix B ObsRO and Appendix C ClinRO for further discussion).

1457

1458 **II. CONCEPTUAL FRAMEWORK EXAMPLE WITH PRO MEASURES**

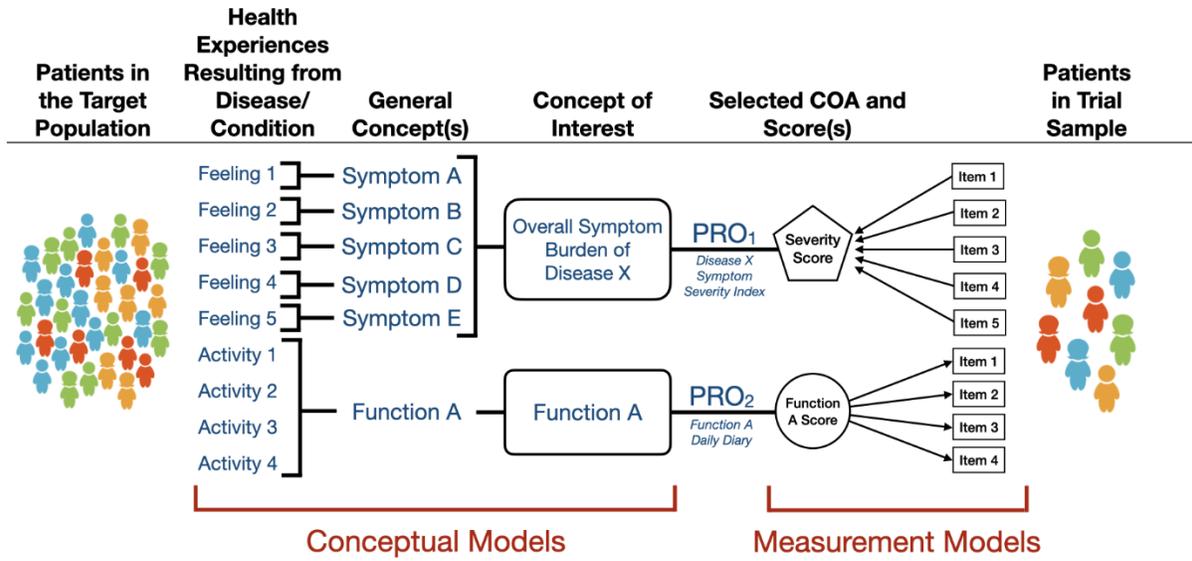
1459

1460 Figure A illustrates a conceptual framework for a study in which two concepts of interest are
1461 assessed using PRO measures. In the example, Disease X can produce multiple symptoms A to
1462 E. One concept of interest is the overall symptom burden of Disease X, and it is assessed using a
1463 multi-item PRO measure, the Disease X Symptom Severity Index. The measurement model
1464 indicates that the five items, corresponding to each of the five symptoms, are combined to create
1465 a summary index (i.e., a composite indicator model). Disease X can also compromise Function
1466 A, which is the other concept of interest. It is measured by a single-item daily diary measure. The
1467 responses to the items by patients in the trial population are used as the basis for an inference
1468 about what patients in the target population might experience if they were given one treatment
1469 versus another.

Contains Nonbinding Recommendations

Draft — Not for Implementation

1470 **Figure A: Illustration of Conceptual Framework for Concepts of Interest Assessed by Two**
 1471 **Patient-Reported Outcomes Measures**



1472

Contains Nonbinding Recommendations

Draft — Not for Implementation

1473 **APPENDIX B: OBSERVER-REPORTED OUTCOME MEASURES**

1474

1475 **I. INTRODUCTION**

1476

1477 An ObsRO measure is a type of COA that assesses observable signs, events, or behaviors related
1478 to a patient's health condition and is reported by someone other than the patient or a health
1479 professional (e.g., parent, caregiver, or someone who cares for the patient the most or spends
1480 significant time with the patient during the relevant observation window in daily life).

1481 An ObsRO measure does not rely on medical judgment or interpretation³⁰ and can be particularly
1482 useful for patients who cannot report for themselves.

1483

Example ObsRO Measures for Use in Clinical Trials

- Rating scales completed by a caregiver, such as:
 - Acute Otitis Media Severity of Symptoms scale, a measure used to assess signs and behaviors related to acute otitis media in infants
- Counts of events recorded by a caregiver (e.g., observer-completed log of seizure episodes)

1484

1485 ***ObsRO versus proxy-reported measures***

1486 A proxy-reported outcome instrument is not an ObsRO instrument; it is an assessment in which
1487 someone other than the patient reports on patient health experiences as if they are the patient or
1488 on the patient's behalf. Proxy-reported outcome instruments are discouraged because they
1489 measure concepts known only to patients and do not necessarily reflect how patients feel and
1490 function in daily life. Concepts that are only known by the patient (e.g., symptoms, feelings)
1491 should be measured by a PRO. FDA acknowledges there are instances when it is impossible to
1492 collect valid and reliable self-report data from the patient. In these instances, it is recommended
1493 an ObsRO instrument be used rather than a proxy-reported outcome instrument.

1494

³⁰ A measure that relies on medical judgment or interpretation is a ClinRO.

Contains Nonbinding Recommendations

Draft — Not for Implementation

Examples of ObsRO Versus Proxy-Reported Item Stem Phrasing

ObsRO items

- “Based on what you observed (saw or what another observer saw), please rate the severity of your child’s abdominal pain-related signs today (such as crying, holding stomach or abdomen).”
- “How frequently did they do household chores (e.g., laundry, washing dishes) in the past week?”
- “Based on what you observed (saw or what was told to you), how often did your child show presence of itch (such as rubbing or scratching) from the time your child woke up today until now?”

Proxy-reported outcome items

- “How severe was your child’s pain from the time your child woke up until right now?”
- “Rate the difficulty they had when shopping for groceries.”
- “Please rate your child’s tiredness over the past 24 hours.”
- “My child felt wheezy and out of breath because of their asthma.”
- “My child felt sad when they had pain.”

1495

1496

ObsRO Selection and Implementation Considerations

1497

1498 Below are key considerations and recommendations for selecting and implementing an ObsRO
1499 measure in a clinical study:

1500

- Conduct qualitative research to explore and define whether a target concept of interest can be reported and observed by someone other than the patient. Such research could include researcher observation of patients along with interviews with caregivers and experts.
- Submit proposed protocols and, as appropriate, interview scripts or observation checklists for FDA review and comment prior to beginning the qualitative research.
- When implementing an ObsRO measure in a clinical study, to the extent feasible, the same observer should complete the assessments throughout the trial to minimize unwanted variability due to different reporters.

1510

II. CONCEPTUAL FRAMEWORK EXAMPLE OF AN ObsRO MEASURE

1511

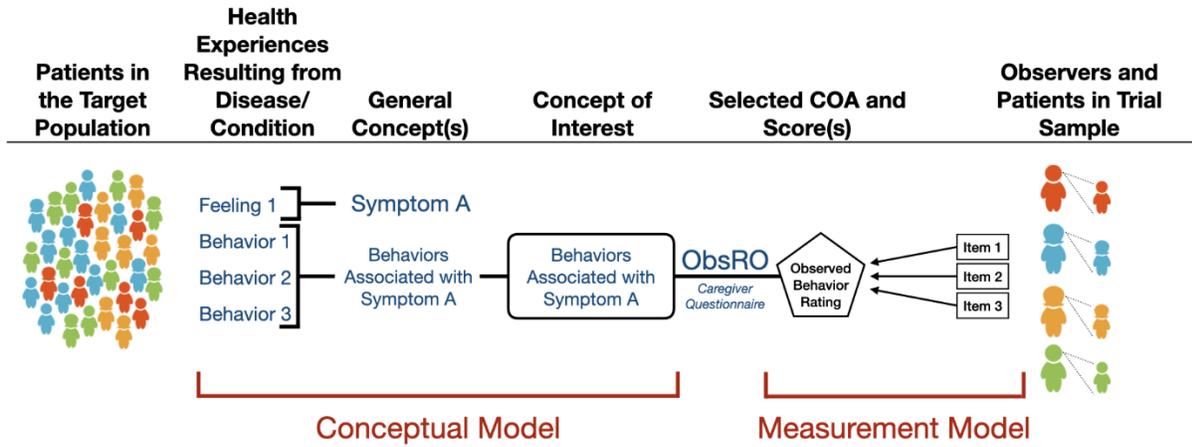
1512
1513 Figure B illustrates a conceptual framework for a multi-item ObsRO measure in the context of
1514 young children with a disease who are unable to reliably and validly self-report. In the example
1515 depicted, Symptom A of the disease causes various behaviors that can be observed by a parent or
1516 caregiver. Parents or caregivers cannot report directly on the symptom severity of their child, but
1517 they can report on these behaviors that are associated with Symptom A, which is the concept of
1518 interest assessed by the ObsRO.

Contains Nonbinding Recommendations

Draft — Not for Implementation

1519

1520 **Figure B: Illustration of a Conceptual Framework for a Concept of Interest Assessed by a**
 1521 **Multi-Item Observer-Reported Outcome Measure**



1522

Contains Nonbinding Recommendations

Draft — Not for Implementation

1523 **APPENDIX C: CLINICIAN-REPORTED OUTCOME MEASURES**

1524

1525 **I. INTRODUCTION**

1526

1527 ClinRO instruments are typically used when clinical judgment is needed to assess some aspect of
1528 a patient's health. ClinROs can include reports of clinical signs or events, ratings of a sign, and
1529 clinician's global assessments of the patient's current status or of the change the patient
1530 undergoes (Powers et al. 2017).

1531

Examples: ClinRO Instruments

- Reports of clinical findings, such as:
 - Counts of skin lesions
 - Presence of swollen lymph nodes
 - Presence or absence of fracture
- Rating scales, such as:
 - Psoriasis Area and Severity Index, a measure used to assess the severity and extent of a patient's psoriasis
 - Clinician global assessment of psoriasis severity, such as through a single-item verbal rating scale

1532

1533 ***ClinRO Selection and Implementation Considerations***

1534

1535 Below are key considerations and recommendations for selecting and implementing a ClinRO in
1536 a clinical study:

1537

- Include a user manual with clear instructions and directions for standardized administration.
- Conduct a standardized training with all clinician raters in the study to help ensure that rating assessments are based on consistent criteria for the ratings to minimize unwanted variability.
- Scales should be developed and tested as they will be used in the registration trial (e.g., it is inappropriate to assume the measurement properties for a dermatology scale used to assess a patient's condition by photographs will be the same when the scale is used during an in-person (non-photographic) assessment)
- Implement standardized case report form for data collection
- Evaluate intra- and inter-rater reliability prior to using a proposed ClinRO measure in a pivotal study
- If visual aids (e.g., photo guides) are used, ensure that they cover a wide variety of patient, condition, and environmental characteristics and pilot test them with clinician raters to ensure they are well understood.

1550

1551

1552

Contains Nonbinding Recommendations

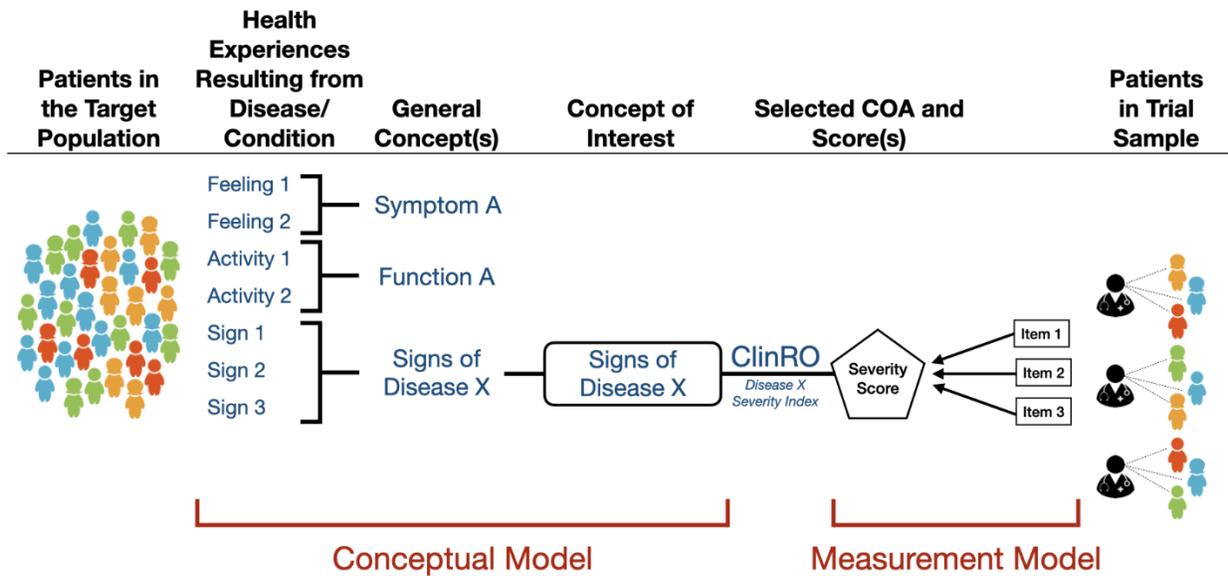
Draft — Not for Implementation

- Use a masked assessor for primary efficacy or effectiveness data collection; in some cases, a centralized blinded review and an adjudication process in the event of rating discrepancies may be necessary to ensure consistent assessment.
- To the extent feasible, the same clinician should conduct the assessments for the same patients throughout the trial to minimize unwanted variability due to different reporters.

II. CONCEPTUAL FRAMEWORK EXAMPLE OF A ClinRO MEASURE

Figure C illustrates an example of a conceptual framework for a ClinRO measure. In the example depicted, three clinical signs are associated with Disease X. These clinical signs are not direct measurements of how the patient with Disease X feels (i.e., Symptom A) or functions (i.e., Function A). Rather, there is an indirect association between the presence of the signs and worse feeling and functioning. Still, there might be interest in assessing treatment-related changes in the signs of Disease X, and so that becomes the concept of interest. Because clinical expertise is required to identify and quantify the signs appropriately, the concept of interest is measured by a ClinRO. In this case, the measure uses three items—one corresponding to each of the signs—and combines the responses in a way to generate an overall severity score. The Figure C also indicates that the items are combined, assuming a composite indicator model (see section IV.E.2).

Figure C: Illustration of a Conceptual Framework for a Concept of Interest Assessed by a Clinician-Reported Outcome Measure



1575
1576

Contains Nonbinding Recommendations

Draft — Not for Implementation

1577 **APPENDIX D: PERFORMANCE OUTCOME MEASURES**

1578

1579 **I. INTRODUCTION**

1580

1581 There are instances when patient experience data are best captured through performance tasks. A
1582 PerfO measure is a type of COA that is used to generate patient experience data through
1583 standardized task(s) performed by a patient. A PerfO measure is administered and evaluated by
1584 an appropriately trained individual or independently completed. PerfO measures are commonly
1585 used to assess patient physical or cognitive functioning, or perceptual/sensory functioning,
1586 through standardized tasks completed by the patient. The patient's performance on these tasks is
1587 then quantified and reported using defined procedures.

1588

1589 A PerfO measure can be considered for use when patient functioning is the concept(s) of interest
1590 (e.g., mobility, memory, attention, visual acuity) and the patient is able to follow the instructions
1591 to perform the required task(s). PerfO measures should not be used to capture information that is
1592 better assessed through other types of COAs, such as the severity of the symptoms of a disease or
1593 condition as captured through a PRO instrument.

1594

1595 Because PerfO instruments are based on patients' actual performance on a set of standardized
1596 tasks, they may be advantageous for the following reasons:

1597

- 1598 • When appropriately designed, PerfO measures may reduce the influence of culture and
1599 language variability on outcome assessment in multinational and multilanguage trials.
- 1600 • By having patients perform standardized tasks in a controlled, standardized environment,
1601 PerfO measures are less influenced by variability between and within patients in the types
1602 and settings of daily activities performed by the patients in their natural environment (e.g.,
1603 driving a car versus taking public transportation, living in rural area versus living in big
1604 cities).
- 1605 • By assessing real-time functioning, PerfO measures are not vulnerable to errors of recall that
1606 can occur for some PRO, ObsRO, and ClinRO measures that use a recall period (e.g., during
1607 the past 7 days).
- 1608 • PerfO measures may be less vulnerable to external changes in the patient's environment,
1609 such as seasonal impacts on daily routines.
- 1610 • Results of PerfO measures can be communicated in units that are familiar and readily
1611 interpretable such as meters (e.g., distance walked in 6 minutes), seconds (e.g., time to climb
1612 a flight of stairs), and frequency counts (e.g., number of words recalled).

1613

1614

1615 ***PerfO Selection and Implementation Considerations***

1616

1617 Although using a PerfO measure can be beneficial in a clinical trial, the following are examples
1618 of unique challenges and recommendations:

- 1619 • ***Potentially less direct relationship to a meaningful aspect of the patient's health.*** Each task
1620 usually assesses a specific function. Therefore, the patient's performance on the standardized
1621 task(s) may provide only limited information about the patient's overall functioning outside
1622 of the assessment setting.

Contains Nonbinding Recommendations

Draft — Not for Implementation

- 1623 • ***Potential interference of functions or abilities that are not part of the concept of interest.***
1624 Some PerfO tasks require multiple functions to complete. For example, fine motor skills
1625 might be important in providing a response to a neuropsychological measure of memory
1626 functioning, and so someone with fine motor impairment might receive a score that does not
1627 reflect the person’s true memory functioning. Care should be taken to ensure that functions
1628 other than the concept of interest do not unduly influence scores on the PerfO. If the patient’s
1629 cognitive ability may interfere with the performance with the tasks, sponsors should consider
1630 whether the selected PerfO measure is fit-for-purpose.
- 1631 • ***Potential for patient fatigue or burden.*** Because a PerfO measure involves assessing how
1632 well and/or how quickly a patient performs a task, it is important to consider how patient
1633 fatigue or burden may impact their performance. This is especially the case when PerfO
1634 measures are time- or effort-sensitive. When developing the clinical trial protocol, sponsors
1635 should consider the cumulative burden on the patient and the placement of the PerfO
1636 assessment. For example, in a trial for a disease in which fatigue is a primary concern for
1637 patients, it may be unwise to administer a 12-minute walk test at the end of a clinic visit day
1638 that included 3 hours of blood draws, other biological tests, and PRO measures.
- 1639 • ***Voluntary.*** Patients might refuse to perform the task at the specified time for a variety of
1640 reasons. Consider gamification to make the task more appealing and ways patients can
1641 complete the task regardless of the severity of their condition. Also consider how to record
1642 the many different types of potential missing data.
- 1643 • ***Standardization.*** If a specific published administrator’s manual is selected for the test, it is
1644 important to conduct the test in accordance with the selected manual.
- 1645 • ***Inaccessible equipment for task administration.*** Required equipment or assessment setup
1646 may not be available or feasible for certain clinical trial sites (e.g., a flight of stairs, air-
1647 conditioned rooms) or the materials may not be consistent across cultures (e.g., random
1648 words that are commonly used versus infrequent words, English words versus French words).
1649 Special attention should be paid to maintaining standardization of PerfO measures, especially
1650 in multisite and multinational clinical trials, to ensure that the assessment results are reliable,
1651 reproducible, and interpretable.
- 1652 • ***Practice effects.*** There are some instances in which patients improve their performance after
1653 repeated exposure to the same tasks, even though their underlying disease state has not
1654 changed. Steps should be implemented in trials to minimize the practice effect so that it does
1655 not confound the assessment results, including increasing the time in between PerfO
1656 assessments and allowing all patients to practice the task prior to randomization. Sponsors
1657 should consider potential learning effects associated with the selected performance-based
1658 tasks. The study protocol should include plans and/or procedures that will be put in place to
1659 minimize potential bias contributed by the learning effects on the interpretation of the PerfO-
1660 based endpoint results.
- 1661 • ***Standardized case report forms, assistive devices, and documentation.*** The use of a
1662 standardized case report form is recommended, which should include information on whether
1663 an assistive device was used during the test. The use of assistive devices should be
1664 standardized, and the type of device, if used, should be recorded. If the test was not
1665 completed, sponsors should collect the reason for not completing the test. These pieces of
1666 information should be part of the analysis data sets and may play a role in analysis and
1667 interpretation of the data.
1668

Contains Nonbinding Recommendations

Draft — Not for Implementation

1669 **II. CONCEPTUAL FRAMEWORK EXAMPLE OF A PerfO MEASURE**

1670
1671 Figure 5 (shown in the body text) illustrates a conceptual framework for a PerfO assessment. In
1672 the example,³¹ the disease impacts activities that are all instances of lower limb-related function.
1673 Perhaps because of the heterogeneity among patients in their activities and environments, the
1674 sponsor selects two subfunctions that are thought to be important to lower limb-related
1675 function—walking capacity and leg muscle strength. Note in Figure 5 dotted lines were used to
1676 represent the indirect relationship between the general health concept and the measured concepts
1677 of interest. Each of these two concepts of interest are then assessed using a PerfO measure.

³¹ Example adapted from Walton et al., 2015.

Contains Nonbinding Recommendations

Draft — Not for Implementation

1678 **APPENDIX E: EXAMPLE TABLE TO SUMMARIZE RATIONALE AND SUPPORT**
 1679 **FOR A COA**

1680
 1681 **Table A. Example Table To Summarize Rationale and Support for a [CHOOSE 1:**
 1682 **PRO/ObsRO/ClinRO/PerfO] to Measure [FILL IN CONCEPT OF INTEREST] in [FILL**
 1683 **IN TARGET POPULATION]**

	Component	Support
A	The concepts of interest, [FILL IN], should be assessed by a [PRO/ObsRO/ClinRO/PerfO], because . . .	
A.1		
A.2		
A.3		
B	The content of the [NAME OF MEASURE] includes all the important aspects of [CONCEPT OF INTEREST].	
C	[PERSON PROVIDING INFORMATION] understand the [e.g., INSTRUCTIONS, ITEMS, AND RESPONSE OPTIONS] as intended by the measure developer.	
D	Scores from the [NAME OF MEASURE] are not overly influenced by processes/concepts that are not part of [CONCEPT OF INTEREST]. <i>[Select and comment on appropriate rows for the type of COA]</i>	
D.1	[ITEM OR TASK] interpretations or relevance do not differ substantially according to respondents' demographic characteristics (including sex, age, and education level) or cultural/linguistic backgrounds or physical environment.	
D.2	Recollection errors do not overly influence assessment of the concept of interest. [PRO, ObsRO, and ClinRO measures]	
D.3	Respondent fatigue or burden does not overly influence assessment of the concept of interest. [PRO, ObsRO, ClinRO, and PerfO measures]	
D.4	The mode of assessment does not overly influence assessment of the concept of interest. [PRO, ObsRO, ClinRO, and PerfO measures]	
D.5	Expectation bias does not unduly influence assessment of the concept of interest. [PRO, ObsRO, ClinRO, and PerfO measures]	
D.6	Practice effects do not overly influence the assessment of the concept of interest. [PerfO measures]	
E	The method of scoring responses is appropriate for assessing [CONCEPT OF INTEREST]. <i>[Select E.2 or E.3 if appropriate. E.1 and E.4 are likely appropriate for all COAs.]</i>	
E.1	Responses to an Individual [ITEM OR TASK]	
E.2	Rationale for Combining Responses to Multiple [ITEMS OR TASKS]	
E.3	Scoring Approaches Based on Computerized Adaptive Testing	
E.4	Approach to Missing [ITEM OR TASK] Responses	
F	Scores from the [NAME OF MEASURE] correspond to the specific health experience(s) the patient has related to [CONCEPT OF INTEREST].	
G	Scores are sufficiently sensitive to reflect clinically meaningful changes within patients over [TIME] in the [CONCEPT OF INTEREST] within [CONTEXT OF USE].	
H	Differences in assessment scores can be interpreted and communicated clearly in terms of the expected impact on patients' experiences	

1684