U.S. FOOD & DRUG ADMINISTRATION

Public Webinar

Patient-Focused Drug Development: Incorporating Clinical Outcome Assessments into Endpoints for Regulatory Decision Making – Draft Guidance

May 4, 2023

#PFDD

U.S. FOOD & DRUG ADMINISTRATION

Welcome and Overview

Shannon Sparklin, MS
Patient-Focused Drug Development
Office of the Center Director
Center for Drug Evaluation and Research

Laura Lee Johnson, PhD
Division Director, Division of Biometrics III
Office of Translational Science
Center for Drug Evaluation and Research

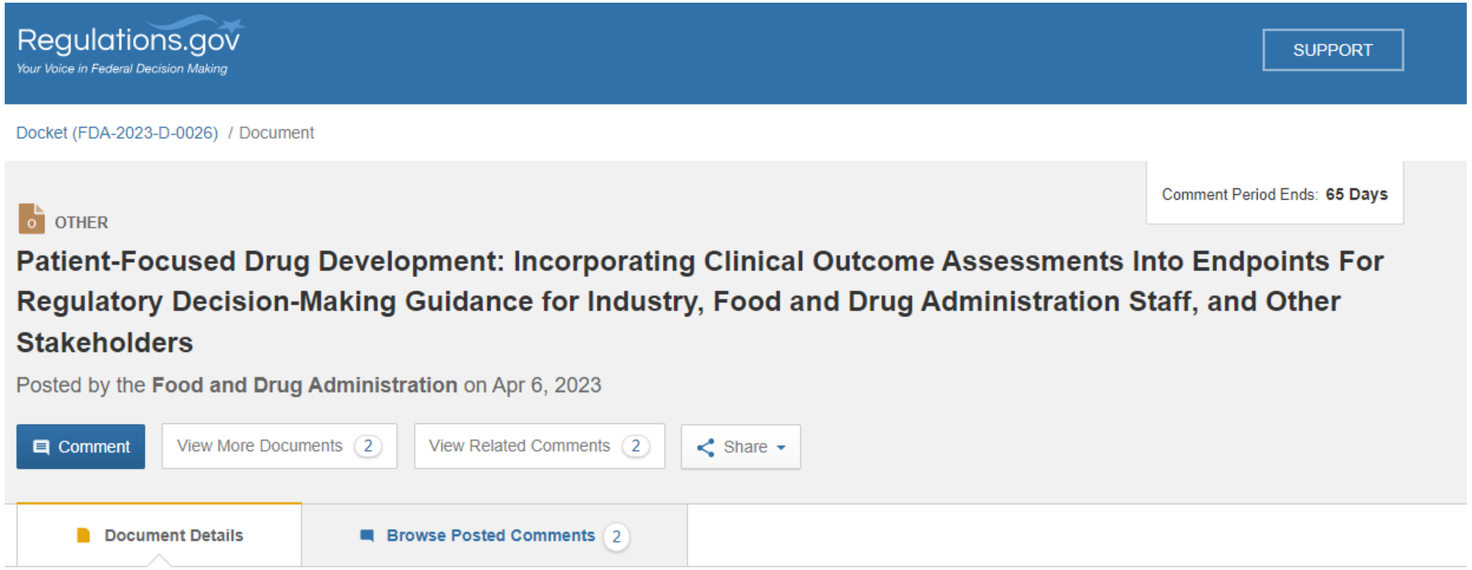
Send us your comments!

Interested stakeholders are invited to submit comments on the draft guidance to the public docket.

The docket will close on July 5, 2023.

How do you submit a comment?

- Please visit: https://www.regulations.gov/docket/FDA-2023-D-0026
- And **Click Comment**

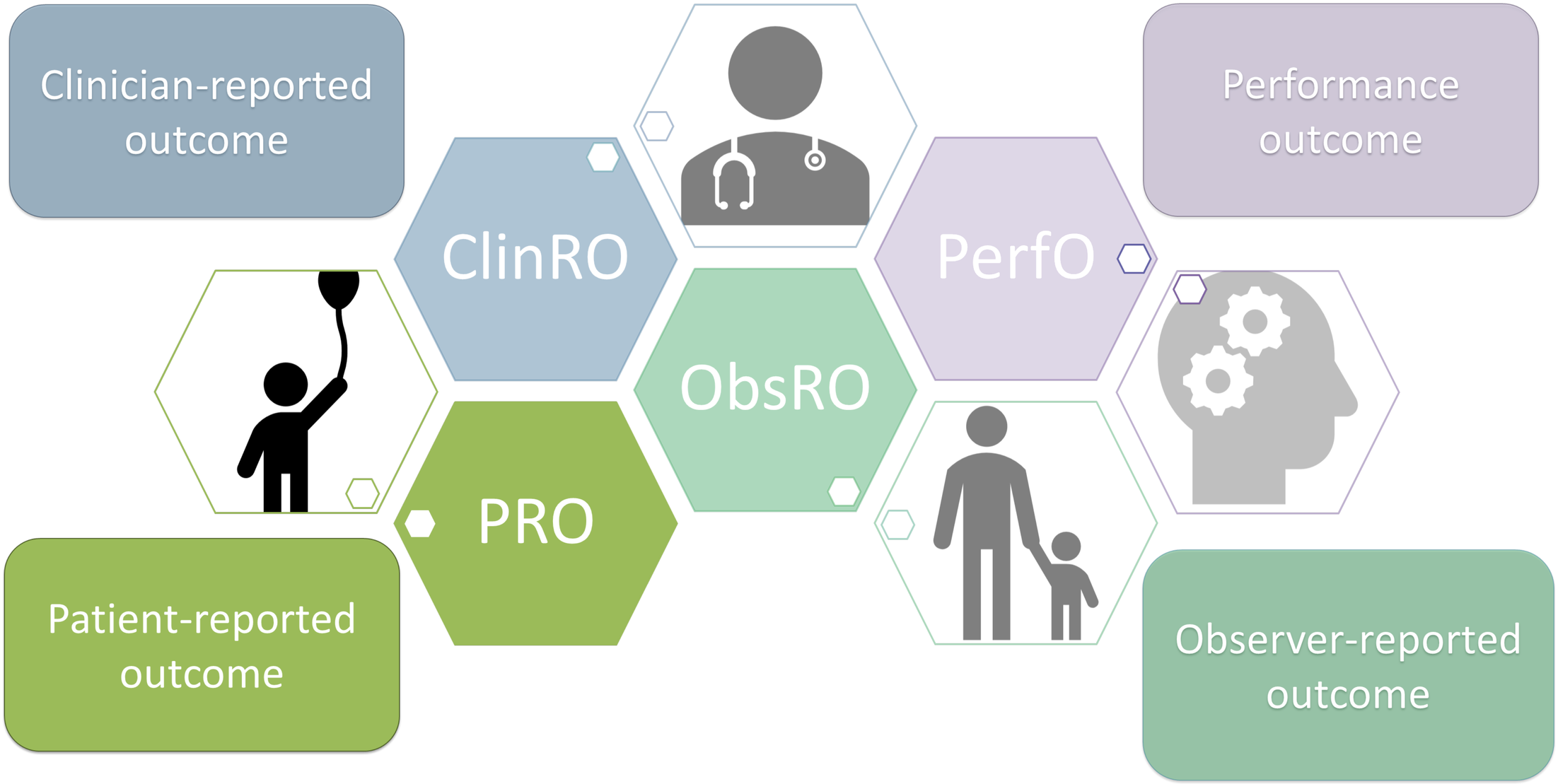


Agenda

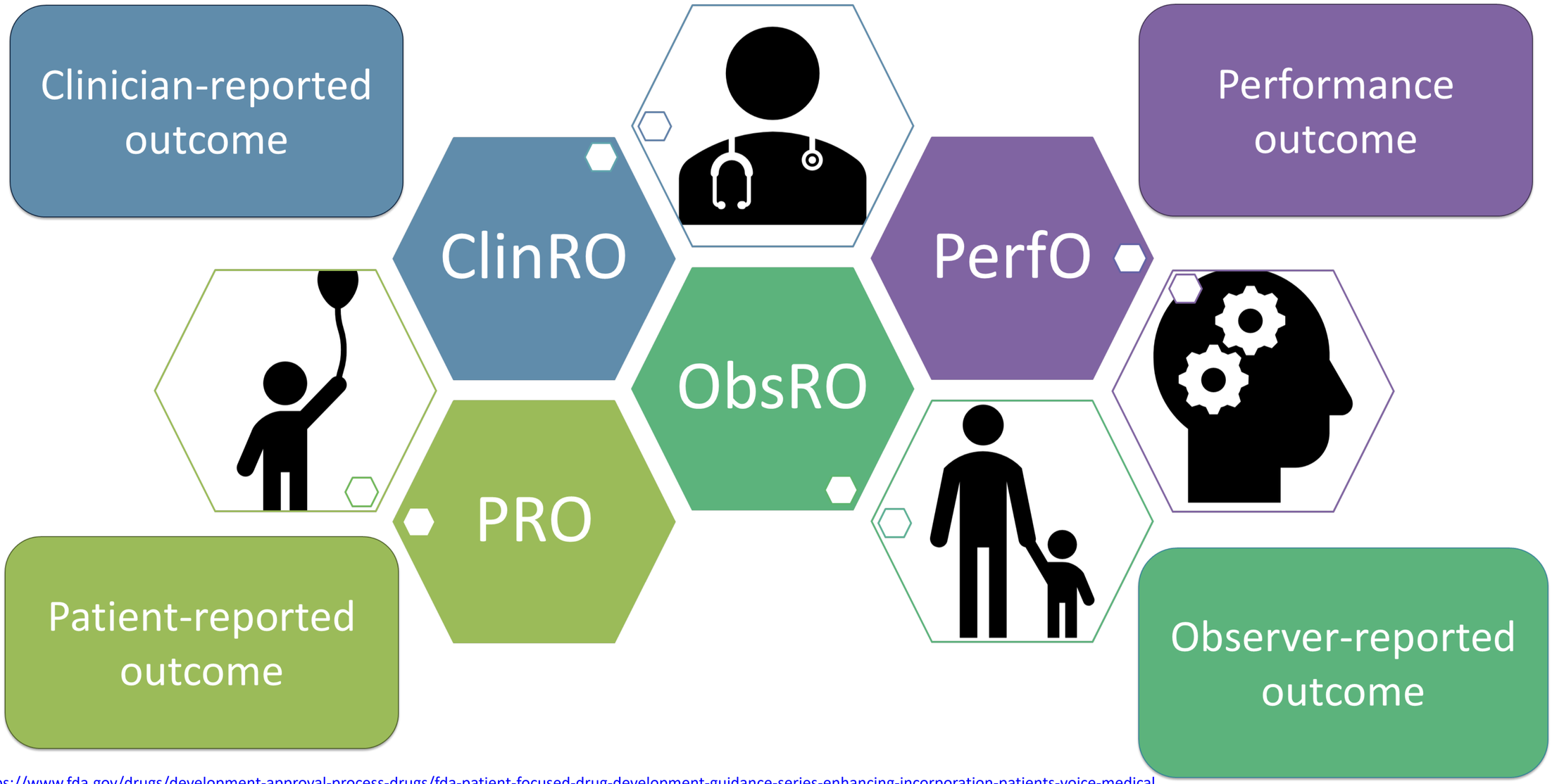
- Welcome and Overview of PFDD Methodologic Guidance Series
- Constructing COA-based Endpoints
- Obtaining Patient Input to Inform Selection of COA-based Endpoints
- Analyzing COA-based Endpoints
- Clinician Perspective
- Introduction to Evaluating Meaningfulness of Treatment Benefit
- Approaches for Collecting Evidence to Support Interpretability of COA-based Endpoints
- Applying Information about Meaningful Score Differences or Meaningful Score Regions to Clinical Trial Data
- Question and Answer

Overview of PFDD Methodologic Guidance Series

2009 PRO Guidance



Four Types of COAs



https://www.fda.gov/drugs/development-approval-process-drugs/fda-patient-focused-drug-development-guidance-series-enhancing-incorporation-patients-voice-medical

https://www.fda.gov/drugs/development-approval-process-drugs/fda-patient-focused-drug-development-guidance-series-enhancing-incorporation-patients-voice-medical

PFDD Guidance Series: Guidance Path

(First) Discussion Phase

- Discussion document(s)
- Public workshop
- Docket comments
- Internal FDA discussions
  - Docket comments
  - Feedback
    - Meetings (e.g., public, professional society meetings)
    - International development programs
    - Daily interactions with many different stakeholders
  - Literature

PFDD Guidance Series: Guidance Path

Discussion phase

- Discussion document(s)
- Public workshop
- Docket comments
- Internal FDA discussions

Draft guidance

- Docket comments
- Internal FDA discussions

PFDD Guidance Series: Guidance Path

Discussion phase

- Discussion document(s)
- Public workshop
- Docket comments
- Internal FDA discussions

Draft guidance

- Docket comments
- Internal FDA discussions

Final guidance

Trainings, public workshops

PFDD Guidance Series

1) **Guidance 1[F]:** Collecting Comprehensive and Representative Input

2) **Guidance 2[F]:** Methods to Identify What is Important to Patients

3) **Guidance 3[D]:** Selecting, Developing or Modifying Fit-for-Purpose Clinical Outcome Assessments

4) **Guidance 4[D]:** Incorporating Clinical Outcome Assessments into Endpoints for Regulatory Decision Making

F: Final Guidance; **D:** Draft Guidance
https://www.fda.gov/drugs/development-approval-process-drugs/fda-patient-focused-drug-development-guidance-series-enhancing-incorporation-patients-voice-medical

Using COAs in Clinical Research

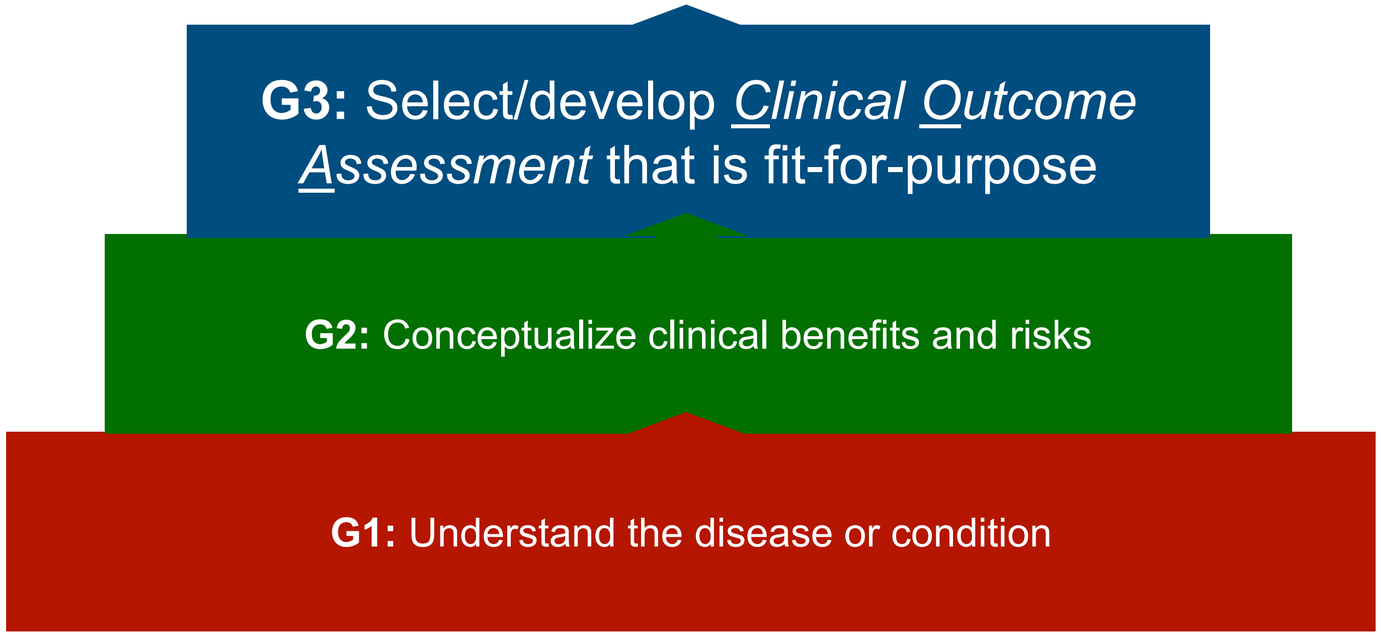
G1: Understand the disease or condition

Using COAs in Clinical Research

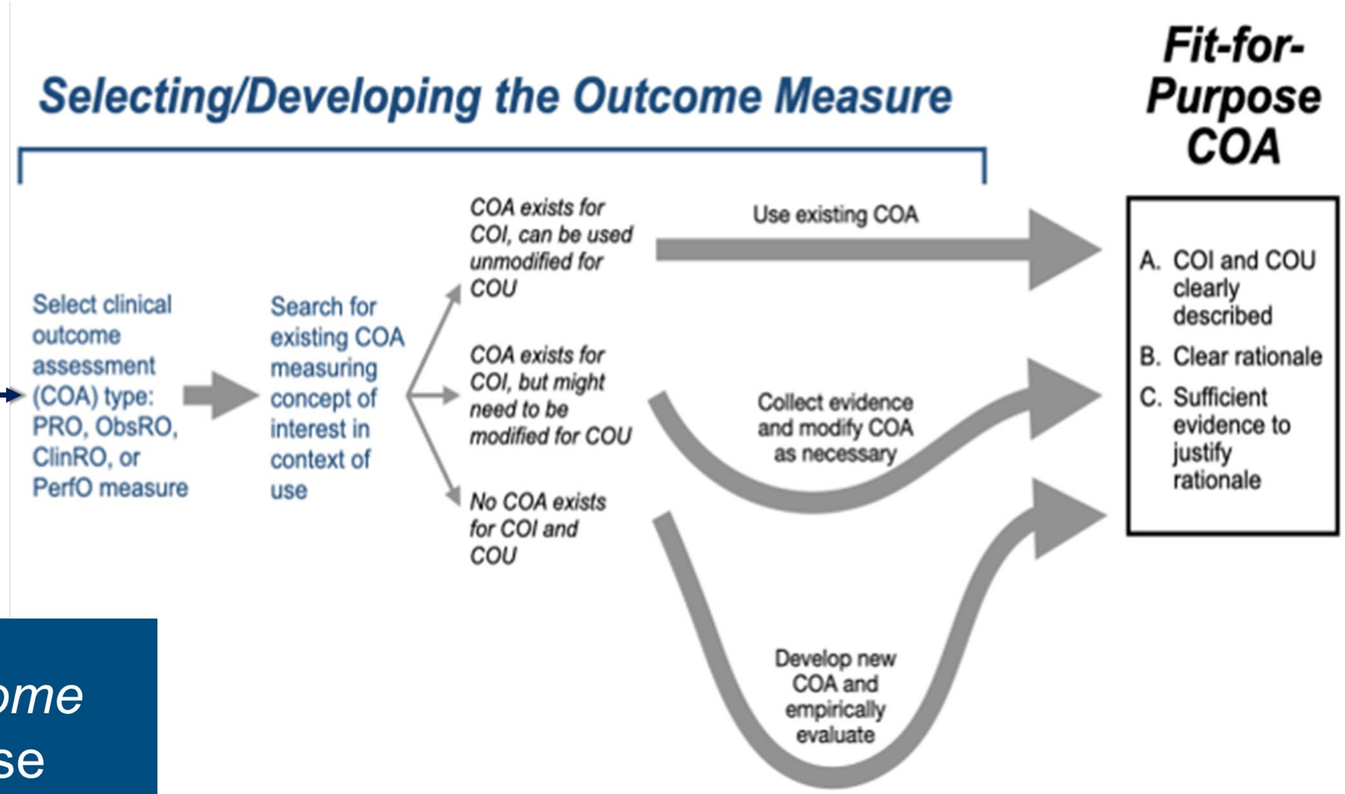
G2: Conceptualize clinical benefits and risks

G1: Understand the disease or condition

Using COAs in Clinical Research



Using COAs in Clinical Research

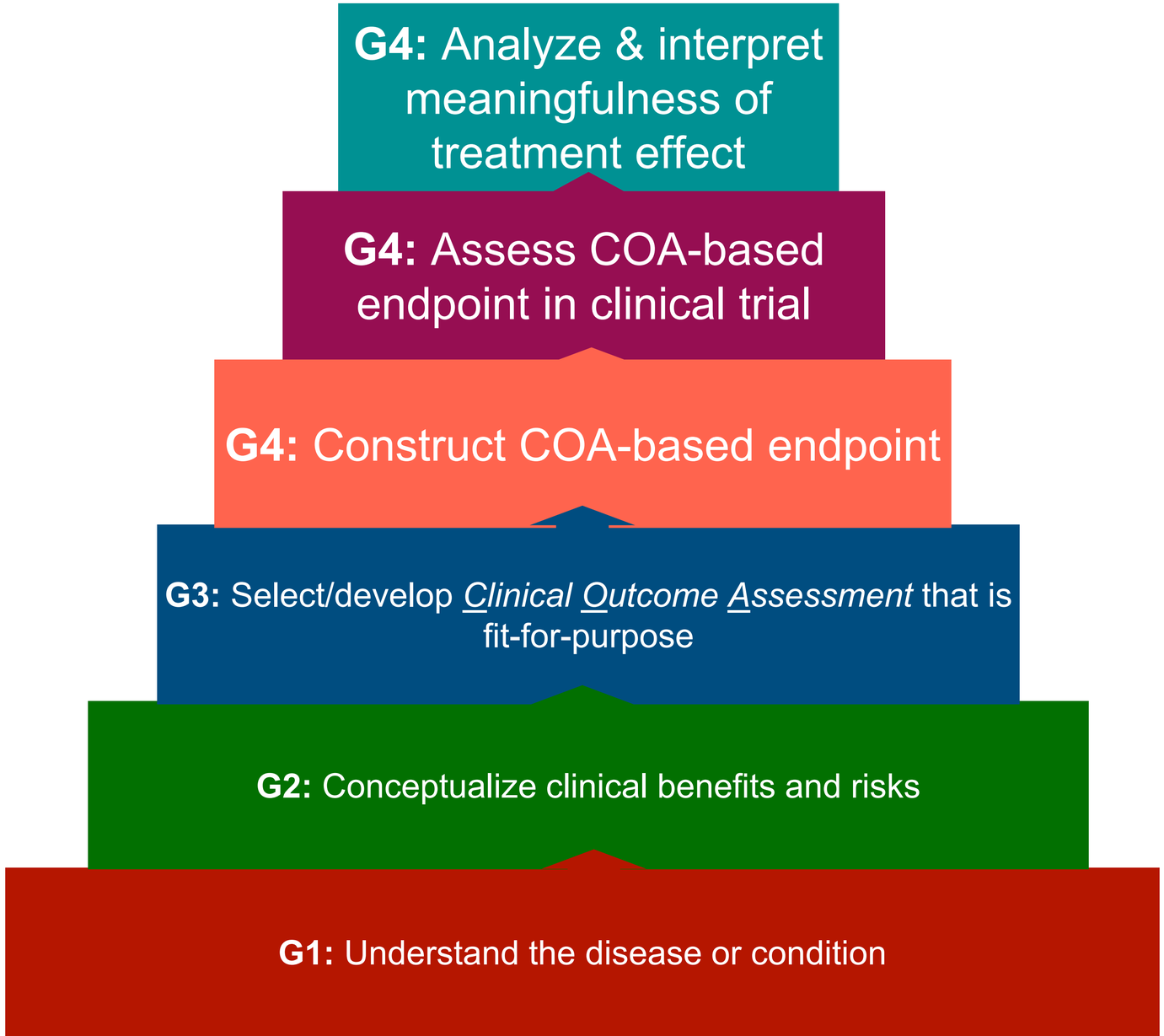


G3: Select/develop *Clinical Outcome Assessment* that is fit-for-purpose

G2: Conceptualize clinical benefits and risks

G1: Understand the disease or condition

Using COAs in Clinical Research



COA-Based Endpoint Considerations

Lili Garrard, PhD
Division of Biometrics III
CDER/OTS/Office of Biostatistics

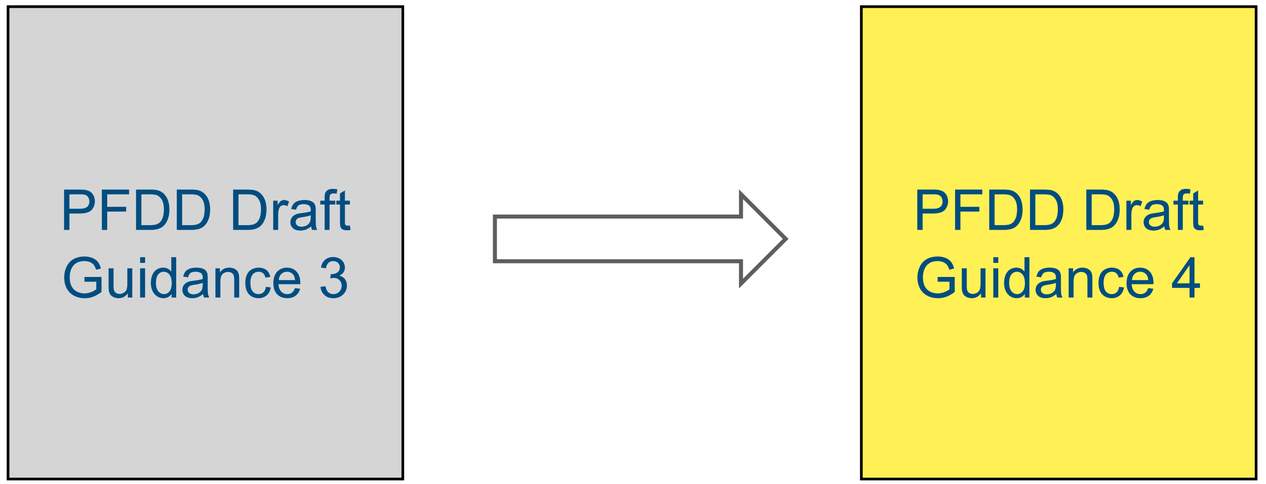
Purpose of A COA-Based Endpoint

- Reflect an aspect of the patient's health that is meaningful
- Be capable of supporting an inference of treatment effect within the context of the planned clinical trial

Definition of A COA-Based Endpoint

- Type of COA assessment(s) made (e.g., a PRO measure)
- The COA(s) used to measure the concept(s) of interest (see draft PFDD Guidance 3 for considerations of fit-for-purpose COAs)
- Specific score(s) from the COA (e.g., a total score)
- Clear definition of baseline, if applicable
- Timing of assessments, timeframe over which COA scores are combined, and how COA scores are combined into an endpoint
- Rules for handling missing item responses or task results for scoring, along with justification for the rules
- If multi-component endpoint, the algorithm used to combine scores from components into a single endpoint

#1: Sponsors should provide a well-justified rationale for the choice of endpoint(s).



Rationale for the interpretation of COA scores as measures of the concept of interest

Rationale for the choice of endpoint based on COA scores

Endpoint Selection: A Well-Justified Rationale

- ✓ Concept(s) of interest (e.g., abdominal pain)
- ✓ Clinical trial objective or hypothesis corresponding to the endpoint
- ✓ Role of the endpoint (e.g., primary, secondary, or exploratory)
- ✓ Intended indication related to the COA-based endpoint
- ✓ Explanation for why the selected COA is fit-for-purpose in the planned trial (see PFDD Draft Guidance 3)
- ✓ Support for importance of endpoint to patients and/or caregivers
- ✓ If a multi-component endpoint, justification for components included and the algorithm for combining them into the endpoint
- ✓ Strengths and limitations of the proposed endpoint

Recommended COA-Based Endpoint(s)

- COA score at a predefined assessment point, i.e., fixed time point
  - Needs justification for the use of, and time at which, an analysis at the fixed time point is to be performed
- COA scores summarized over predefined assessment period
  - Different summaries may be appropriate
  - Needs Justification and should consider
    - Robustness of the summary (or model) and any modeling assumptions
    - Handling of missing COA scores
    - Power
    - Interpretability

#2: Endpoints defined as COA scores at a fixed f/u time (analyzed by conditioning on baseline COA score) are generally preferred over endpoints defined in terms of responder status, change-from-baseline scores, or percent change-from-baseline scores.

"In most situations in which a COA produces ordinal or continuous (interval or ratio scale) scores, the best and recommended endpoint will be the COA score at a predefined assessment point or summarized over some predefined post-baseline assessment period,

#2: Endpoints defined as COA scores at a fixed f/u time (analyzed by conditioning on baseline COA score) are generally preferred over endpoints defined in terms of responder status, change-from-baseline scores, or percent change-from-baseline scores.

"In most situations in which a COA produces ordinal or continuous (interval or ratio scale) scores, the best and recommended endpoint will be the COA score at a predefined assessment point or summarized over some predefined post-baseline assessment period, and the most straightforward analysis will be a comparison of randomized groups with respect to the follow-up score(s) after adjusting for the baseline value (e.g., with a linear model to compare average follow-up scores)."

Other Common COA-Based Endpoint(s)

Responder Status

Change-from-Baseline Scores

Percent Change-from-Baseline Scores

Draft Guidance 4 describes specific limitations and considerations for each of these approaches

Responder Status

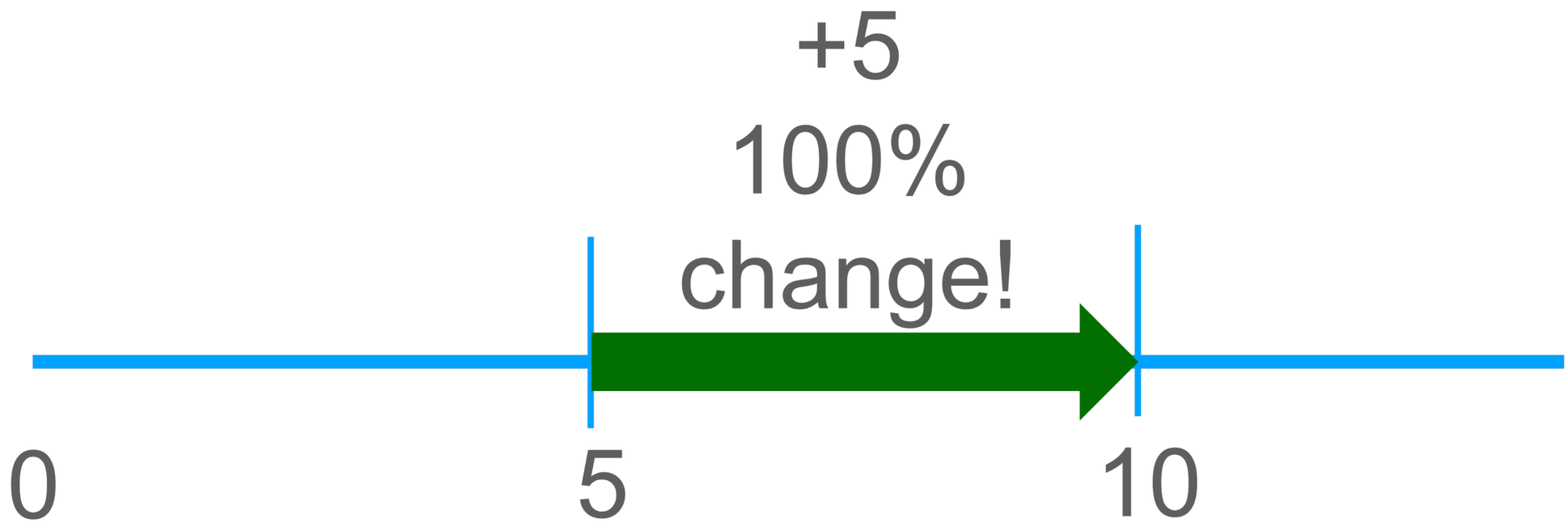
- Constructed by dichotomizing ordinal or continuous COA score(s)
- May use less information and reduce power
- In some cases, may be reasonable when evaluating effect of treatment on probability of achieving clearly defined and important health state (e.g., complete symptom resolution)
- Important to **prespecify** a single score threshold and **provide compelling justification** for dichotomization
  - Example: strong qualitative evidence that patients and/or caregivers view health states above the threshold to be meaningfully different from health states below the threshold
  - **Data used to derive a score threshold should differ from data used to demonstrate effectiveness**

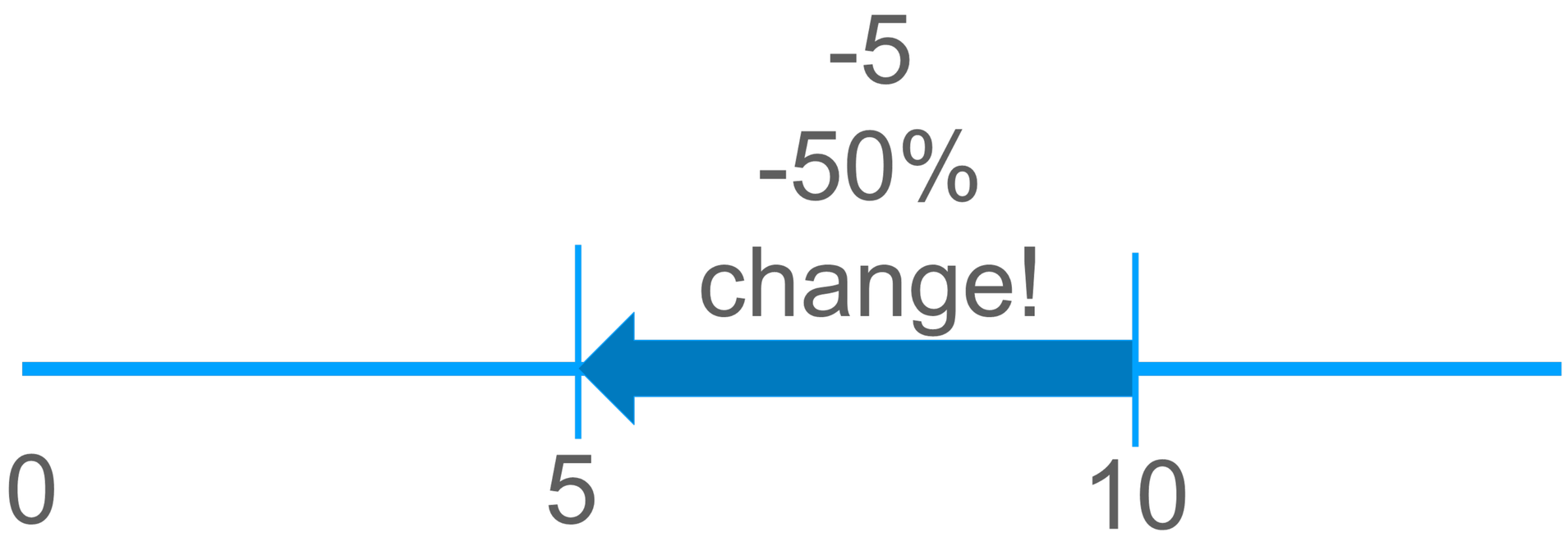
Change-from-Baseline Scores

- May be challenging to interpret for ordinal COA scores
  - Linear relationship between ordinal values and true symptom severity level may not exist
- Preferred method for adjustment for baseline is in context of a model (e.g., linear model)
- For situations in which a single-arm trial is the only viable option, a change-from-baseline endpoint might be best available option

Percent Change-from-Baseline Scores

- Interpretation complicated by asymmetric nature of endpoint: treats baseline and follow up COA scores differently


+5
100%
change!
0
5
10


-5
-50%
change!
0
5
10

Percent Change-from-Baseline Scores

- 100% improvement, 50% decline -> averages to 25% improvement!
    - Or an average of 0 change
- Undefined if baseline COA score is zero
- May have highly non-normal distributions that can be challenging to model
- If effect of treatment is expected to be multiplicative rather than additive, then log or similar transformation could be applied to continuously distributed COA scores

When Should These Other Endpoints Be Used?

Responder Status

Change-from-Baseline Scores

Percent Change-from-Baseline Scores

If considered because results will be easier to understand for some stakeholders, then sponsors can…

(1) First, estimate treatment effect using a model for ordinal or continuous COA scores, and then

(2) Additionally communicate results of the model in terms of thresholds, change-from-baseline, percent change-from-baseline, or whatever helps

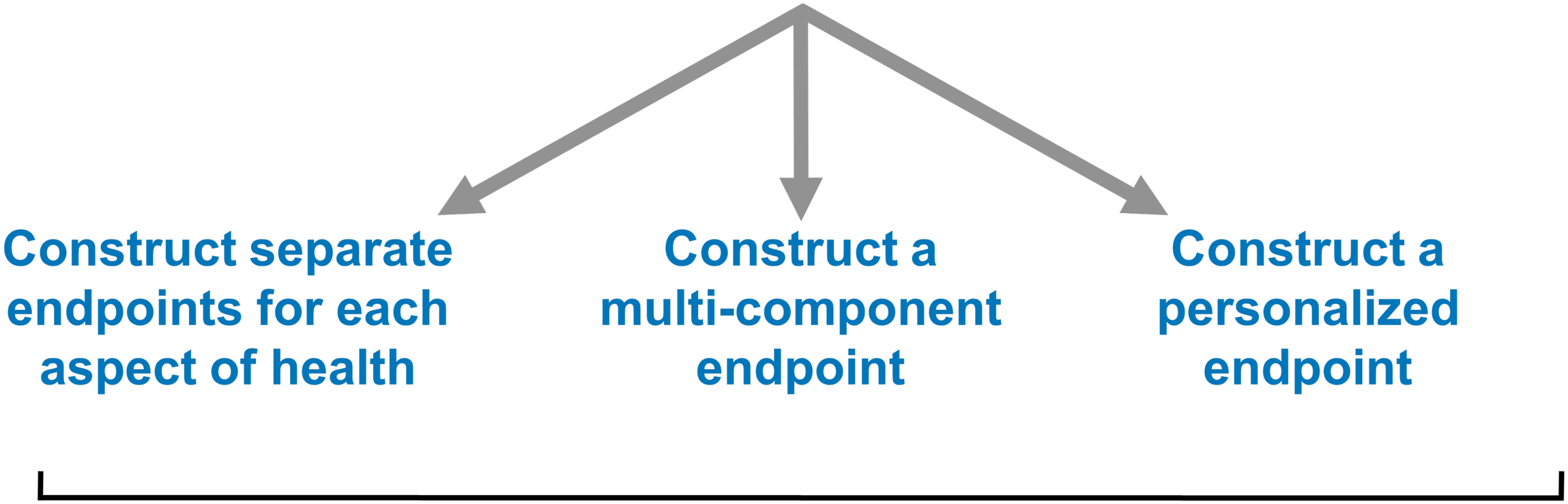
Heterogeneity In Diseases

No perfect endpoint strategy when a disease affects multiple aspects of feeling and functioning

- Maybe necessary to consider several different aspects to adequately assess benefit
- Should consider the strengths and limitations of various approaches
- When possible, evaluate multiple endpoints in earlier studies to inform endpoint selection for later studies

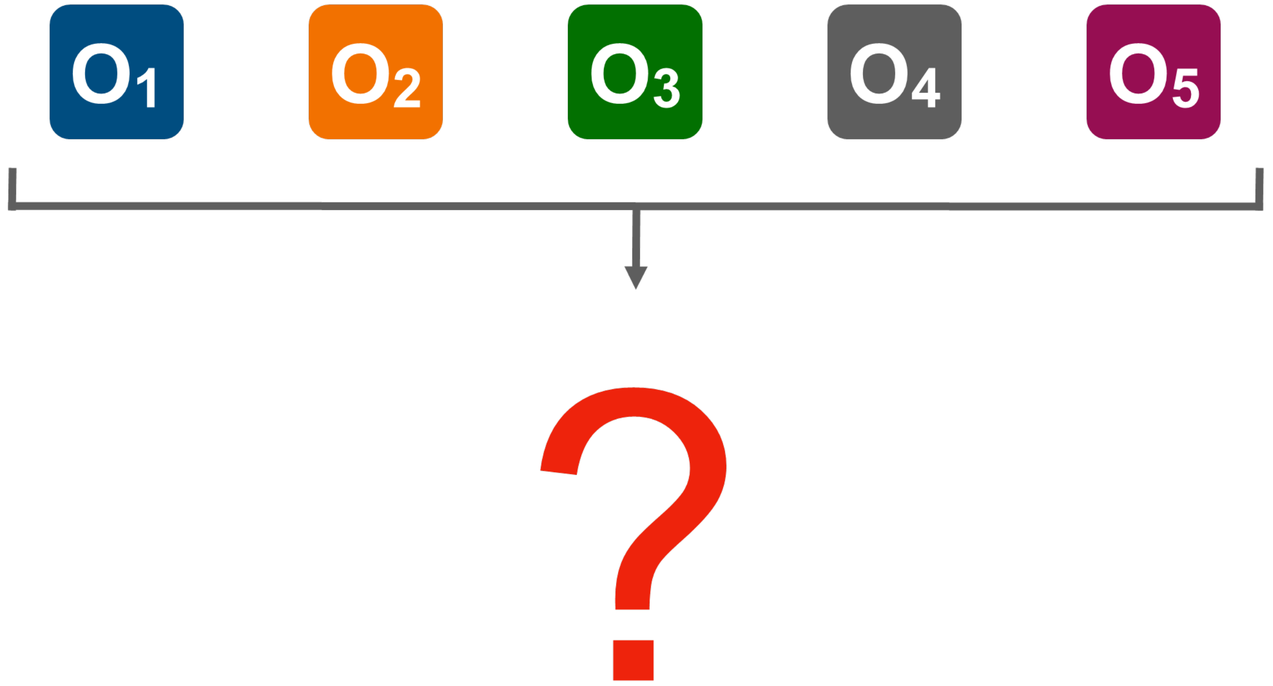
Also see FDA guidance for industry *Multiple Endpoints in Clinical Trials* (October 2022)

#3: There is no perfect endpoint strategy when a disease affects multiple aspects of feeling and functioning, so sponsors should choose the best for their context of use.



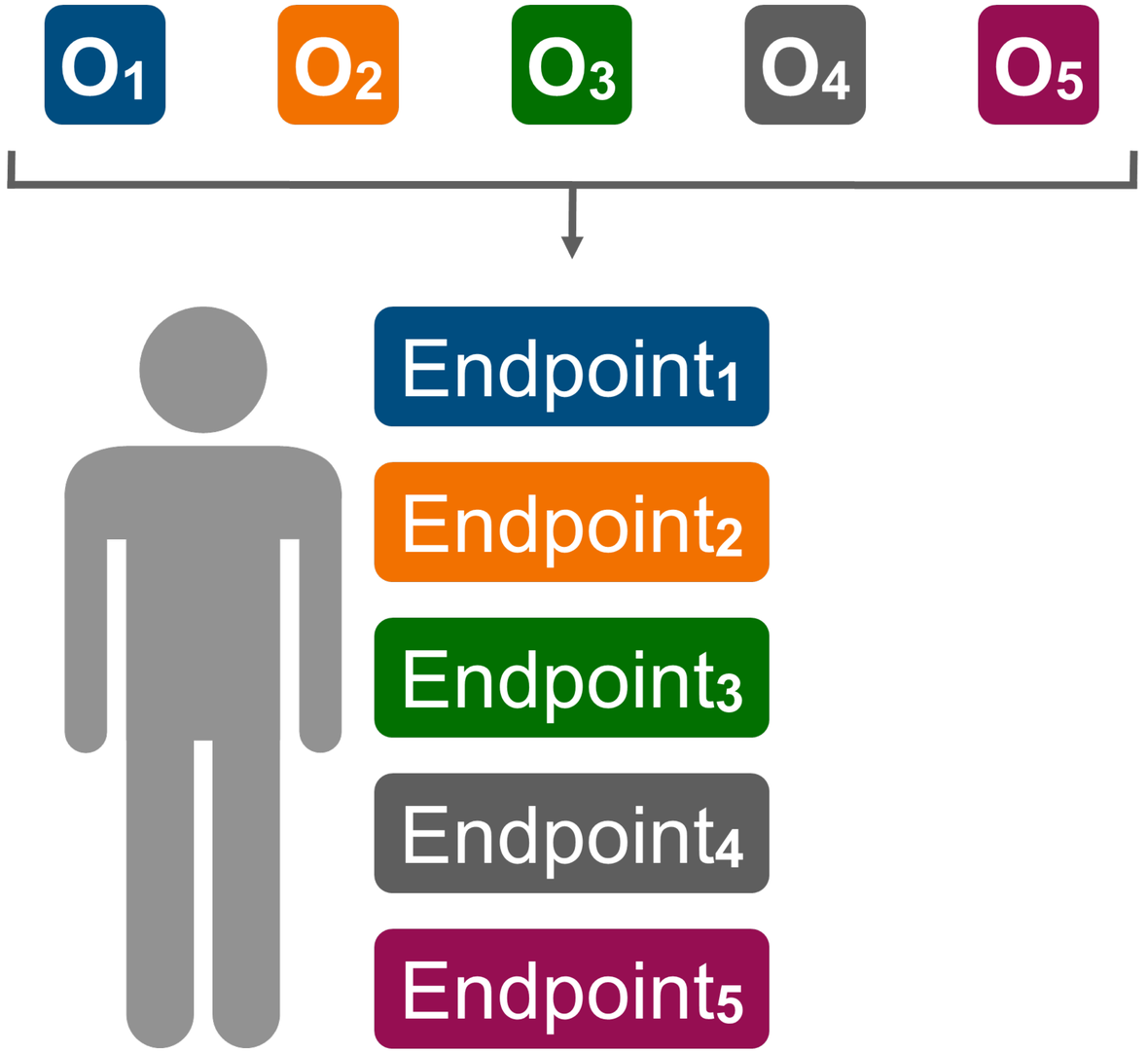
Heterogeneity In Diseases

Multiple outcome variables associated with a disease



Heterogeneity In Diseases

Multiple outcome variables associated with a disease



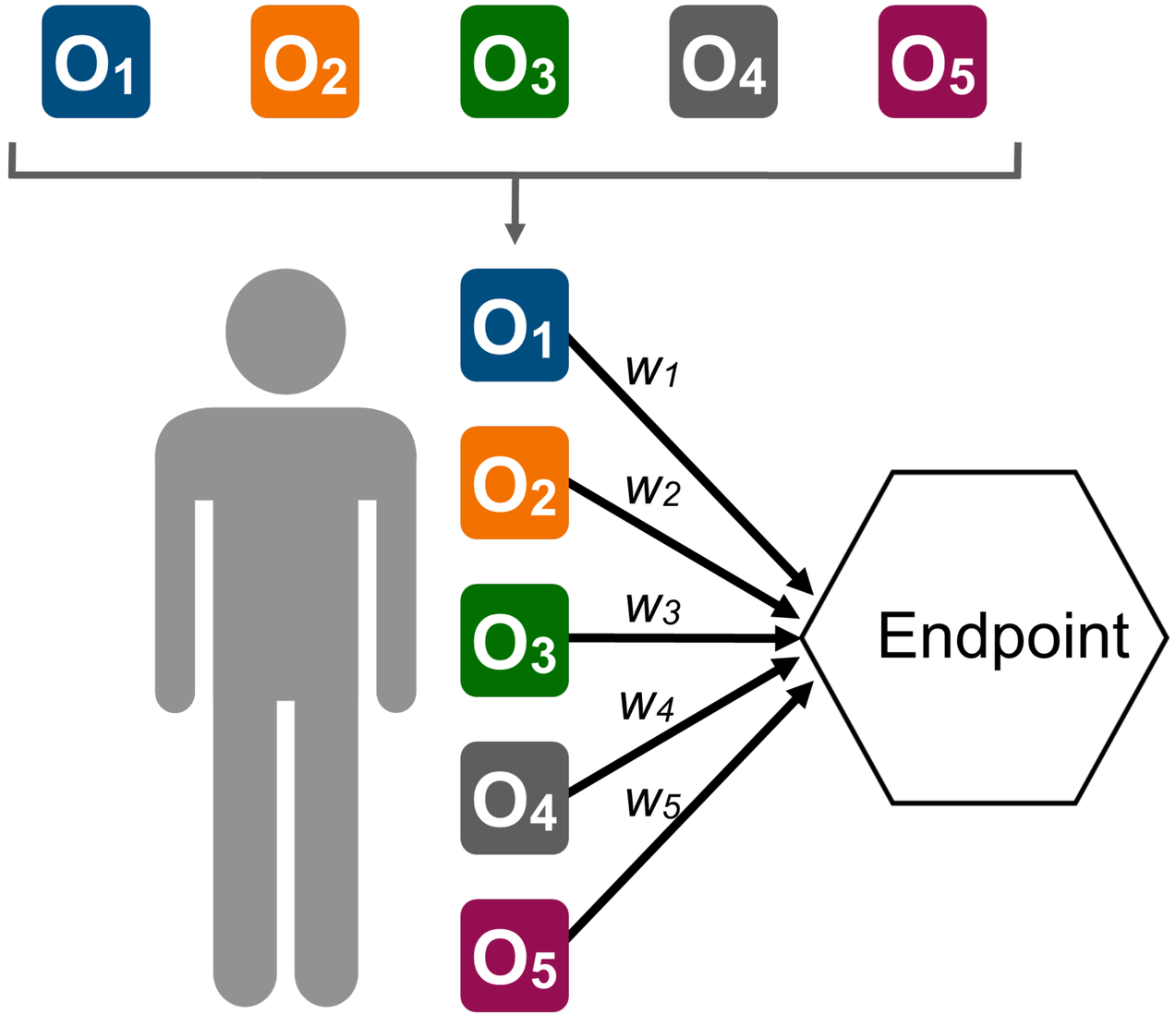
Construct separate endpoints for each aspect of health

Separate Endpoints For Each Aspect of Health

- Strength: Clarity about which aspect of health is affected by medical product
- Challenges
  - Aspect(s) of health affected by medical product not always known ahead of time
  - Depending on role of endpoints, multiplicity adjustments might be needed, resulting in larger sample size
  - If patients differ in aspect of health affected, then treatment effect for any one endpoint will be diluted

Heterogeneity In Diseases

Multiple outcome variables associated with a disease

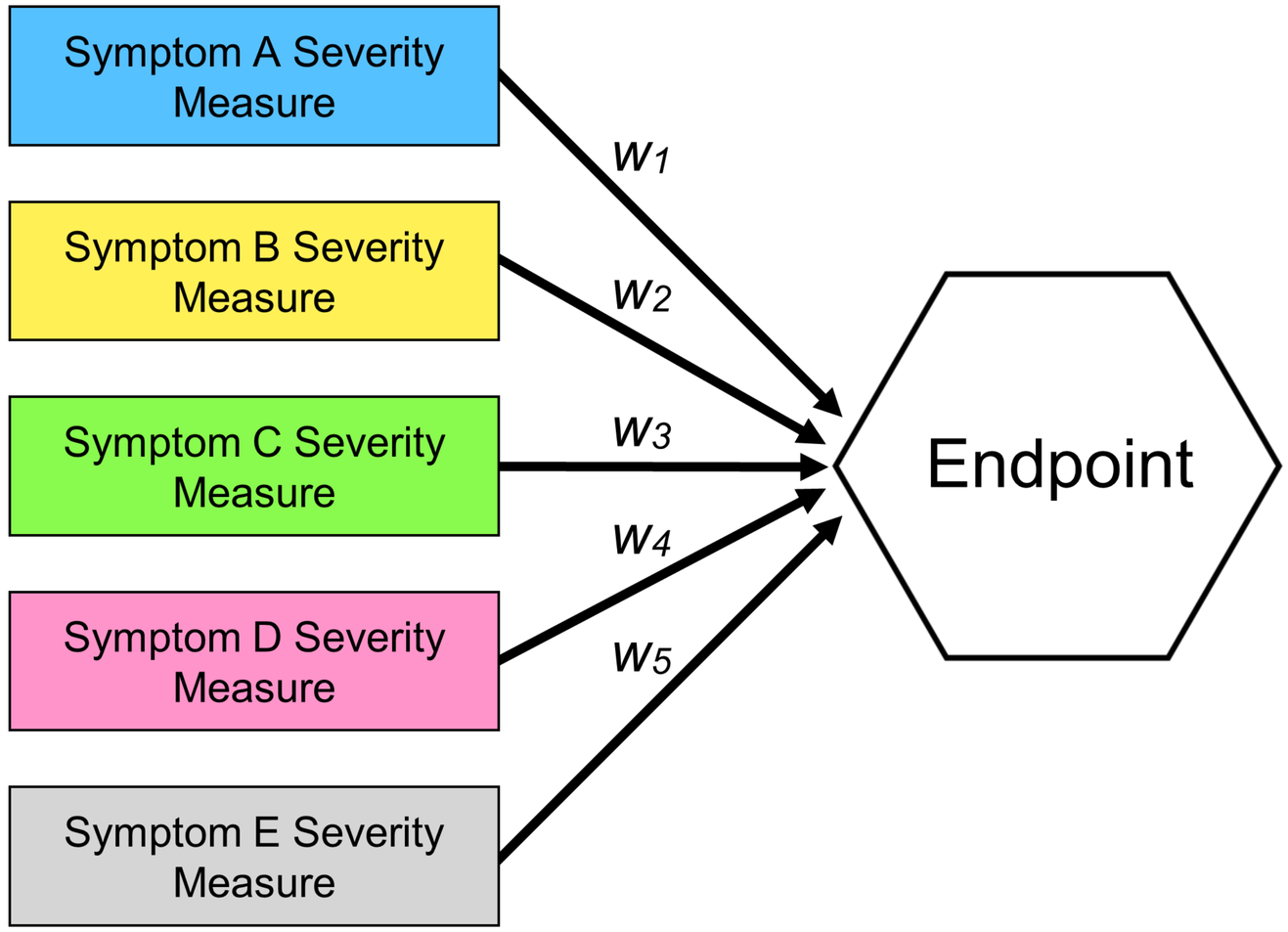


Construct a multi-component endpoint

"[A] within-subject combination of two or more components"

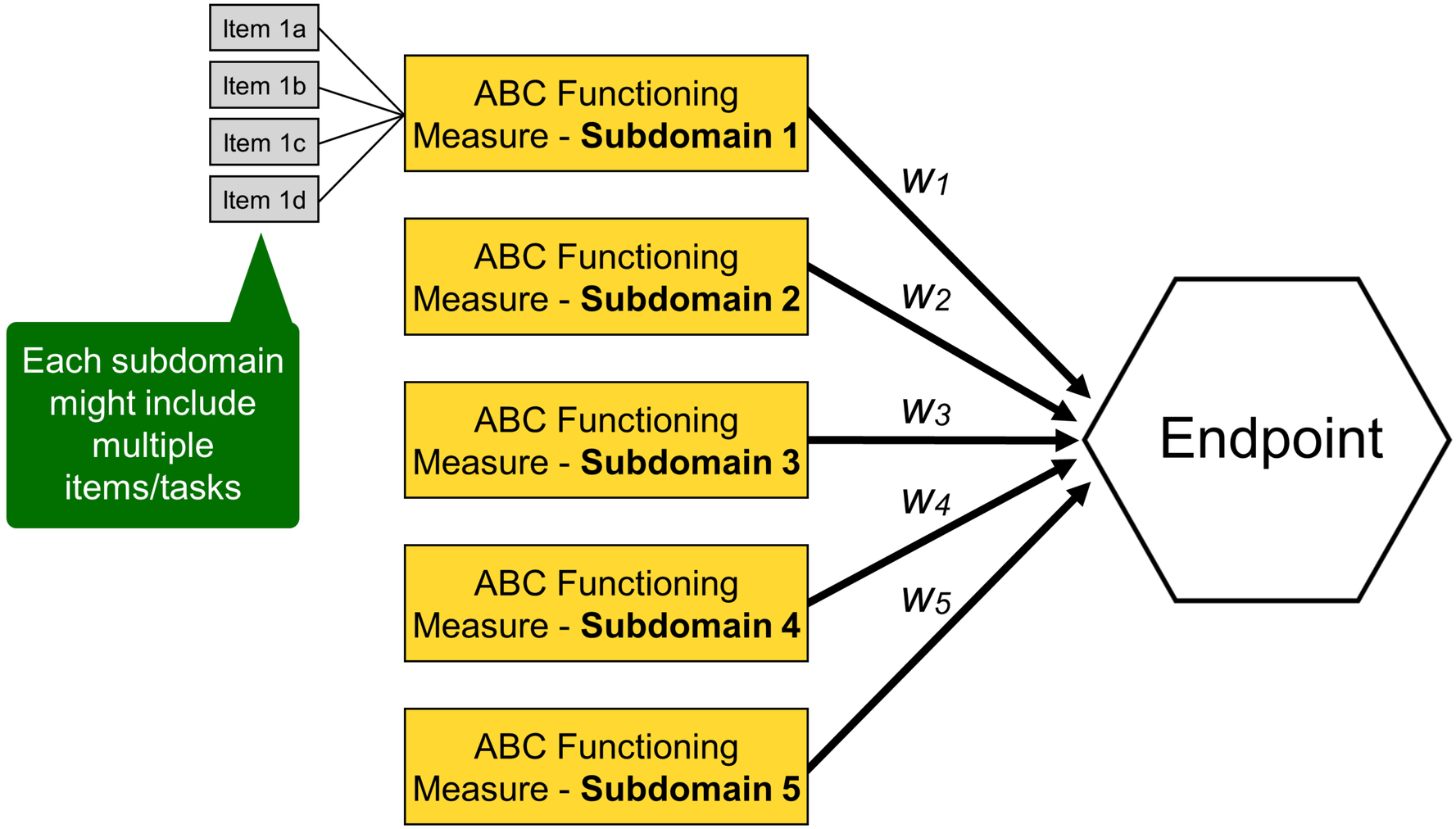
Multi-Component Endpoint

<u>Option 1</u>: Each component could be the score from a different COA



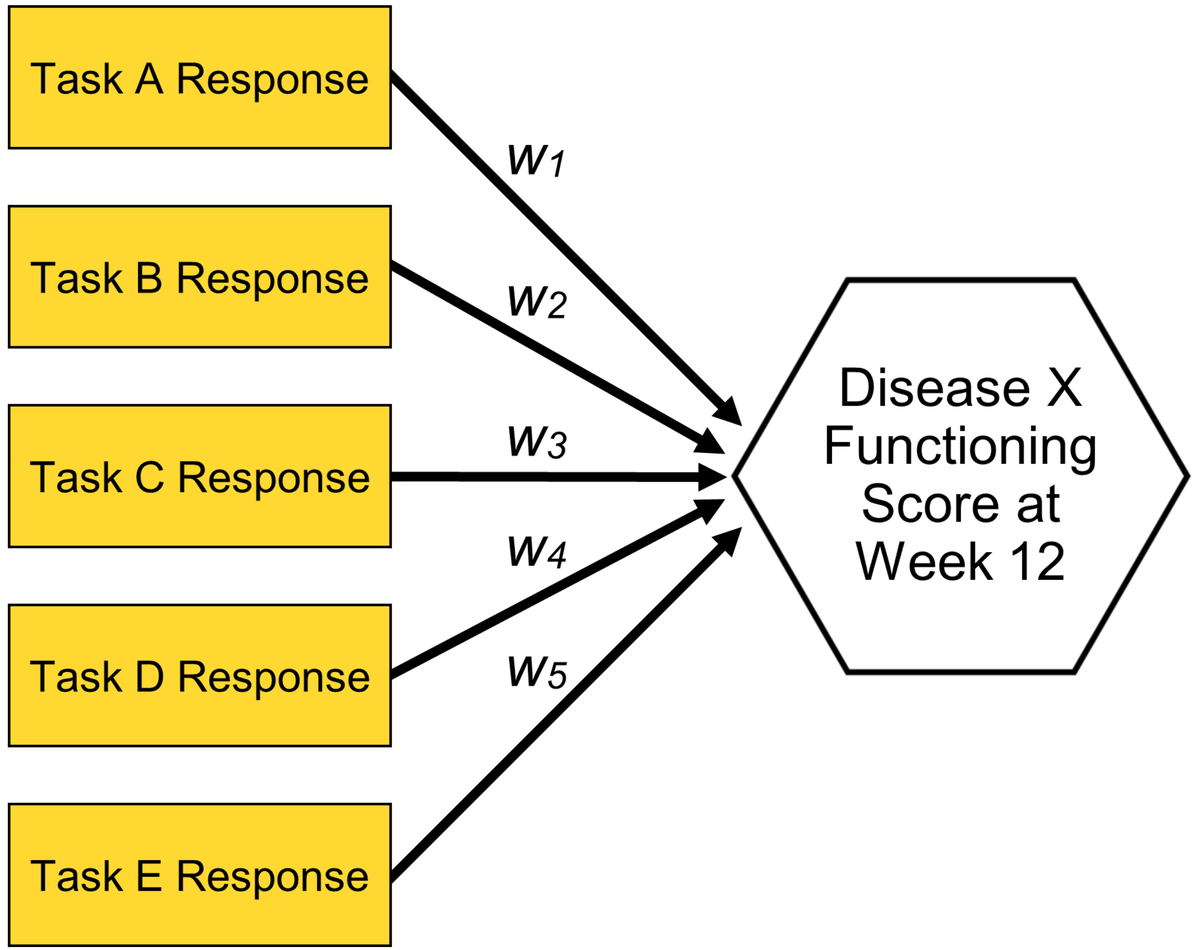
Multi-Component Endpoint

<u>Option 2</u>: Each component could be the score from a subdomain of a single, multidimensional COA



Multi-Component Endpoint

<u>Option 3</u>: Each component could be the response to an item/task from a single COA

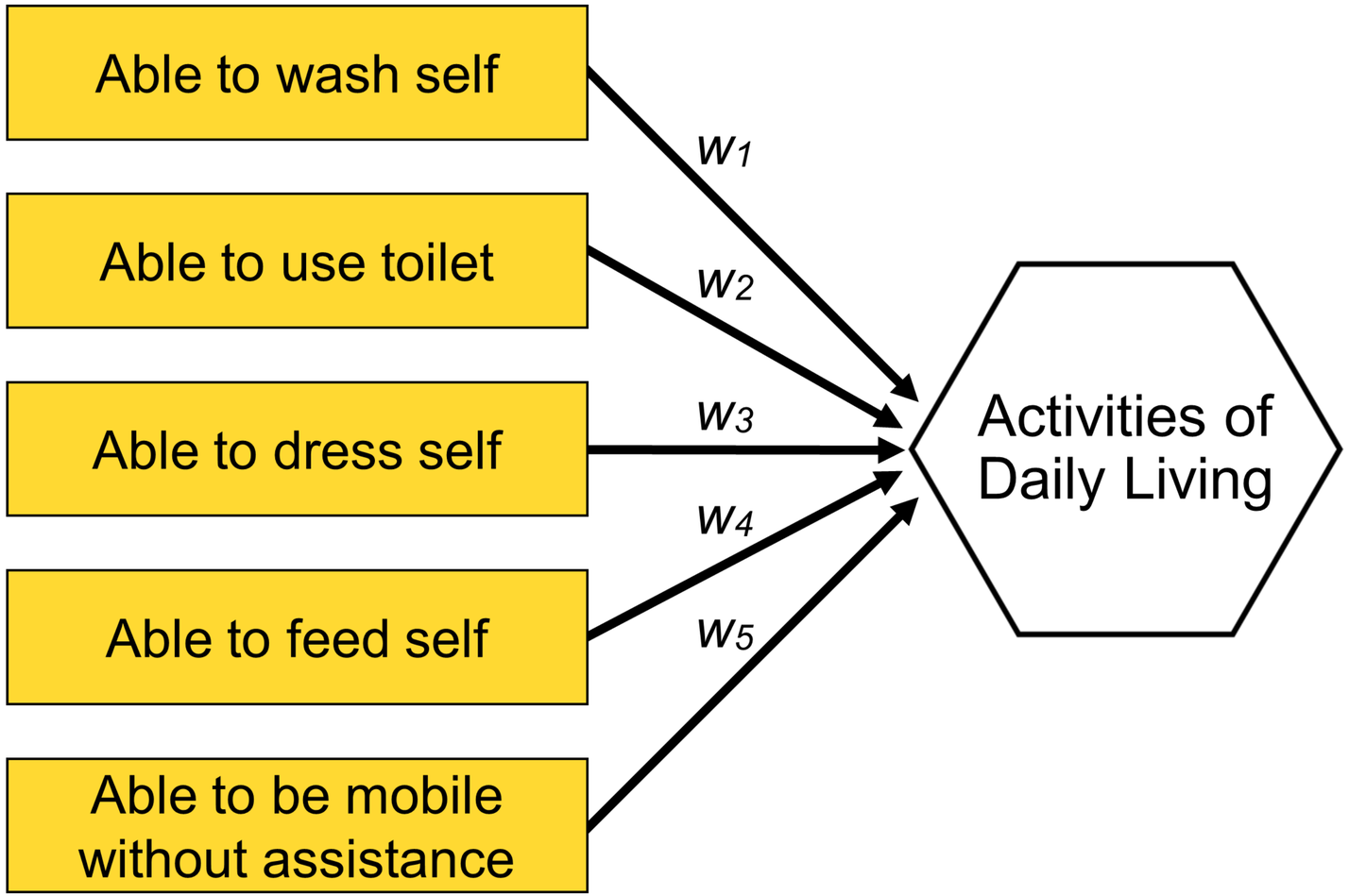


Multi-Component Endpoint

Unique consideration for Option 3 (*each component could be the response to an item/task from a single COA*) – only when the COA is based on a **composite indicator model**

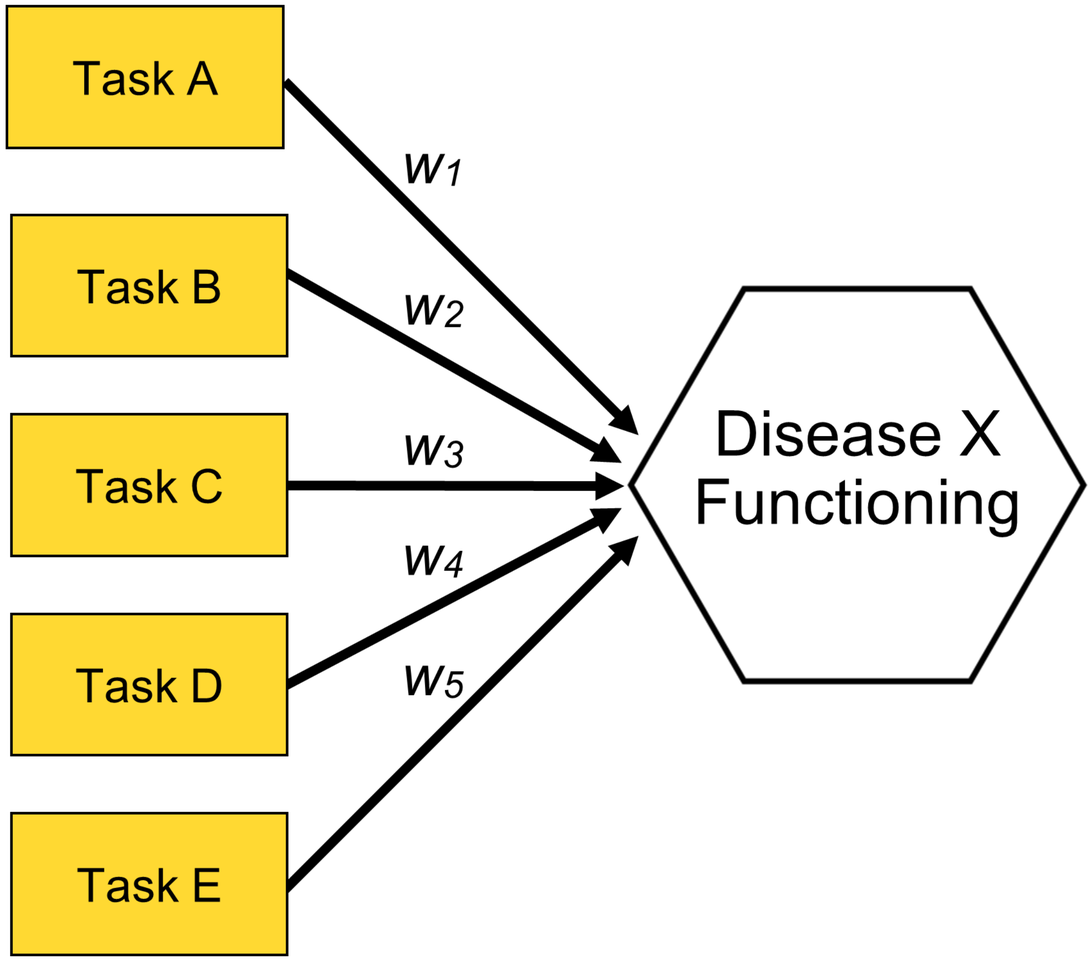
- See draft PFDD Guidance 3
- Responses to the items or tasks are not assumed to be reflective of or caused by a single underlying aspect of health
- Each item or task addresses a separate health concept and, when combined, responses to all the items or tasks define the overall concept of interest

Example Composite Indicator Model For A COA

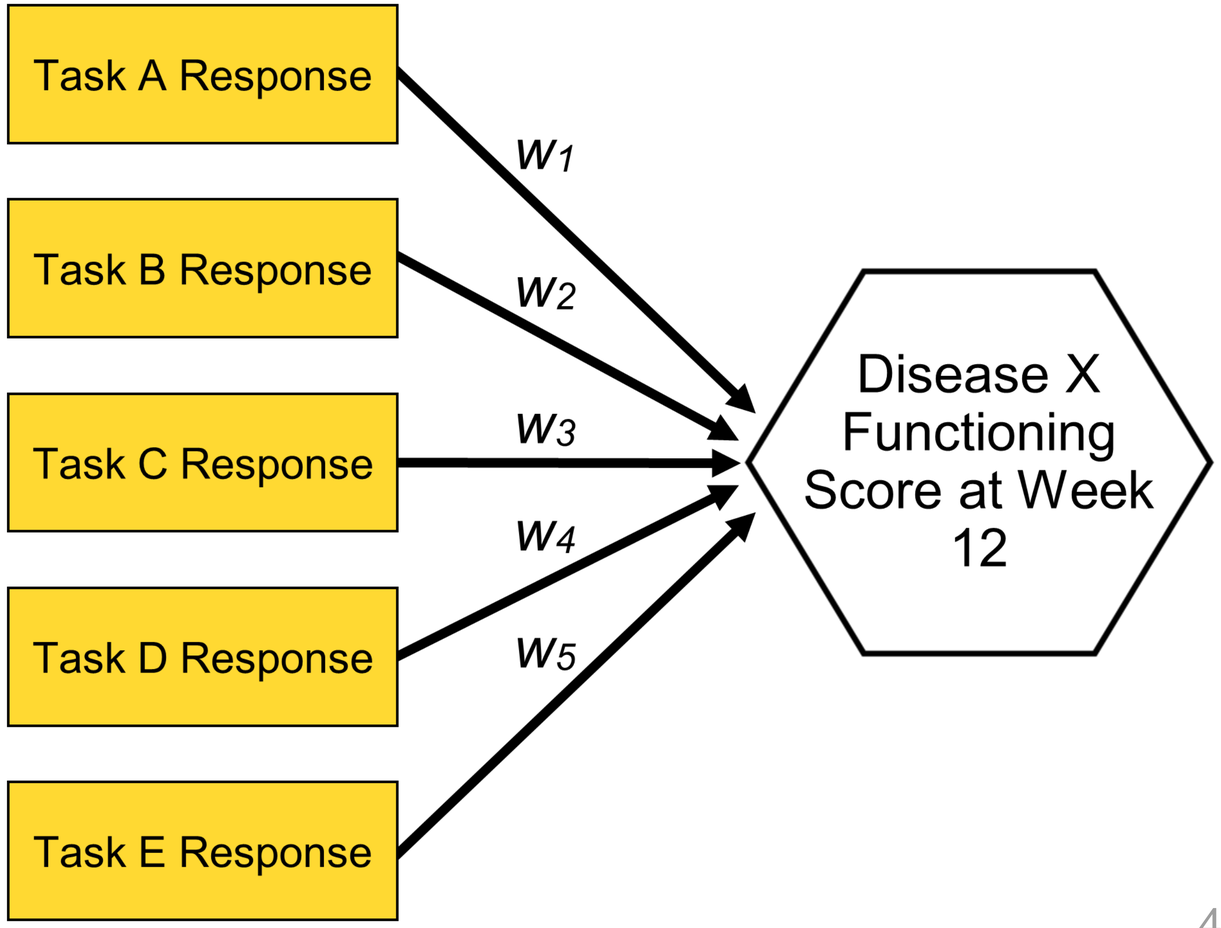


Option 3: Each component could be the response to an item/task from a single COA (based on a composite indicator measurement model)

PRO Measure (Disease X Functioning Index) Based on Composite Indicator Model



Multi-Component Endpoint Based on Scores from Same PRO Measure at Fixed Time Point

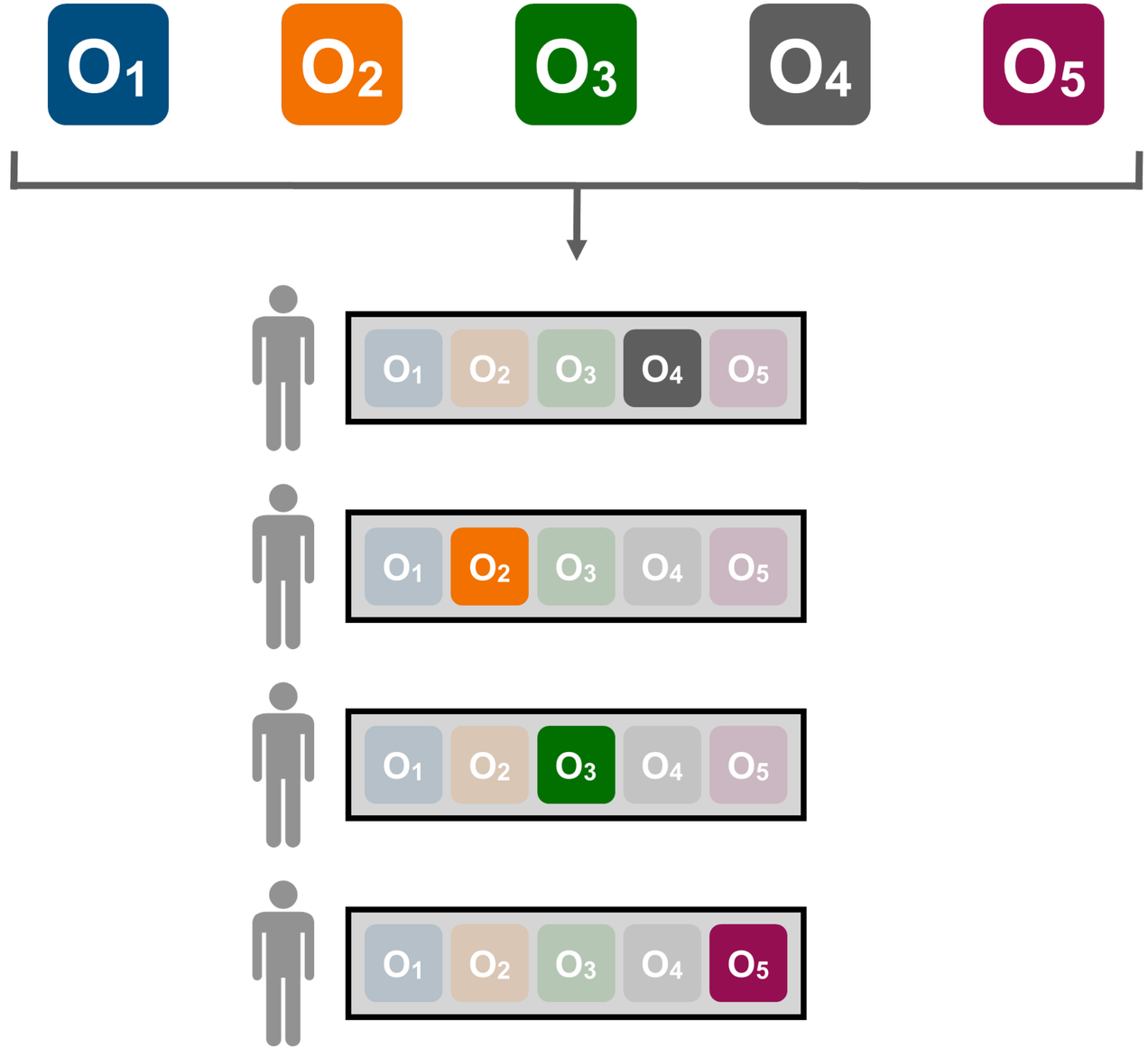


Multi-Component Endpoint

- Advantages
  - Has the potential to evaluate the entire range of important disease manifestations
  - No multiplicity adjustment needed
  - Can be efficient if the treatment effects on the different components are generally concordant
- Challenges
  - Creating a reliable and reasonable scoring algorithm can be difficult
  - Other important concerns and limitations for different types of multi-component endpoints
    - See draft Guidance 4 for detailed discussion

Heterogeneity In Diseases

Multiple outcome variables associated with a disease



Construct a personalized endpoint

Personalized Endpoint: Advantages

- Very patient focused
- Aspects of health that are not a problem for patient are not included in endpoint, meaning no dilution of treatment effect
- Could be considered along with another endpoint to inform decisions about effect of medical product
  - See FDA guidance for industry *Migraine: Developing Drugs for Acute Treatment* (February 2018) for an example of personalized endpoint for a specific context of use

Personalized Endpoint: Concerns

- For “most bothersome” or “most severe” symptoms, might be hard for patient to pick a single symptom
- Patient’s view of “most bothersome” or “most severe” symptom might change during duration of a trial
- For goal attainment scaling (GAS), patients might choose symptom/functioning that is unlikely to be affected by medical product during trial; or the chosen goal might change during the trial

Personalized Endpoint: Reminders

- Assess the same set of outcome measures for all patients regardless of their own personalized endpoint
- Standardized process for eliciting personalized endpoints, including standardized criteria to select outcome measures
- In addition to symptoms/functioning identified as most important to the individual patients, important to measure all relevant symptoms and areas of functioning to support additional analyses

Thank you!

Bringing the Patient Perspective to the Selection of Clinical Trial Outcomes

FDA, May 2023

Arthur A. Stone
Professor of Psychology, Economics, and Public Policy
Director, USC Dornsife Center for Self-Report Science
University of Southern California

Disclosures: Gallup Organization; HRA Pharma
Grant support: NIAMS AR0662200 (Stone & Schneider, MPIs)

USC Dornsife
Center for Self-Report Science

Background

- I appreciate the FDA's invitation to present today in support of the new COA Guidance
- The study I will describe was conducted prior to Guidance, and I view it as a small step that is consistent with the recommendations you will hear today
- Let me say that I fully recognize that the methods described are only one of *many* ways to approach bringing the patient perspective into outcome development
- Let me next mention the investigative team and study citations

Investigative Team and References

- Stone, A.A., Broderick, J.E., Goldman, R., Junghaenel, D.U., Bolton, A., May, M., & Schneider, S. I. Exploring indices of pain intensity derived from ecological momentary assessments: rationale and stakeholder preferences. Journal of Pain, 2021, 22, 359-370

- Schneider, S., Junghaenel, D.U., Broderick, J.E., Ono, M., May, M., & Stone, A.A. II. Indices of pain intensity derived from ecological momentary assessments and their relationships with physical, emotional, and social functioning: an individual patient data meta-analysis. Journal of Pain, 2021, 22, 371-385.

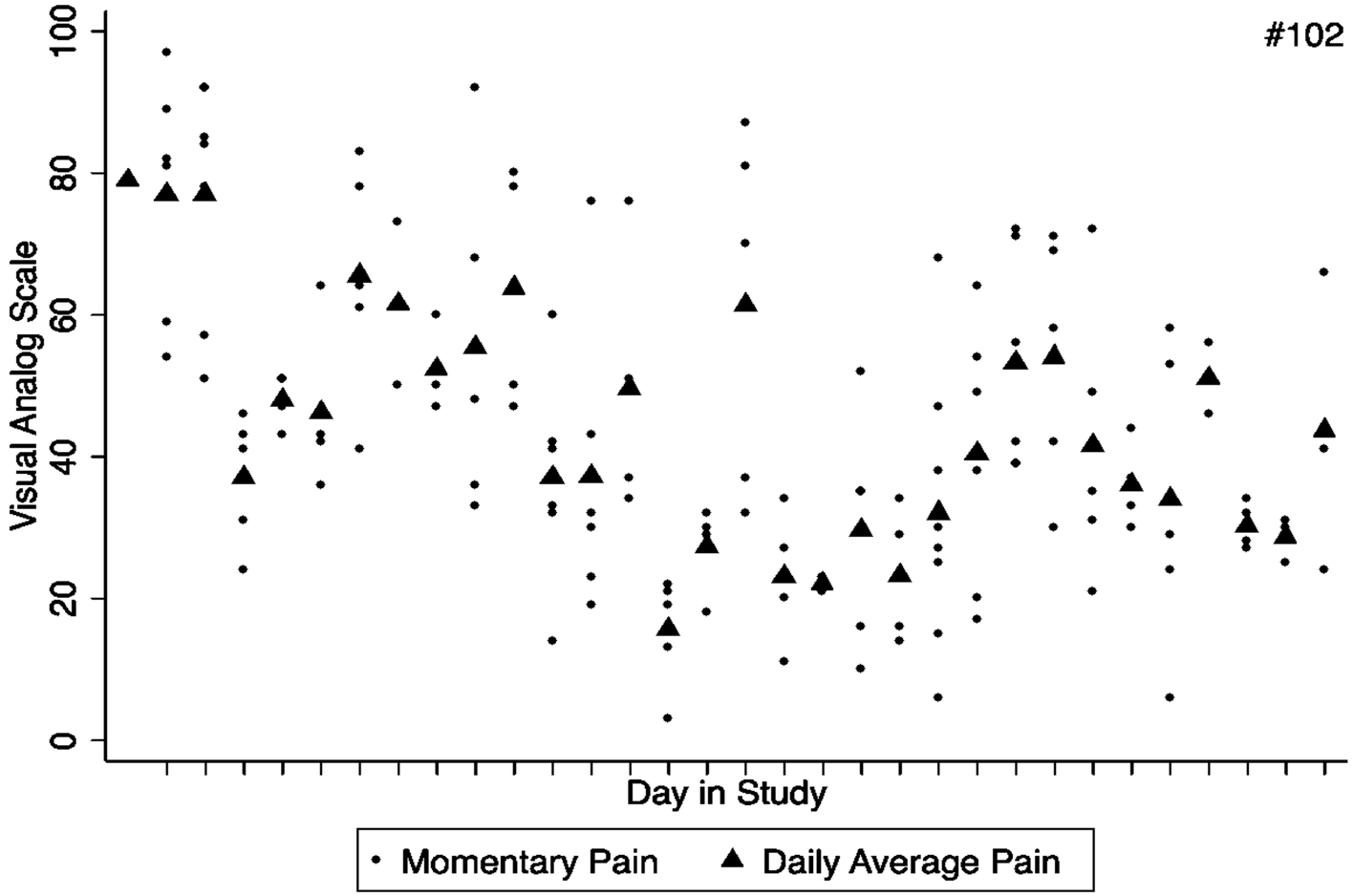
- Schneider, S., Junghaenel, D.U., Ono, M., Broderick, J.E., & Stone, A.A. III. Detecting treatment effects in clinical trials with different indices of pain intensity derived from ecological momentary assessment. Journal of Pain, 2021, 22, 386-399.

Background

- The context of the study was as a part of the development of new outcomes for assessing pain were based on momentary assessments (EMA) of pain intensity
- Current view of pain outcomes: “Average pain intensity” or “Worse pain intensity” for a designated period (say, 1-week); measurements could be diary-based or retrospective
- Our goal was to use granular, repeated measurements of pain intensity over a 1-week period to generate new ways of characterizing the pain experience with the hope of improving medical treatment
- Proposed outcomes were based, in part, on prior literature (e.g., knowing that pain variability is associated with poor functioning) and knowledge about what is missed or difficult to assess by retrospective recall (e.g., duration neglect)

Background

- Just a bit more on the rationale: We need to realize that Pain is quite variable over the day and over longer periods



Rationale for Innovative Pain Metrics from EMA

- As mentioned, we usually measure the average or worst symptom level over the reporting period
- Is this actually what makes a difference to patients?
- Does it provide us with optimal information about treatment effect?
- Alternatives to the average and worst have rarely been examined, so there is little information about their potential utility
- Objective was to explore other ways of characterizing the experience of pain over the course of a week based extracted from momentary pain reports

Potential New Outcomes

Pain Intensity Index	Definition/Explanation
Average pain intensity over a week	If we take many ratings of a patient's pain intensity during a week, add them up and then divide by the number of ratings, this would give us an average of a patient's pain during that week.
Level of pain intensity when it is at its worst during a week	If we take many ratings of a patient's pain intensity during a week, we could see what a patient's highest pain level was. This would indicate the level of pain intensity when it was at its worst.
Level of pain intensity when it is at its least during a week	If we take many ratings of a patient's pain intensity during a week, we could see what a patient's lowest pain level was. This would indicate the level of pain intensity when it was at its least.
Amount of time patient spends with no or low pain during a week	This refers to how much of the time during the week a patient didn't feel any or felt very little pain. That is, if we were to take many ratings of a patient's pain intensity, we could figure out the amount of time during a week that a patient had no pain or almost no pain.

Arm-chaired New Indices

Pain Intensity Index	Definition/Explanation
Amount of time patient spends in high pain during a week	If we were to take many ratings of a patient's pain intensity during the week, we could figure out the amount of time when a patient had ratings of pain intensity at very high levels.
How much pain intensity fluctuates or changes during a week	If we take many ratings of a patient's pain intensity during a week, we can get a sense of how much a patient's pain intensity varies from moment-to-moment or day-to-day over the week. That is, whether the intensity is more or less constant or how much a patient's pain fluctuates (that is, goes up and down).
Amount of unpredictability of pain levels during a week	This refers to the degree to which a patient's pain intensity changes for reasons that the patient can't identify. If a patient doesn't know when and why his/her pain changes, then a patient's pain levels are unpredictable.

Development of New Pain Intensity Indices brought into sharp focus the question about how trial outcomes and endpoints are chosen

How do various Stakeholders view the Importance of the Indices?

Methods

- Stakeholder Samples
  - 32 Patients with chronic pain conditions
    - Recruited from Internet Panel (SSI); Qualified as self-report Chronic Pain condition sufferer; US-wide; $30
  - 20 Healthcare Providers
    - Recruited through Am Academy of Pain Medicine; at least 8 hours care of chronic pain patient/week; MDs, PhDs, NPs. PAs; $150-200
  - 20 Clinical Trialists (Pain)
    - Recruited from NIH Research Portfolio Online Reporting Tools; $150-200
  - Understandably, Regulators were reluctant

Methods

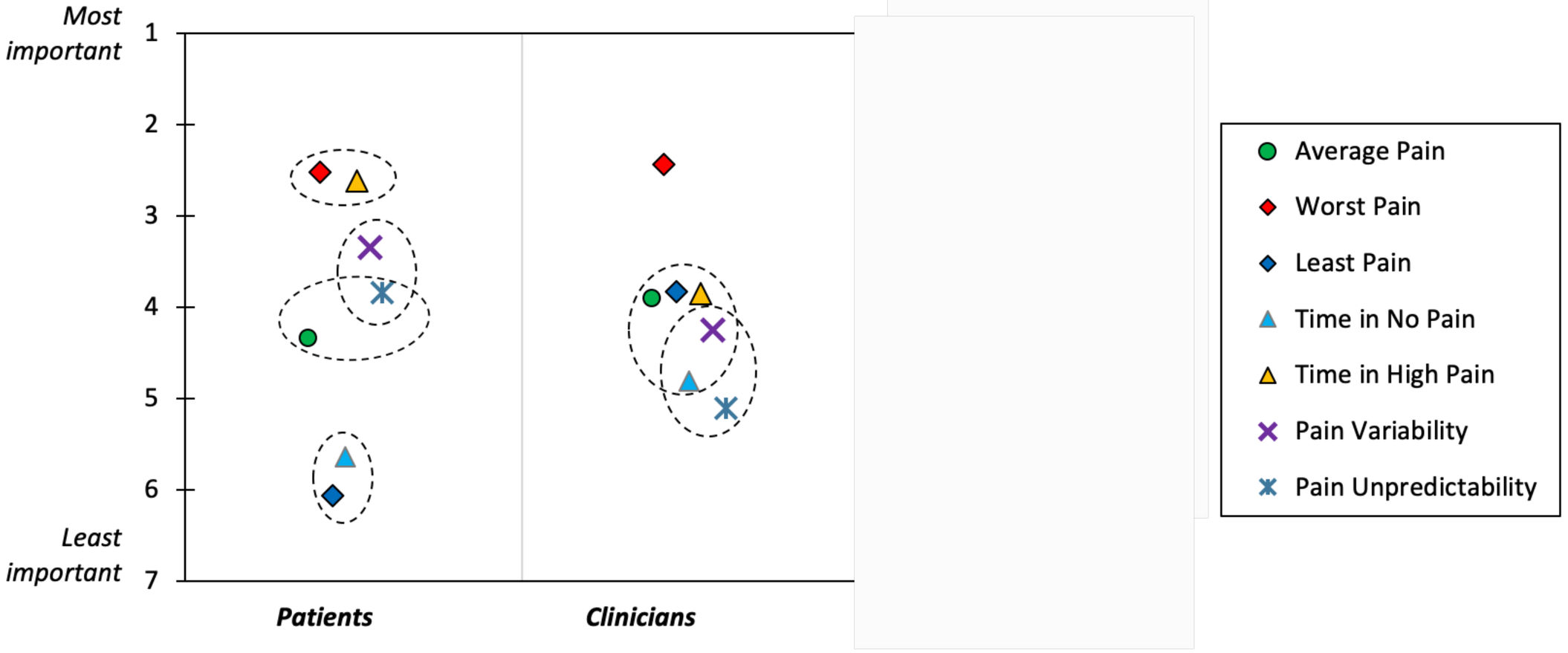
- Interviewed by telephone for 20m
- Read definitions of Pain Indices
  - Previously mailed cards with definitions
  - Probed for comprehension
- Rated “Importance” for each Index
  - Patients: “Most hoping to achieve from treatment”
  - Others: “Importance for evaluating treatment outcome”
  - Rankings: 1= Most important, 7=Least important
- Qualitative discussion of ratings
  - Stakeholders verbatim text can enhance understanding

Who were the respondents?

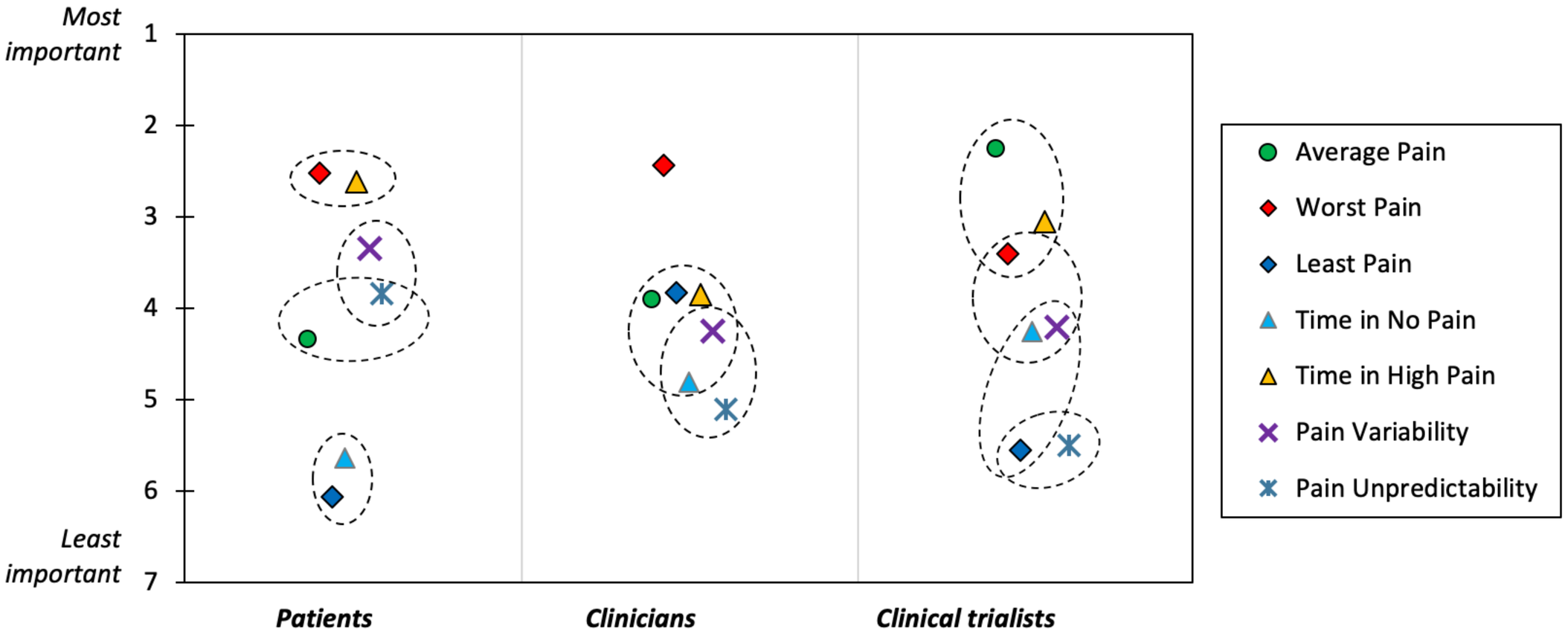
- Patient characteristics
  - 72% chronic back pain, arthritis, fibromyalgia
  - Average years since diagnosis: 15.5
  - 59% in current pain treatment
- Clinicians
  - Predominately male
  - Practicing from 1-30 years
- Trialists
  - Conducted between 1-30 trials
  - Years in clinical research: 4-40 years

Results of the Importance Sortings

Importance Ratings



Importance Ratings



Comment

- New momentary-based pain indices were understood by stakeholders
- Clear and differing opinions expressed by stakeholder groups
  - Worst pain preferred by Patients and Clinicians, whereas Average by Trialists
  - Patients thought Variability and Unpredictability were important qualities of pain intensity
  - Duration in High Pain important to Patients, but less so Clinicians

Comment

- How could this approach be improved?
  - Better representation of stakeholder populations
  - Increased sample size to allow subgroup analyses
    - E.g., by age, language comprehension, racial groups, gender
  - Preference for face-to-face, although remote interviews are scalable
- Conceptual Considerations
  - If treatments are directed at patient-chosen outcomes, will patients be better off?
  - Patient preferences are one aspect – an important one – of a multi-faceted approach to outcome(s) selection

Comment

- I believe these data show the importance of considering the opinions of multiple stakeholders, *including patients*, for selecting outcome measures.
- They highlight the importance the various ways that pain intensity can be summarized when granular data is available.
- They highlight the need for an *empirical*, *systematic,* and *transparent* approach to the development and selection of outcomes.

Analyzing COA-based Endpoints

Yuqun Abigail Luo, Ph.D.
Therapeutics Evaluation Branch 2 (TEB 2), Division of Biostatistics (DB), Office of Biostatistics and Pharmacovigilance (OBPV)
Center for Biologics Evaluation and Research (CBER), FDA

Overview

A. Analysis at a Fixed Time Point

B. Analyzing Ordinal Data

C. Missing Data

Overview

A. Analysis at a Fixed Time Point

B. Analyzing Ordinal Data

C. Missing Data

Analysis at a Fixed Time Point

- Statistical power of the treatment group comparison is generally better when the comparison is statistically adjusted for patients' baseline scores on the COA
- Also applies when the endpoint is the change in COA score from baseline to a predefined time point

See the draft guidance for industry *Adjusting for Covariates in Randomized Clinical Trials for Drugs and Biological Products* (May 2021)

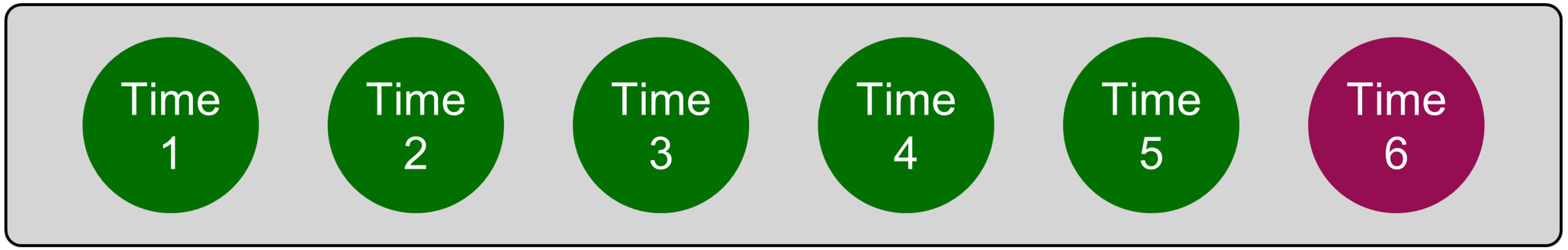
Analysis at a Fixed Time Point

If a COA-based endpoint is collected repeatedly . . .

Analysis at a Fixed Time Point

If a COA-based endpoint is collected repeatedly . . .

Data from intermediate time points can still be included in a longitudinal (e.g., mixed-effects or generalized estimating equations) model in which a treatment contrast is made for a prespecified fixed time point.



Prespecified time point for treatment contrast

Overview

A. Analysis at a Fixed Time Point

B. Analyzing Ordinal Data

C. Missing Data

Analytic Approach Might Depend Upon Type of Ordinal Endpoint

<u>Type #1</u>: Ordinal endpoint based on a COA measuring a single aspect of health

<u>Example</u>

Single item measuring musculoskeletal pain intensity

None = 0

Mild = 1

Moderate = 2

Severe = 3

Some Analytic Options

- Model mean differences (e.g., via ANCOVA)
  - *Can be challenging to interpret if steps between successive levels do not reflect equal increments in pain*
- Dichotomize (e.g., [0 or 1] vs [2 or 3])
  - *Risks ignoring important information about patients' status on concept of interest*
- Use an ordinal model (e.g., cumulative logistic regression)

Analytic Approach Might Depend Upon Type of Ordinal Endpoint

<u>Type #2</u>: Multi-component endpoint constructed by assigning ordinal values based on scores reflecting multiple aspects of health

- e.g., endpoint that combines symptom levels, hospitalization, and death

- Treatment could be beneficial for one aspect of health (e.g., symptoms) but harmful in terms of another (e.g., mortality)
  - Typical analyses might suggest overall treatment benefit and obscure harmful effects of treatment
- Sponsors should consult with FDA when developing analytic plans for such endpoints

Overview

A. Analysis at a Fixed Time Point

B. Analyzing Ordinal Data

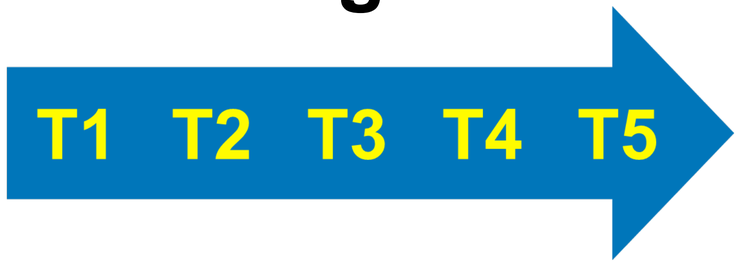
C. Missing Data

Two Types of Missingness for COA-based Endpoints

Missing Items or Tasks

Item 1
Item 2
Item 3
Item 4
Item 5
Item 6

Intermittently Missing COA

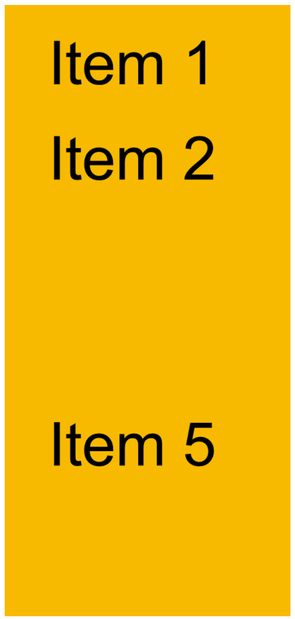


Preventing and Managing Missing Data

- Collect only the COAs needed to assess endpoint
- Make data collection plan easy and low burden for patients/caregivers
  - Counsel respondents on importance of completing the COA
  - Provide reminders when COAs are to be completed
- When COA is missing, site should be notified so that research staff can address
- Collect reasons for missingness

Two Types of Missingness for COA-based Endpoints

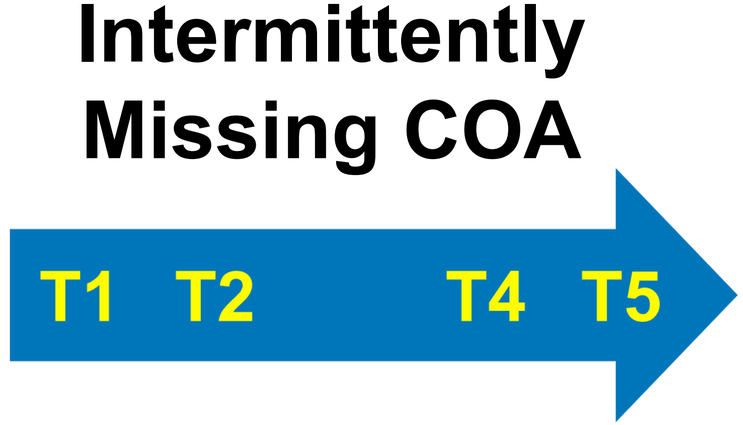
Missing Items or Tasks



Handle based on the scoring algorithm for the instrument

Two Types of Missingness for COA-based Endpoints

Sponsors should propose statistical methods that properly account for missing data with respect to a particular estimand



Patient-Focused Drug Development (PFDD) Guidance 4

Incorporating Clinical Outcome Assessments into Endpoints for Regulatory Decision-Making

May 4, 2023, Webinar

“Clinical Perspective”

Hylton V. Joffe, MD, MMSc
Director, Office of Cardiology, Hematology, Endocrinology, and Nephrology
Office of New Drugs
Center for Drug Evaluation and Research

My Perspective

- Office director and prior division director within FDA's Office of New Drugs
  - Signatory authority for certain drugs and biologics
- I've seen use of clinical outcome assessments (COAs) across a variety of diseases
  - Cardiovascular
  - Non-malignant hematology
  - Endocrinology
  - Rare diseases
  - Urology
  - Gynecology
- I'll share some common issues I've seen with COA measures (covered in PFDD guidance 3) and COA-based endpoints in trials for establishing drug effectiveness

FDA's PFDD Guidances

- I won't be able to cover all the clinical aspects of COAs as trial endpoints in this 10-minute talk

- I recommend reviewing the four PFDD guidances, which incorporate principles from FDA's experience with a wide range of drug development programs

Suboptimal Instruments

- A suboptimal instrument usually adds “noise,” biasing results towards the null
  - Harder to show a drug effect when there is one
  - Can underestimate the drug’s effect or lead to uncertainty about what the instrument is actually assessing
  - Can lead to questions about clinical meaningfulness or impact the benefit/risk assessment

Suboptimal Instruments: Examples

- Vague, confusing, or ambiguous questions
- Distal concepts that could be impacted by things other than the drug
- Trying to measure a broad concept (e.g., cognition) with a single item
- Concepts that are not relevant to the disease
- Concepts not sensitive to change with the drug
- Multi-barreled items – more than one question in an item
- Response options that are hard to tell apart – e.g., “slight” vs. “mild”

Recall Period

- Ensure the recall period is appropriate for patients to validly recall the requested information
- We often see recall periods that are too long
  - Typically increases “noise”
  - Patient recollection may be more heavily influenced by more recent experiences
- In certain circumstances may need even shorter recalls (e.g., disease affecting cognition)

Instrument Burden

- Avoid overly burdensome instruments
  - Can cause diary fatigue, leading to missing data
- Consider:
  - How many items are critical for the instrument
  - Frequency of instrument administration
  - Minimizing other trial burdens (e.g., instruments being used for exploratory endpoints)

Floor Effects

- Floor effect: high percentage of subjects select the least severe response for an item
  - These patients cannot improve on this item, which could obscure a drug effect
- Enroll patients with sufficient severity at baseline
  - Pay attention to inclusion/exclusion criteria, using same instrument at baseline that will be used for the endpoint

Blinding

- The trial should be blinded
  - Concern for bias when responding to instrument items while knowing treatment assignment
- Discuss with FDA if blinding is not possible (there may be situations where bias can be overcome – e.g., very large treatment effects)

And Don’t Forget…

- Patient input is critical when developing the COA, recall period, determining the effect size that is clinically meaningful
- Engage early with the FDA
- Ideally introduce COAs in earlier trials – opportunity to see how the COA performs and inform use in pivotal trial(s)
- A statistical win alone isn’t sufficient to establish benefit
  - The treatment effect must also be shown to be clinically meaningful
  - More on this topic in the next session


FDA
U.S. FOOD & DRUG
ADMINISTRATION

Evaluating the Meaningfulness of Treatment Benefit

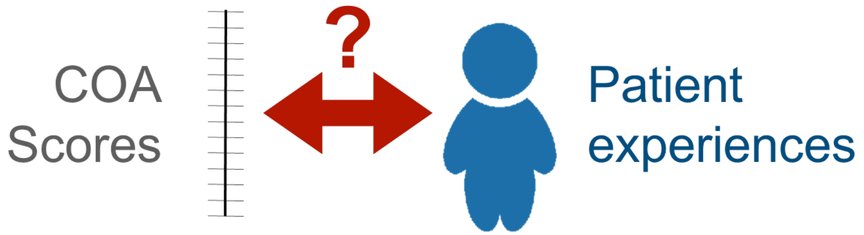
David S. Reasner, PhD

Division Director, Division of Clinical Outcome Assessment

FDA|CDER|OND|ODES

Overview

- FDA evaluates how well results of a COA-based endpoint correspond to a clinically meaningful effect of an intervention
  - Clinical benefit is a positive effect on how an individual feels, functions, or survives (BEST Glossary)
- Statistical significance does not, by itself, indicate whether the detected result corresponds to a meaningful treatment effect
- To interpret the meaningfulness of a COA-based endpoint result, we need to know how COA scores relate to patients' experiences



Easier to interpret

Clinical Meaningfulness of Scores on Which Endpoints are Based

Harder to interpret

Metric: *Number of times waking up per night*

Metric: *None, Mild, Moderate, Severe*
Patient-reported pain intensity

Metric: Severity 0-*10*
sleep disturbance

Metric: Severity 1-*12*
Item score is the product of frequency (range 1 to 4) and severity (range 1 to 3)

Metric: *Number of times waking up per night*

Metric: *None, Mild, Moderate, Severe*
Patient-reported pain intensity

Metric: Severity 0-*10*
sleep disturbance

Metric: Severity 1-*12*
Item score is the product of frequency (range 1 to 4) and severity (range 1 to 3)

Need for Empirical Evidence to Interpret Meaningfulness of Scores

May not be necessary for interpretation

Recommend additional evidence to justify

General Considerations

- Begin by reviewing existing evidence to support interpretability of COA scores used to construct the COA-based endpoint
- If existing evidence is not sufficient, conduct one or more empirical studies to support interpretability
- Empirical approaches will generate a range of plausible thresholds
- Sponsors should prespecify a range of thresholds to be used to interpret treatment effects in registration trials

Given a COA that is fit-for purpose, is the meaning (i.e., what you are trying to measure) preserved in the COA-based endpoint?

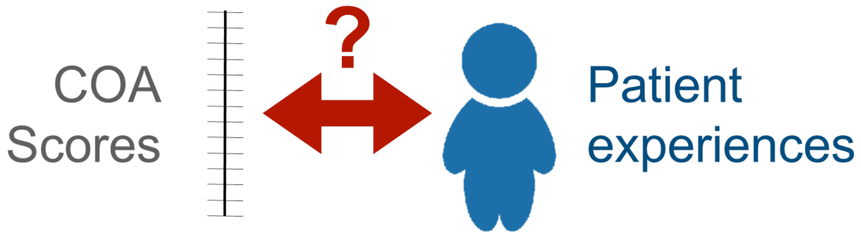
Approaches for Collecting Evidence to Support Interpretability of COA-based Endpoints

Fraser Bocell, MEd, PhD

Psychometrician, Patient Science and Engagement
Office of Strategic Partnerships and Technology Innovation
Center for Devices and Radiological Health

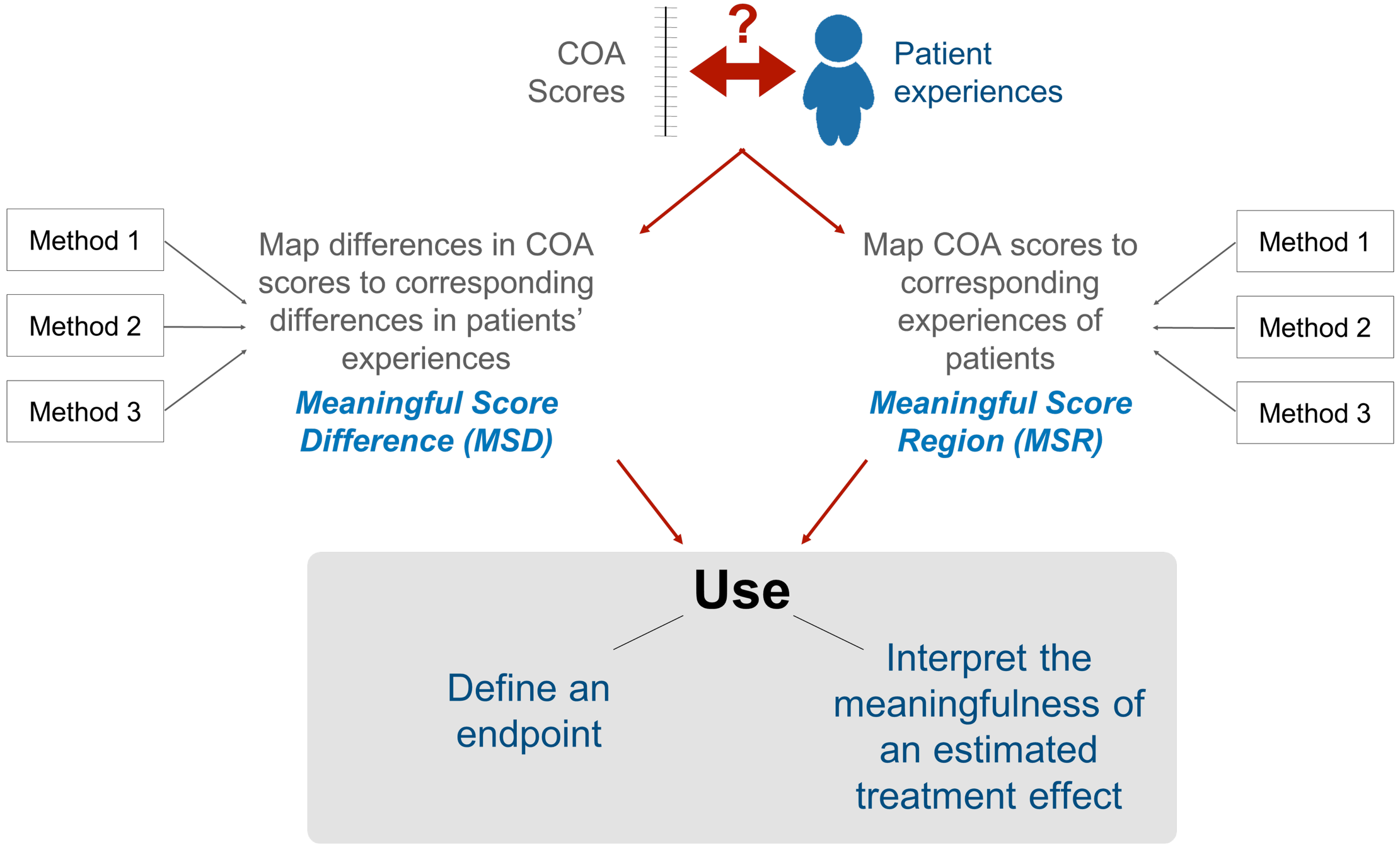
Considerations

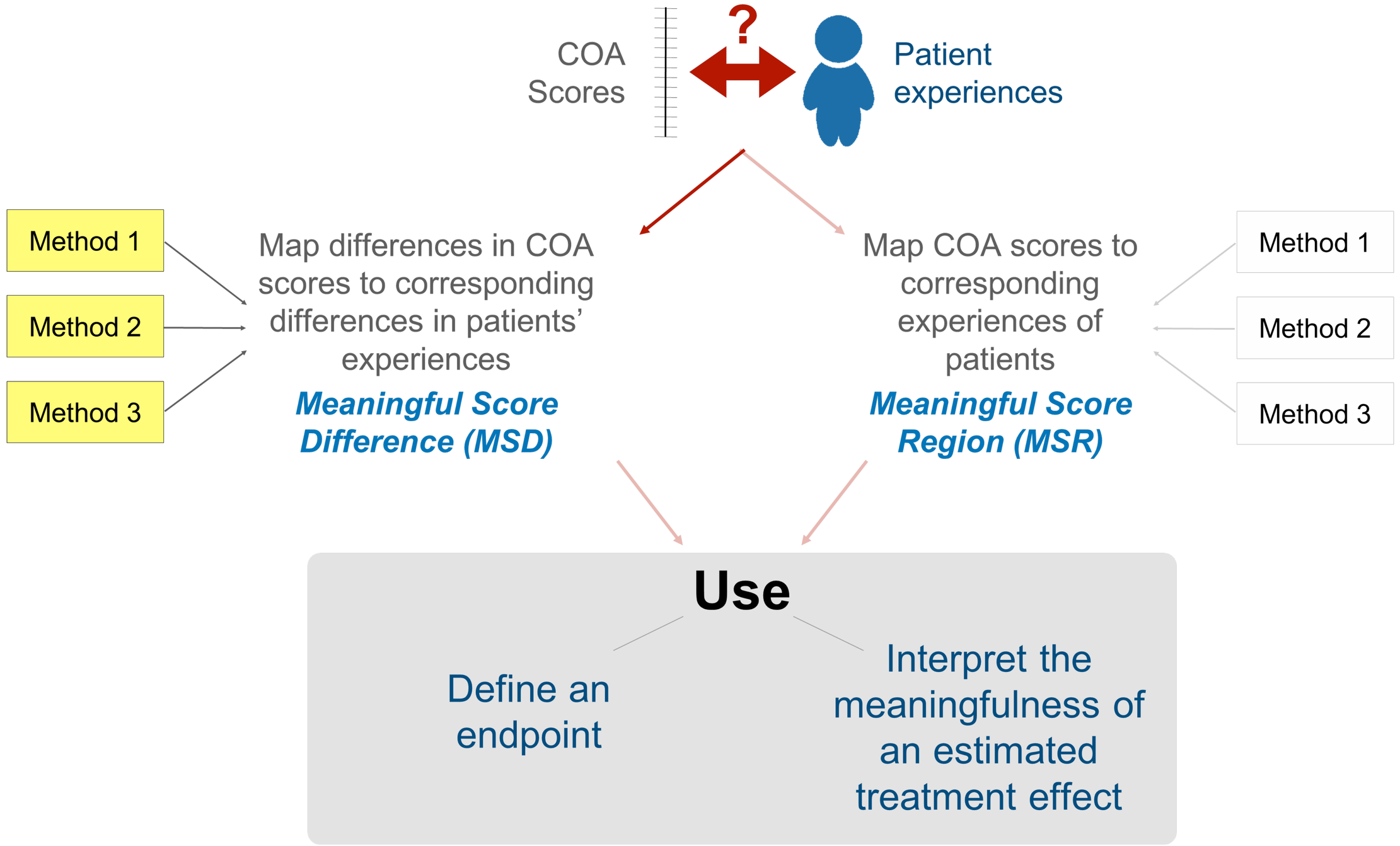
- Interpretation of COA scores can be challenging
- No single method works in all situations
  - Guidance is designed to encourage rigor and flexibility in approaches
  - A variety of methods can be used to help connect COA scores to patient experiences
- Guidance 4 includes terminology that accommodates a broader range of methods



Use

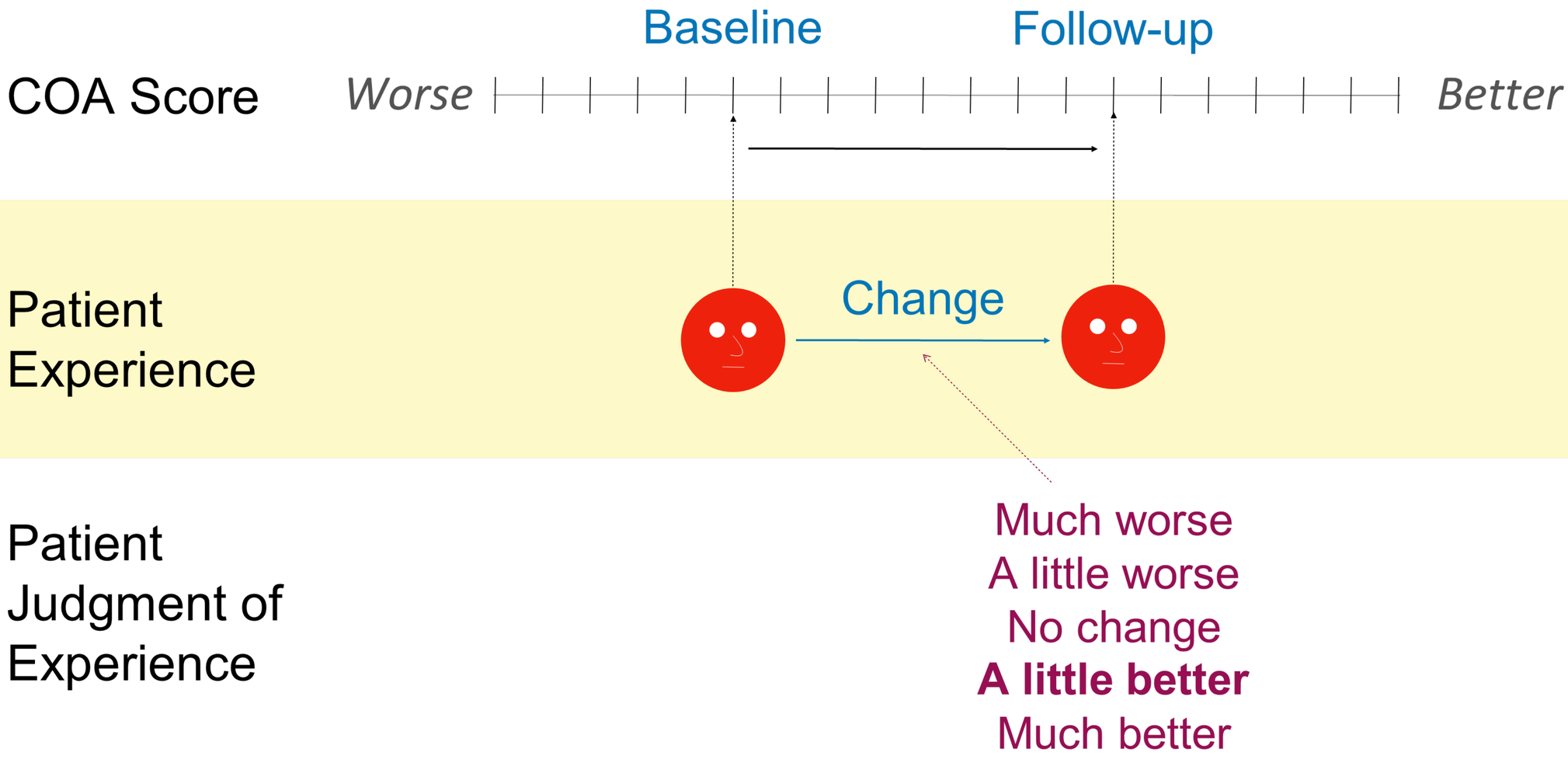
- Define an endpoint
- Interpret the meaningfulness of an estimated treatment effect


COA Scores
?
Patient experiences
Method 1
Method 2
Method 3
Map differences in COA scores to corresponding differences in patients' experiences
Meaningful Score Difference (MSD)
Map COA scores to corresponding experiences of patients
Meaningful Score Region (MSR)
Method 1
Method 2
Method 3
Use
Define an endpoint
Interpret the meaningfulness of an estimated treatment effect


COA Scores
?
Patient experiences
Method 1
Method 2
Method 3
Map differences in COA scores to corresponding differences in patients' experiences
Meaningful Score Difference (MSD)
Map COA scores to corresponding experiences of patients
Meaningful Score Region (MSR)
Method 1
Method 2
Method 3
Use
Define an endpoint
Interpret the meaningfulness of an estimated treatment effect

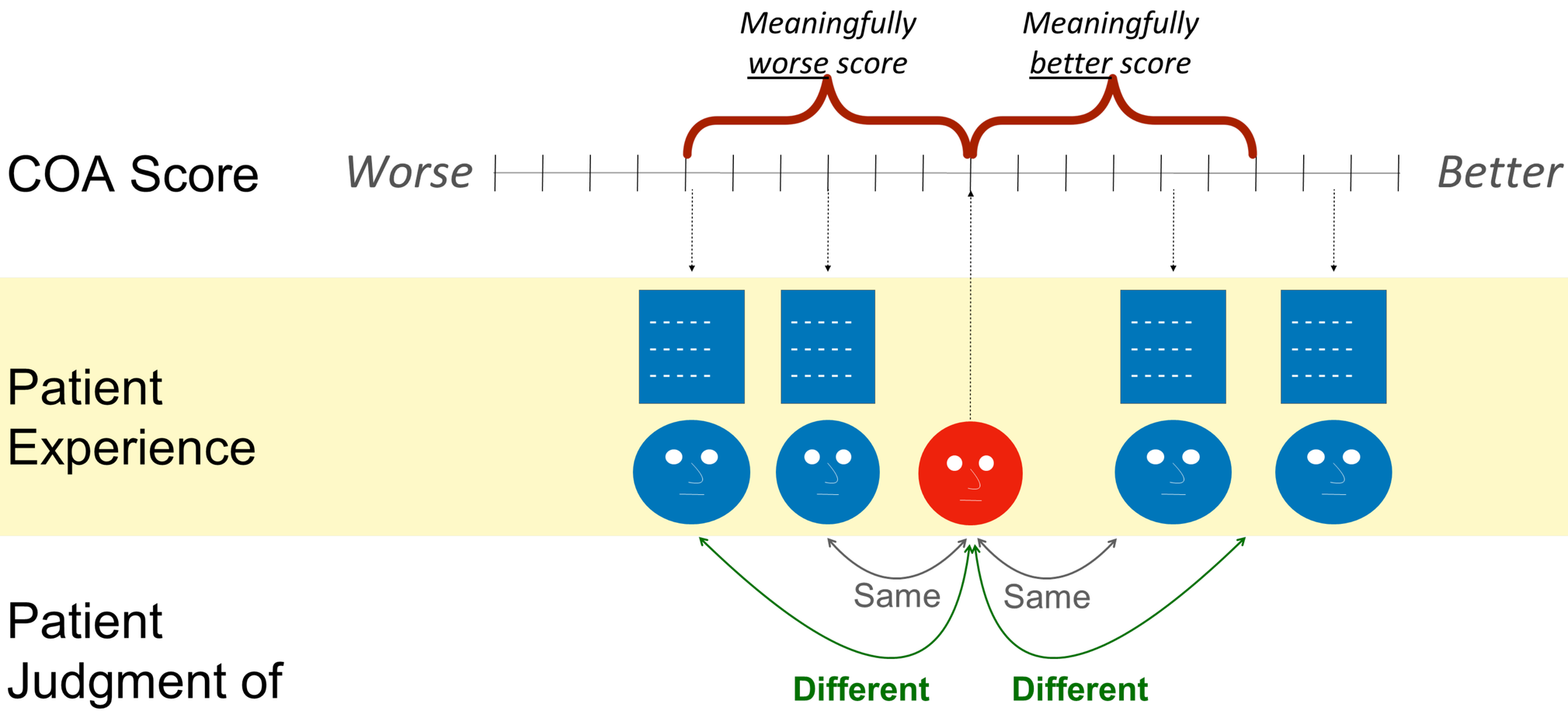
Meaningful Score Difference Approach

Anchor-based Method Using Patient Global Impression of Change

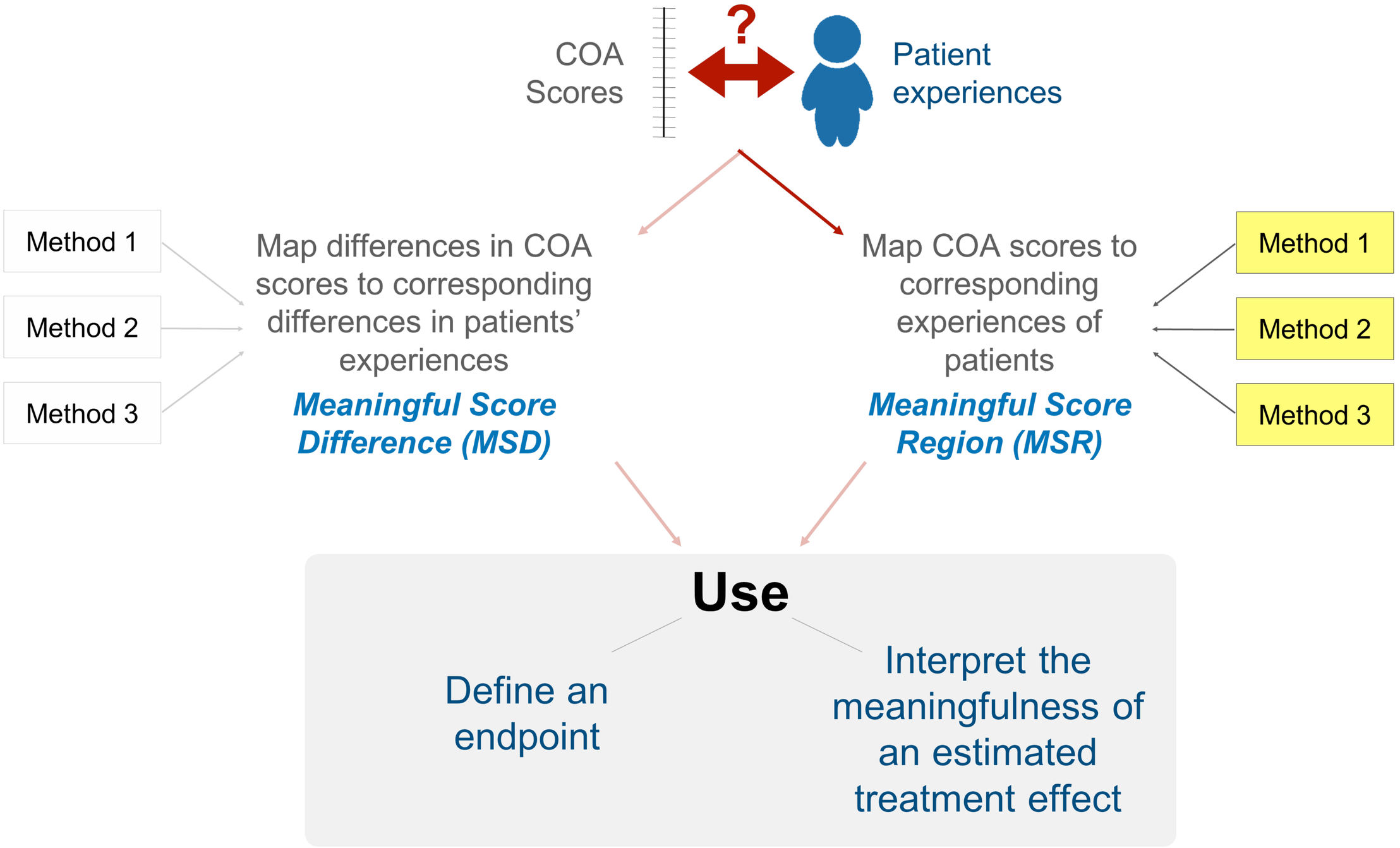


Meaningful Score Difference Approach

Idio Scale Judgment

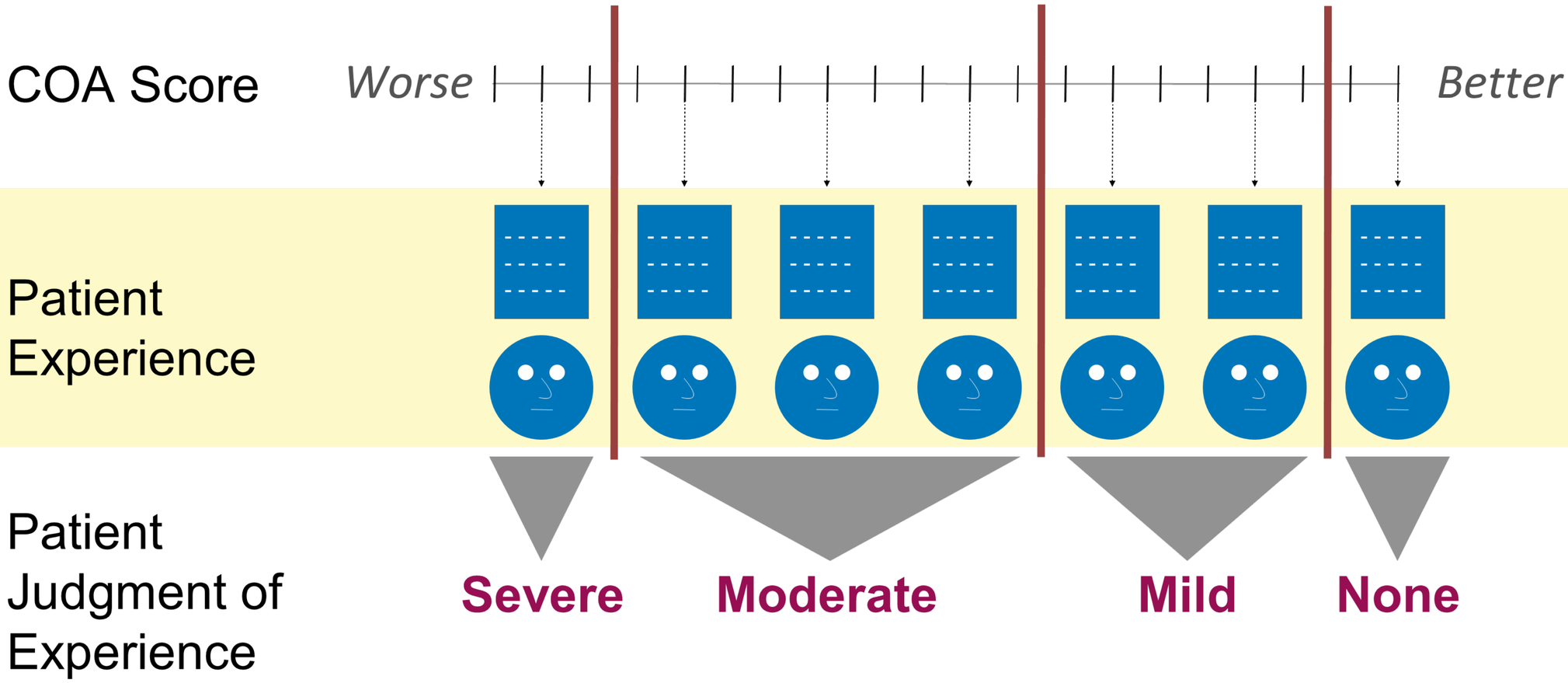


Cook et al., 2017


COA Scores
?
Patient experiences
Method 1
Method 2
Method 3
Map differences in COA scores to corresponding differences in patients' experiences
Meaningful Score Difference (MSD)
Map COA scores to corresponding experiences of patients
Meaningful Score Region (MSR)
Method 1
Method 2
Method 3
Use
Define an endpoint
Interpret the meaningfulness of an estimated treatment effect

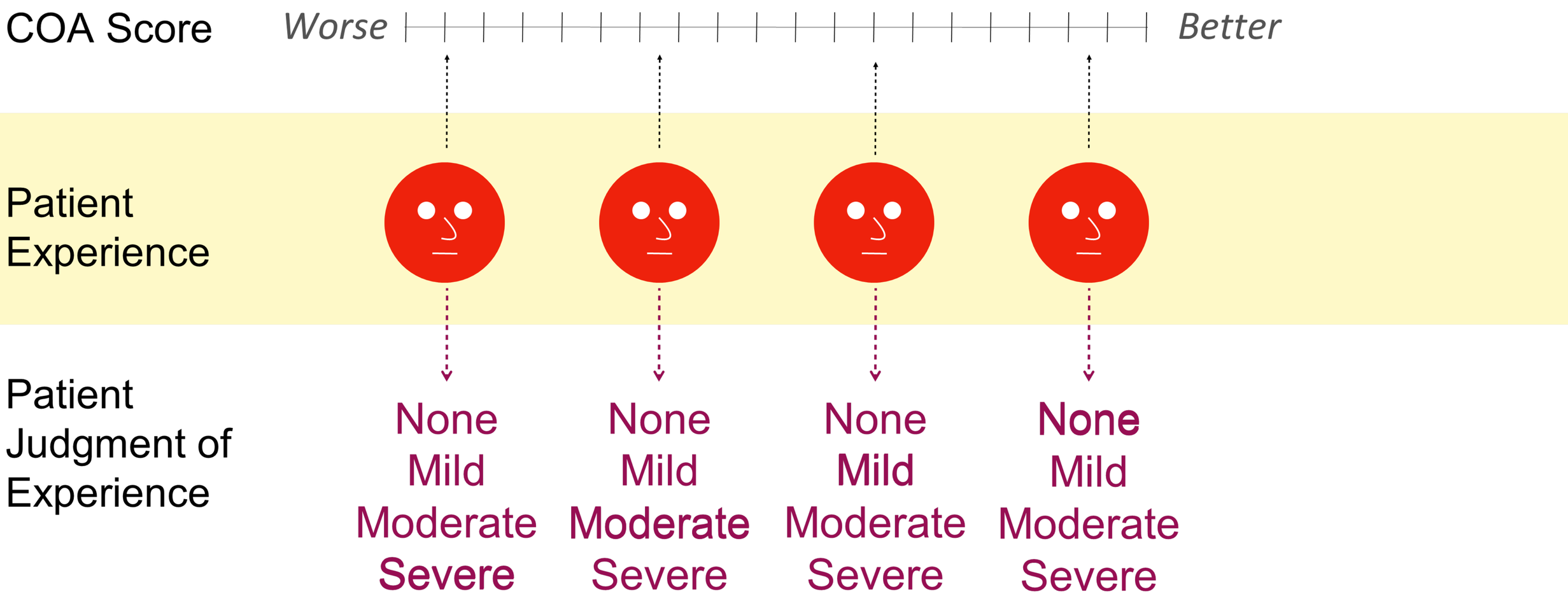
Meaningful Score Regions Approach

Bookmarking Method



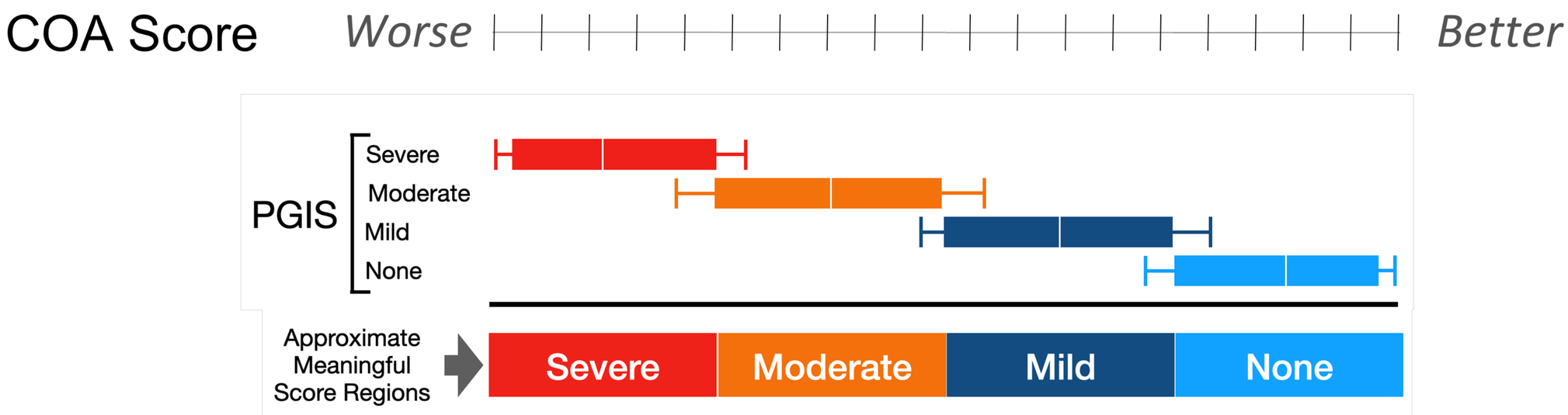
Meaningful Score Regions Approach

Anchor-based Method Using Patient Global Impression of Severity



Meaningful Score Regions Approach

Anchor-based Method Using Patient Global Impression of Severity



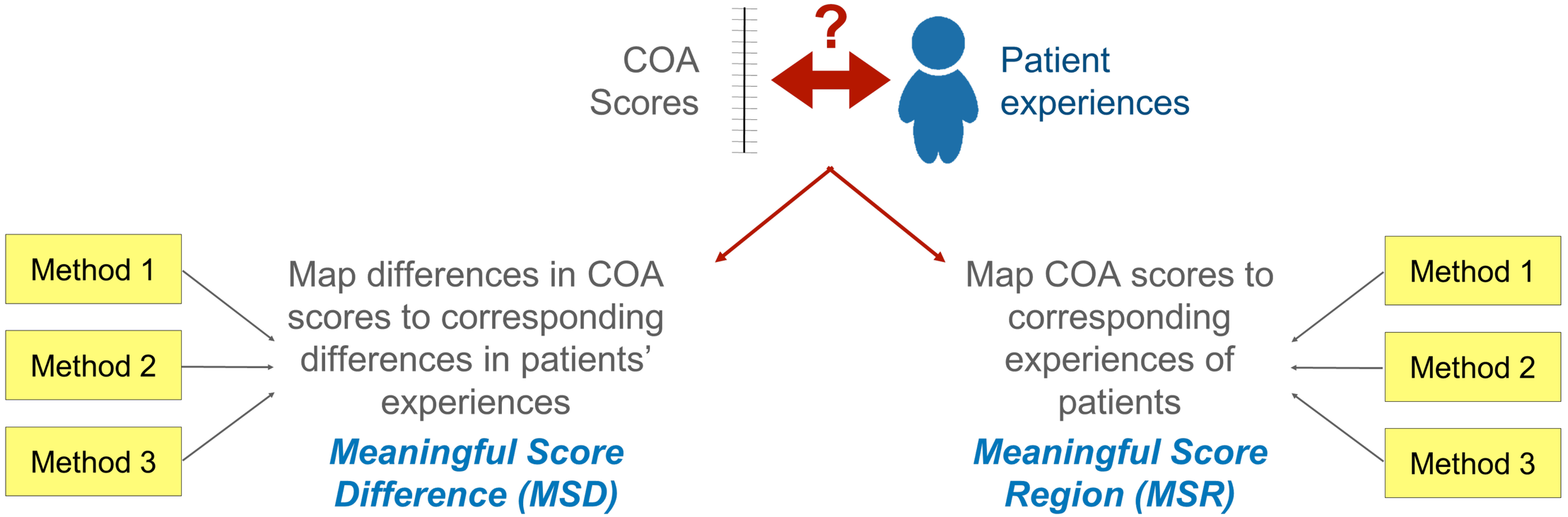
Meaningful Score Regions Approach

Other Methods

Illustrative Item*

Qualitative interviews

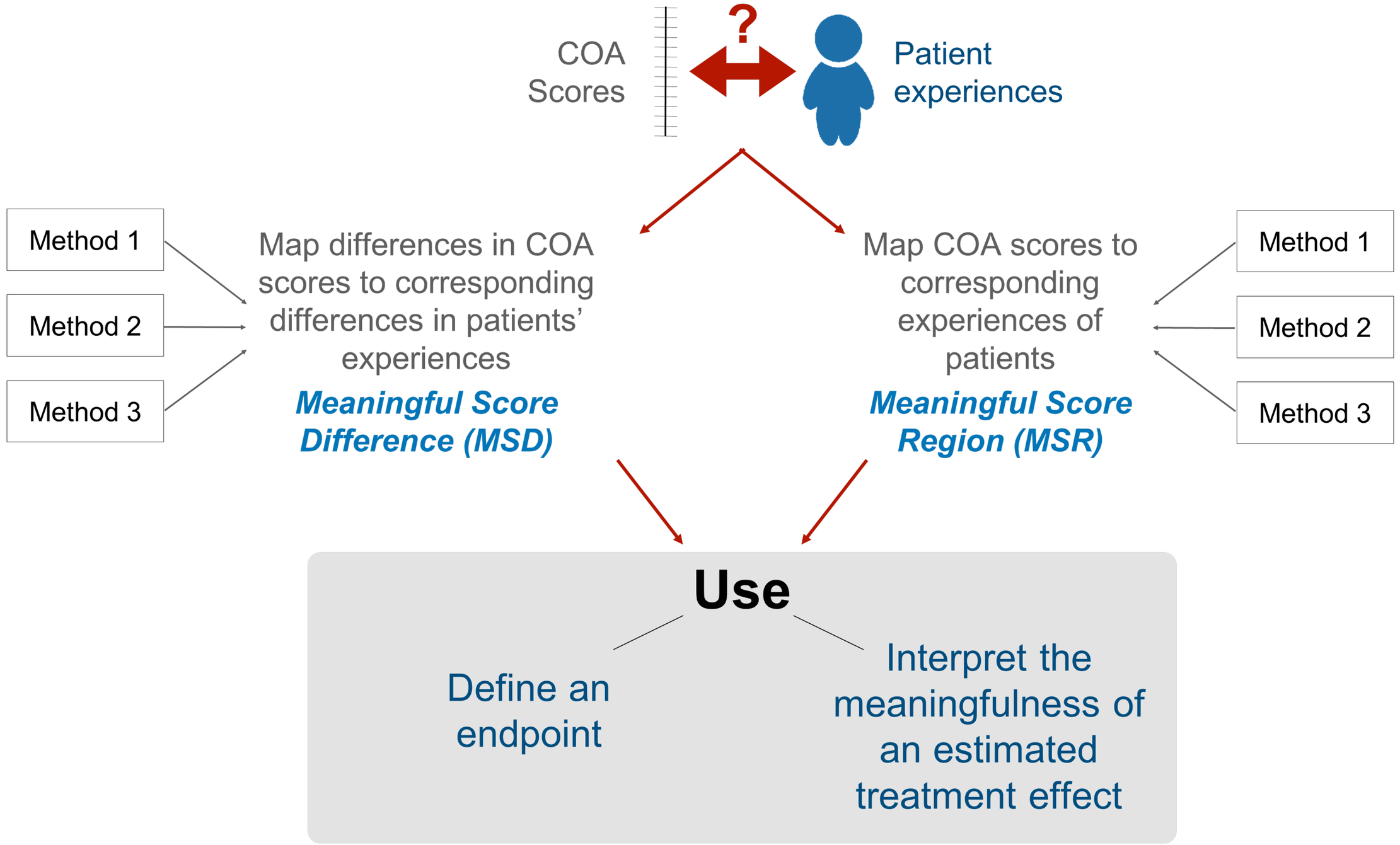
*Also known as “content-based interpretation” by Cappelleri & Bushmakin (2014)



All methods have important assumptions that should be evaluated

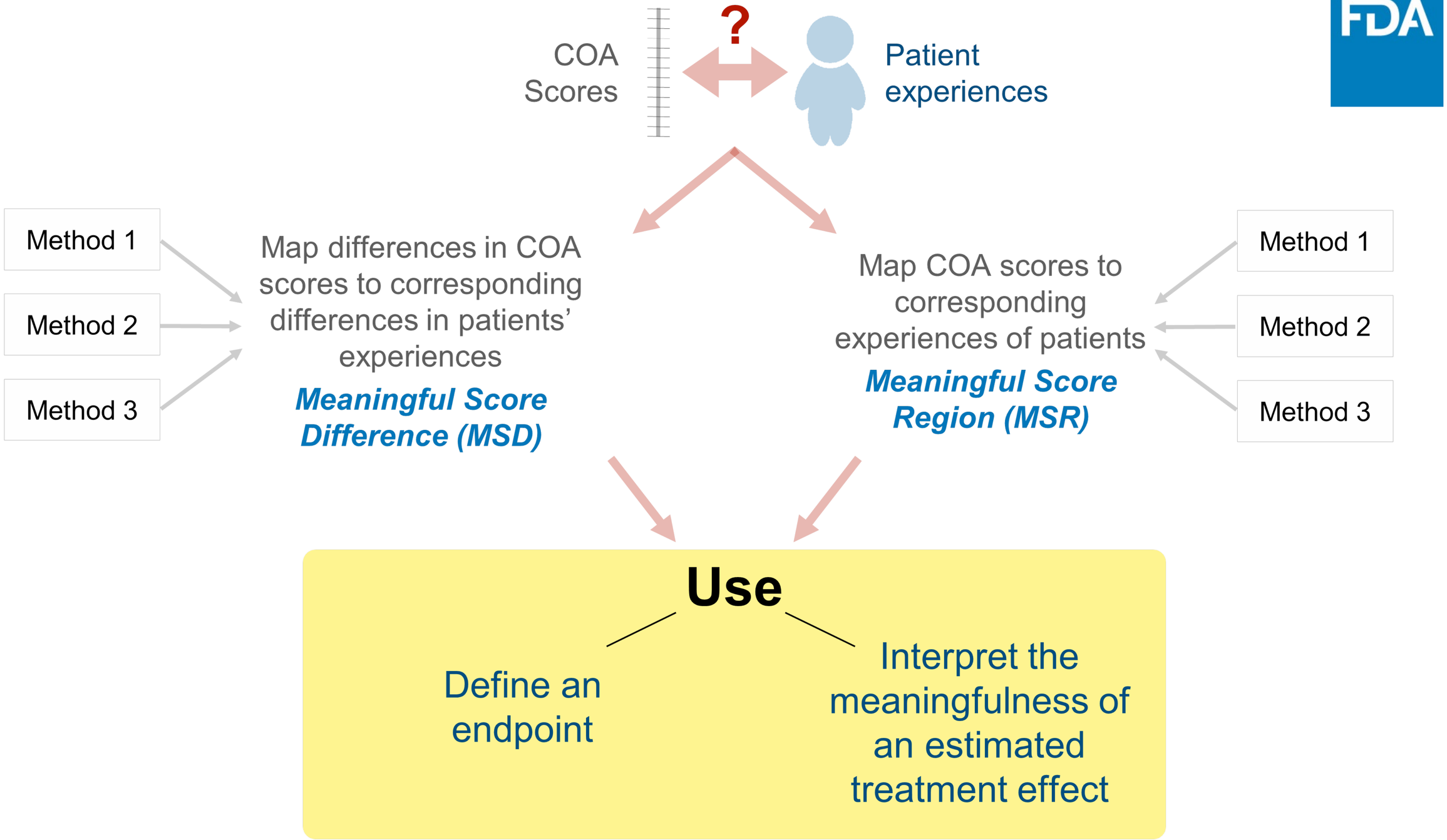
Multiple methods (including multiple anchors) are encouraged

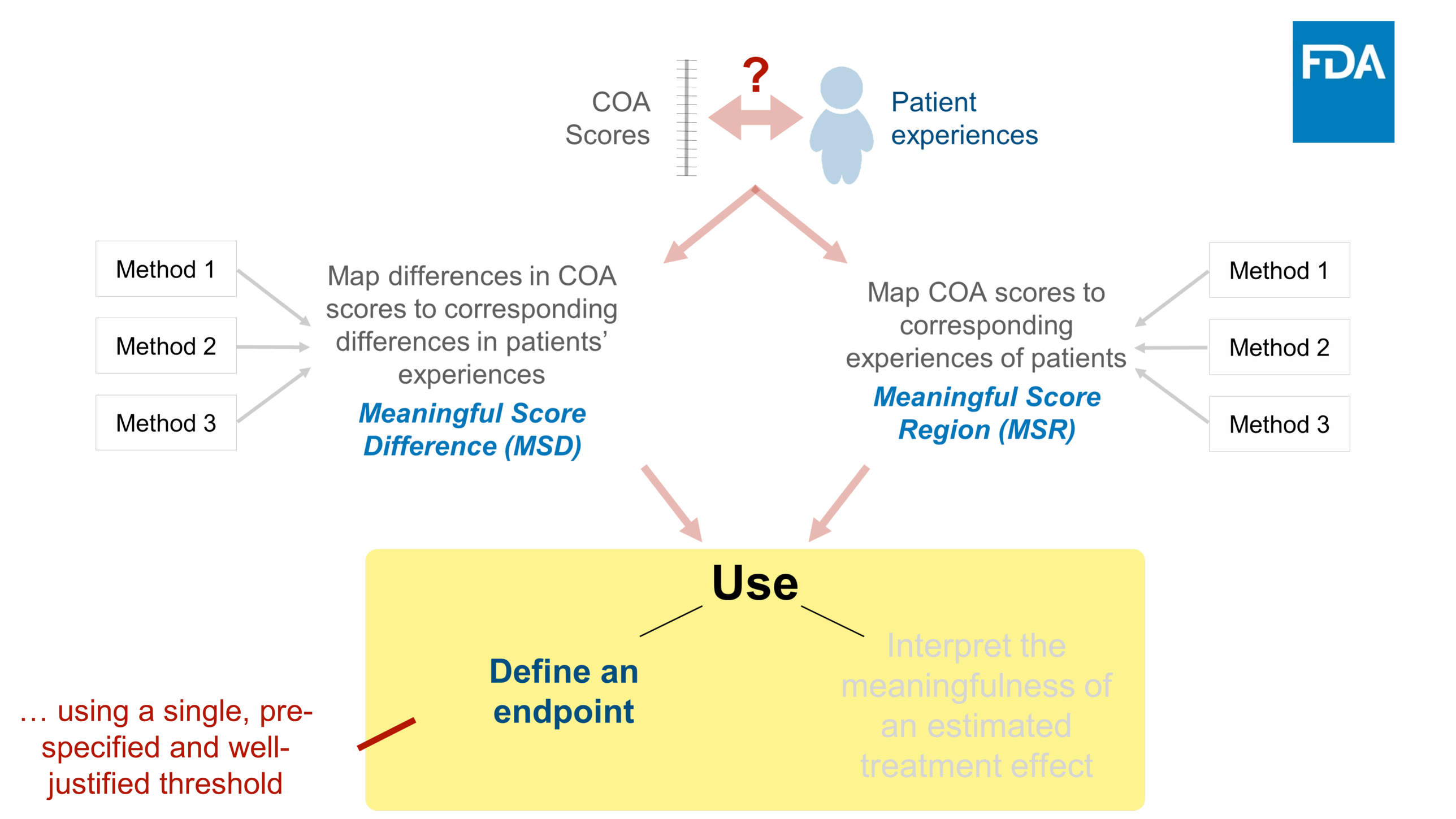
Additional research comparing methods is needed


COA Scores
?
Patient experiences
Method 1
Method 2
Method 3
Map differences in COA scores to corresponding differences in patients' experiences
Meaningful Score Difference (MSD)
Map COA scores to corresponding experiences of patients
Meaningful Score Region (MSR)
Method 1
Method 2
Method 3
Use
Define an endpoint
Interpret the meaningfulness of an estimated treatment effect

Applying Information about Meaningful Score Differences or Meaningful Score Regions to Clinical Trial Data

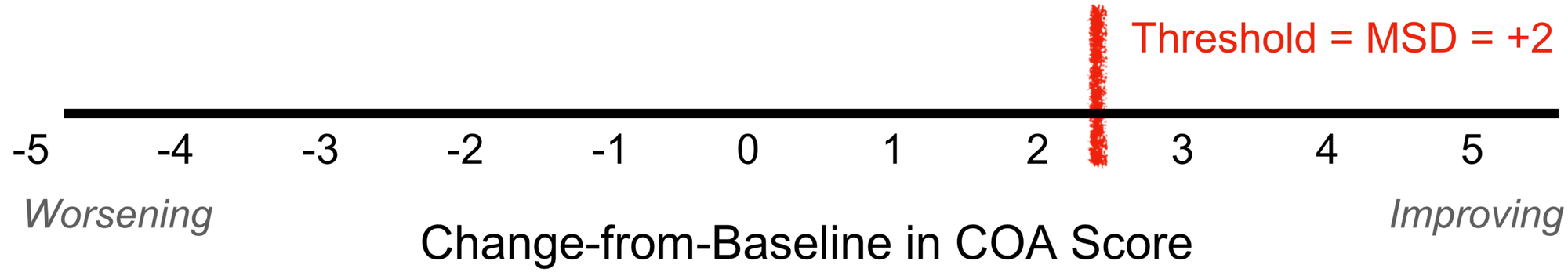
Monica Morell, PhD


COA Scores
?
Patient experiences
Method 1
Method 2
Method 3
Map differences in COA scores to corresponding differences in patients' experiences
Meaningful Score Difference (MSD)
Map COA scores to corresponding experiences of patients
Meaningful Score Region (MSR)
Method 1
Method 2
Method 3
Use
Define an endpoint
Interpret the meaningfulness of an estimated treatment effect


COA Scores
?
Patient experiences
Method 1
Method 2
Method 3
Map differences in COA scores to corresponding differences in patients' experiences
Meaningful Score Difference (MSD)
Map COA scores to corresponding experiences of patients
Meaningful Score Region (MSR)
Method 1
Method 2
Method 3
Use
Define an endpoint
Interpret the meaningfulness of an estimated treatment effect
… using a single, pre-specified and well-justified threshold

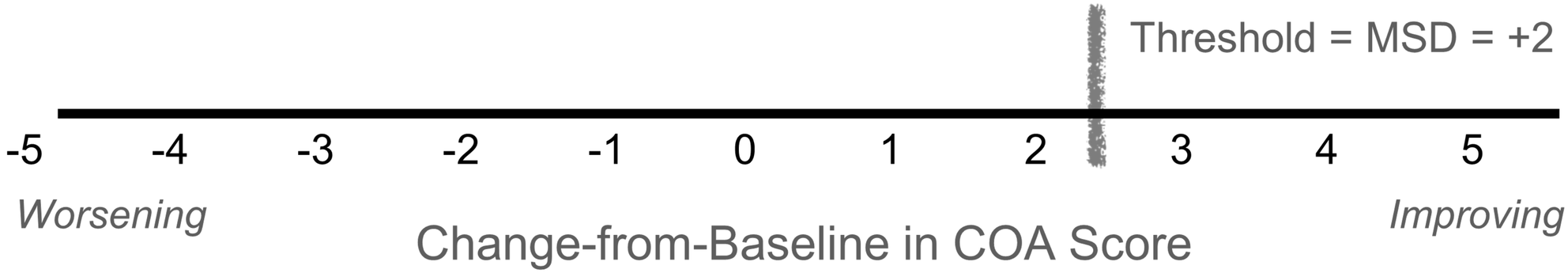
Defining a COA-based Endpoint in Terms of a Threshold

COA Threshold in terms of Meaningful Score Differences

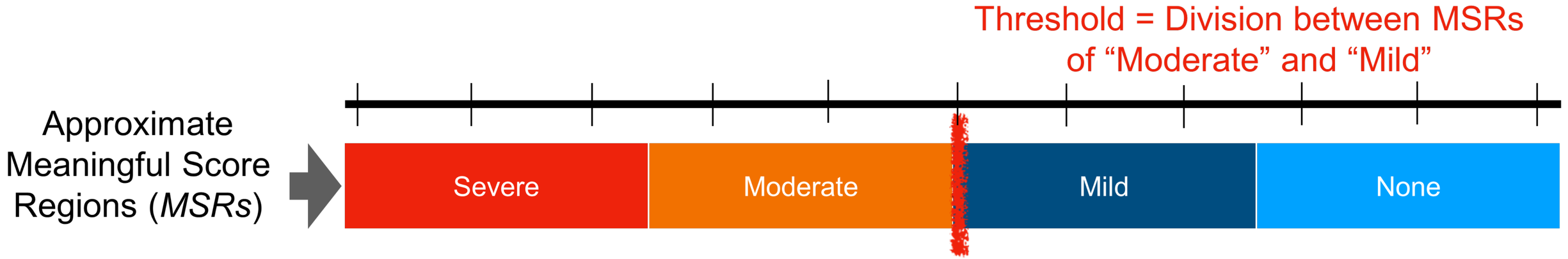


Defining a COA-based Endpoint in Terms of a Threshold

COA Threshold in terms of Meaningful Score Differences



COA Threshold in terms of Meaningful Score Regions



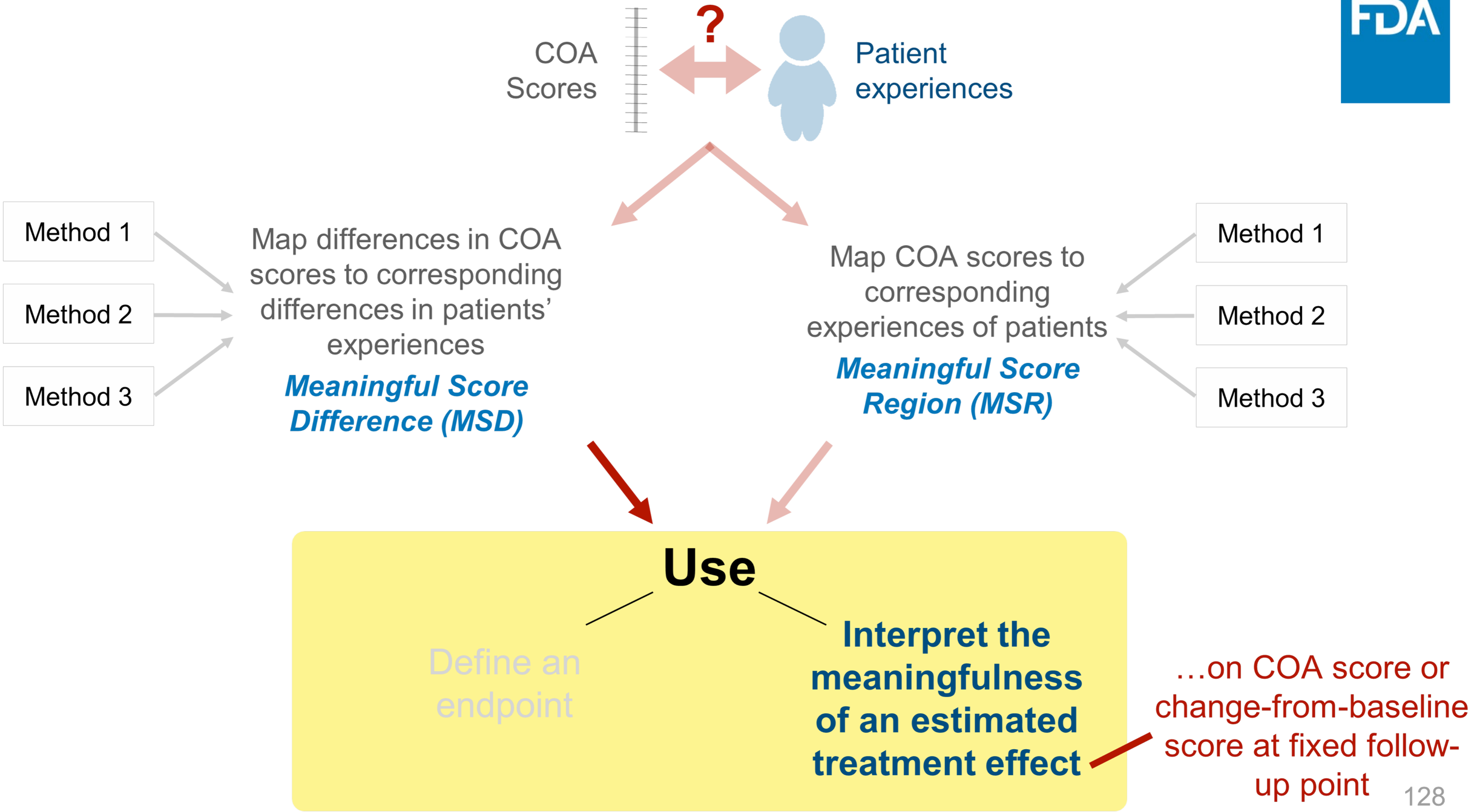
When Using MSDs or MSRs to Define an Endpoint

- Before deciding to use a COA score threshold to construct a responder endpoint, review the special considerations and concerns about responder endpoints discussed earlier
- When a COA score threshold is used to define an endpoint, a single pre-specified threshold is required

Single threshold is required because patients get to have only one value for an endpoint.

Psoriasis Investigator Global Assessment Scale

0	1	2	3	4
Clear	*Almost clear*	*Mild*	*Moderate*	*Severe*


COA Scores
?
Patient experiences
Method 1
Method 2
Method 3
Map differences in COA scores to corresponding differences in patients' experiences
Meaningful Score Difference (MSD)
Map COA scores to corresponding experiences of patients
Meaningful Score Region (MSR)
Method 1
Method 2
Method 3
Use
Define an endpoint
Interpret the meaningfulness of an estimated treatment effect
…on COA score or change-from-baseline score at fixed follow-up point

Statistically significant treatment effect on COA-based endpoint

Is the estimated treatment effect meaningful to patients?

How the treatment effect is estimated will depend on the endpoint and trial design

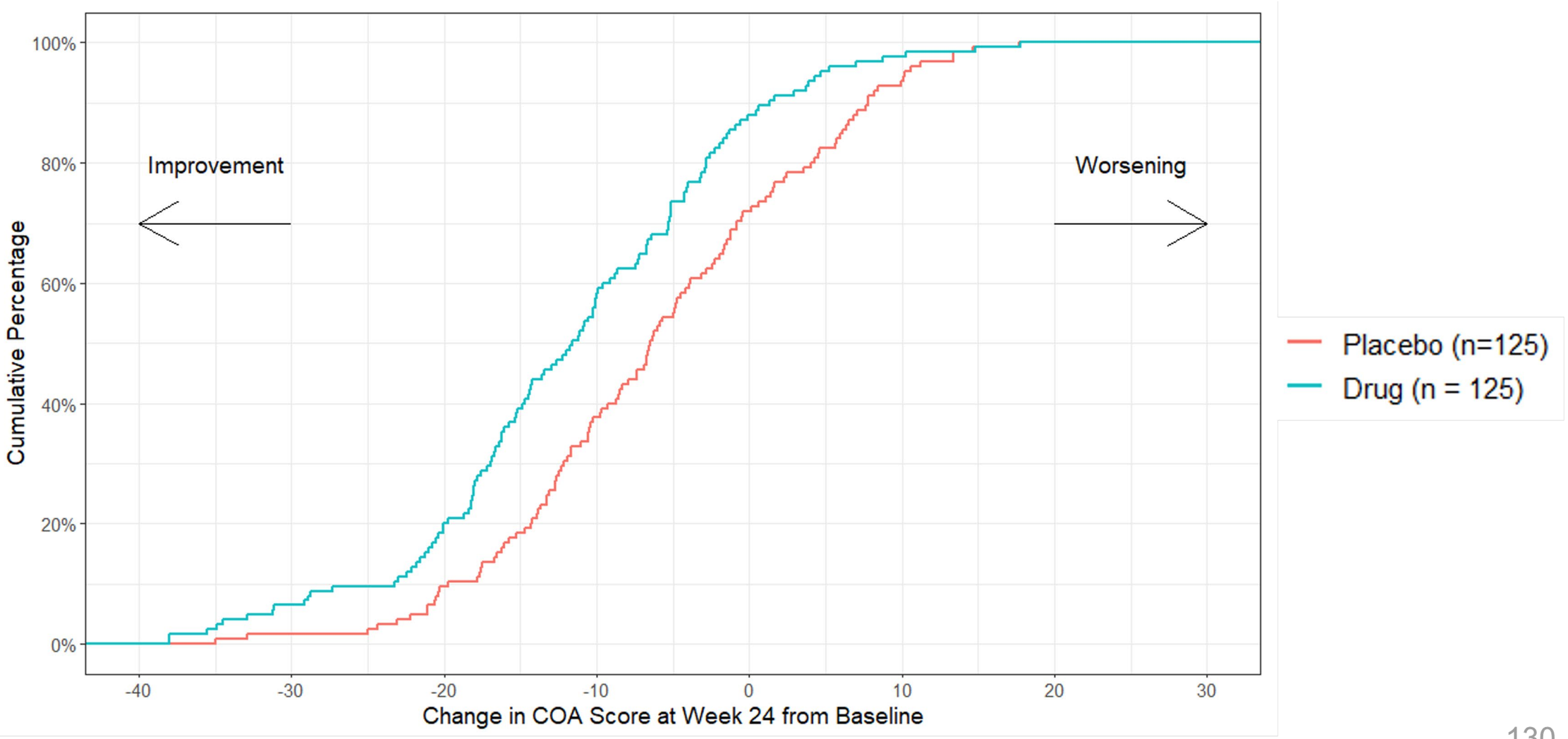
Assume endpoint based on COA score or change-from-baseline score at fixed follow-up time

Assume parallel groups design

Treatment effect is estimated by comparing study groups

How we examine the meaningfulness of the treatment effect depends on how we look at the group comparison

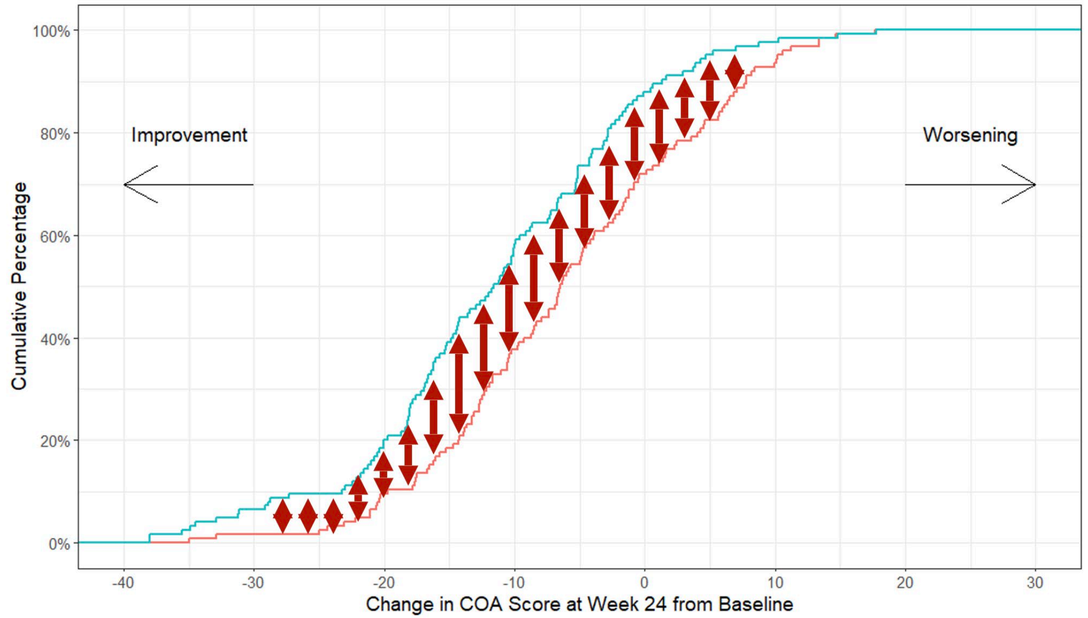
How we examine the meaningfulness of the treatment effect depends on how we look at the group comparison



How we examine the meaningfulness of the treatment effect depends on how we look at the group comparison

Expected Difference in the Probability of Exceeding One or More Score Thresholds:

The Vertical Gap Between Groups

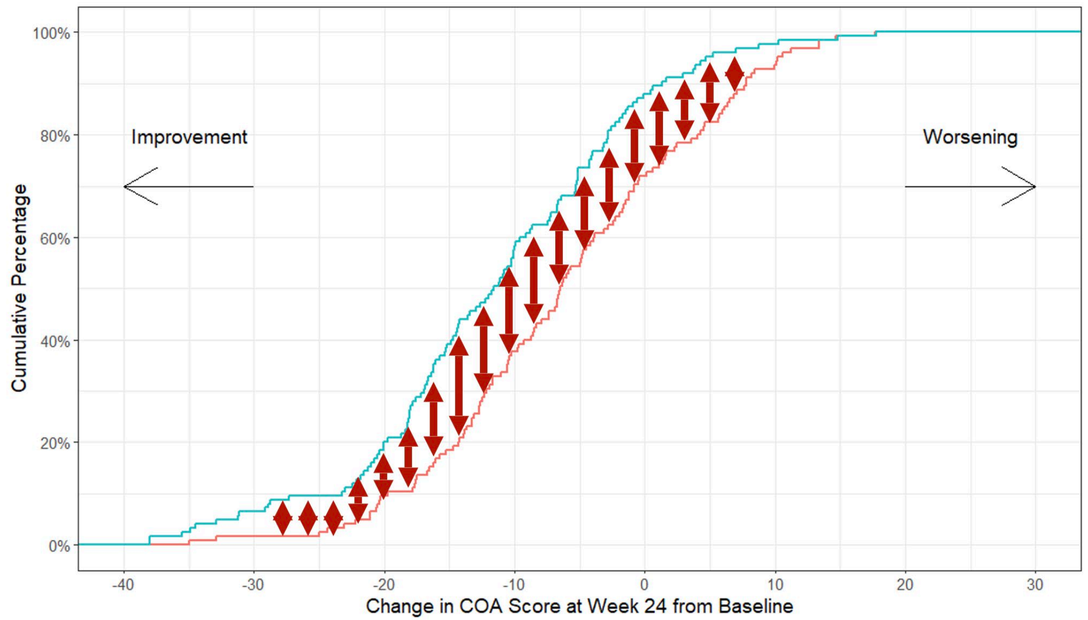


How much more likely is the average patient to experience a meaningful improvement in how they feel if given Drug rather than Placebo?

How we examine the meaningfulness of the treatment effect depends on how we look at the group comparison

Expected Difference in the Probability of Exceeding One or More Score Thresholds:

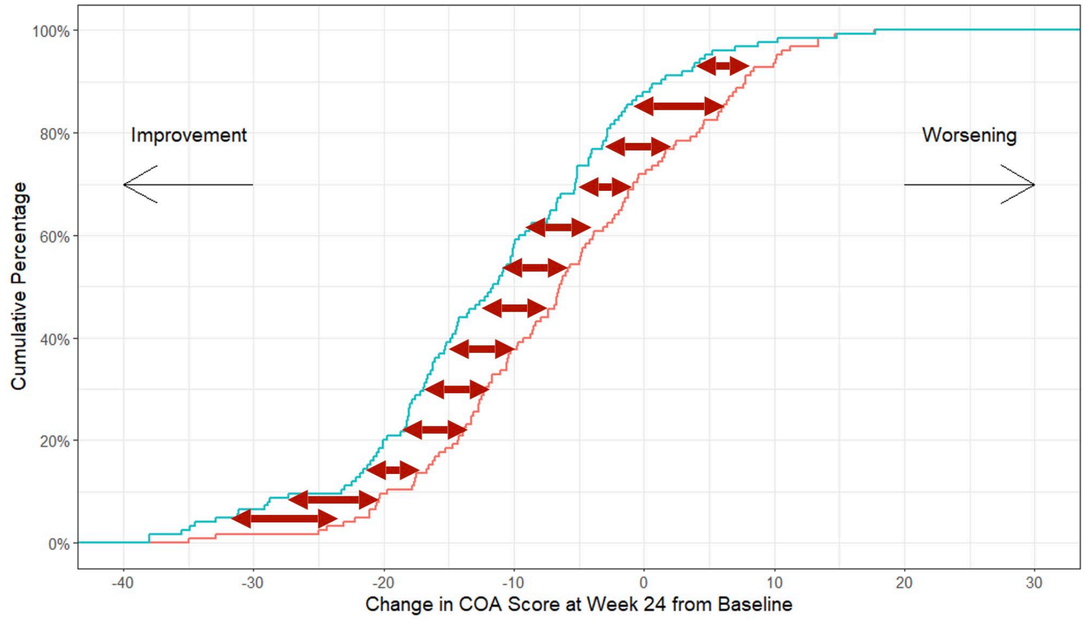
The Vertical Gap Between Groups



How much more likely is the average patient to experience a meaningful improvement in how they feel if given Drug rather than Placebo?

Expected Difference in Scores (or in Change-From-Baseline Scores):

The Horizontal Gap Between Groups

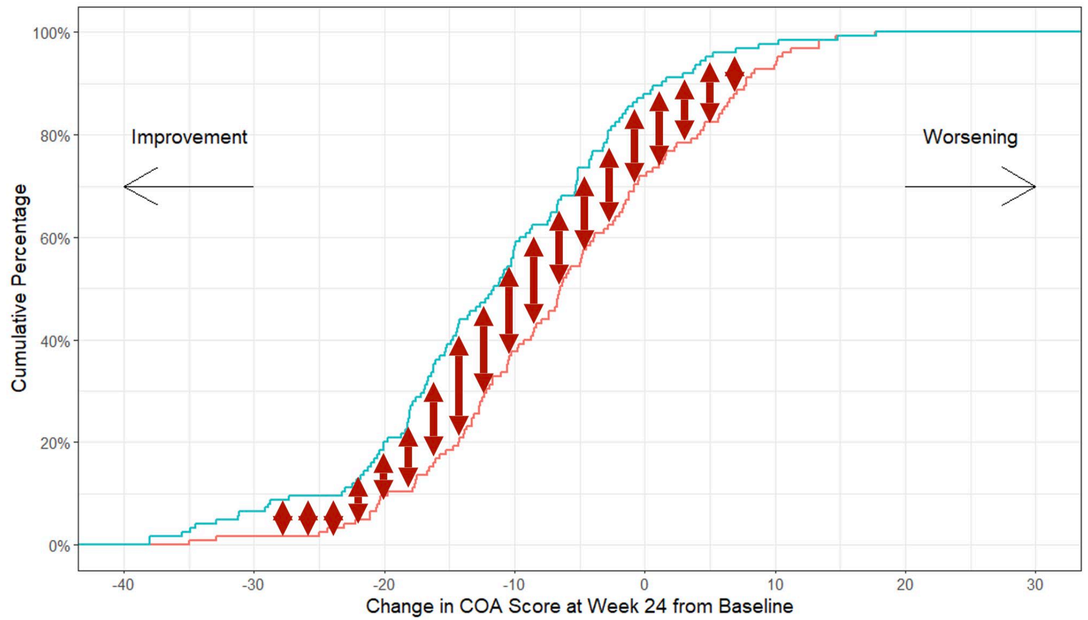


How much better is the average patient likely to feel if they receive Drug rather than Placebo?

How we examine the meaningfulness of the treatment effect depends on how we look at the group comparison

Expected Difference in the Probability of Exceeding One or More Score Thresholds:

The Vertical Gap Between Groups



How much more likely is the average patient to experience a meaningful improvement in how they feel if given Drug rather than Placebo?

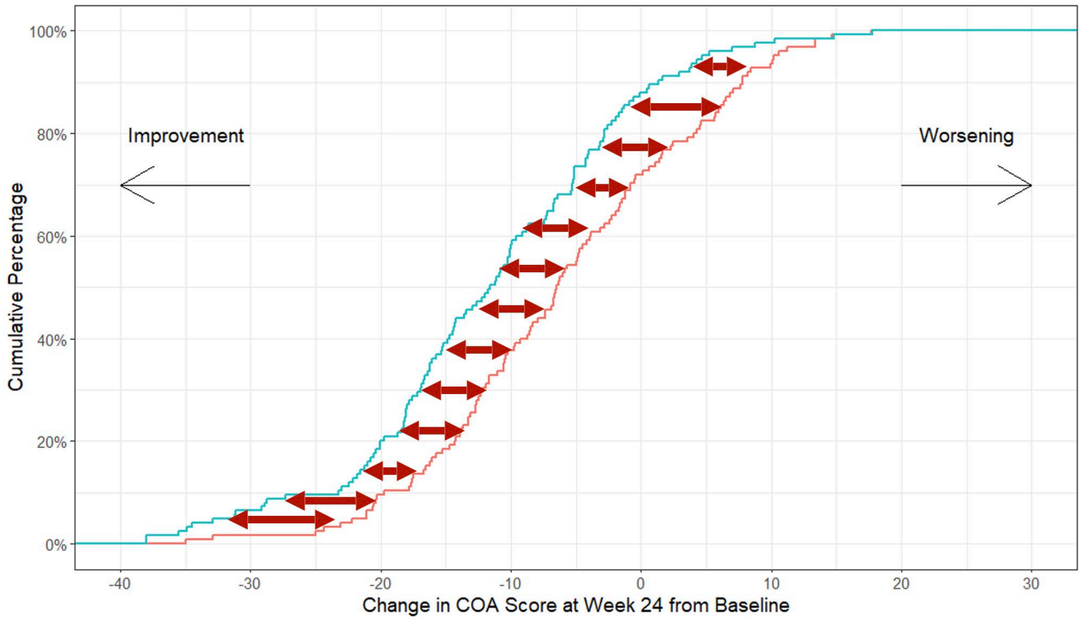
- Differences across treatment groups in the likelihood of exceeding some score threshold(s) are examined
- An examination of the vertical gap between two group's empirical cumulative distribution function (eCDF) curves
- Estimated MSDs or MSRs inform threshold(s)
- A determination must be made about whether the magnitude of difference is clinically meaningful to patients

How we examine the meaningfulness of the treatment effect depends on how we look at the group comparison

- Differences in expected scores (or change-from-baseline scores) between groups are examined
- The average horizontal gap is equal to the difference between means of the two score distributions[1]
  - Average horizontal gap corresponds to the overall treatment effect
- Range of MSDs or MSRs viewed in conjunction with treatment effect to aid in determination of whether estimated treatment effect is meaningful to patients

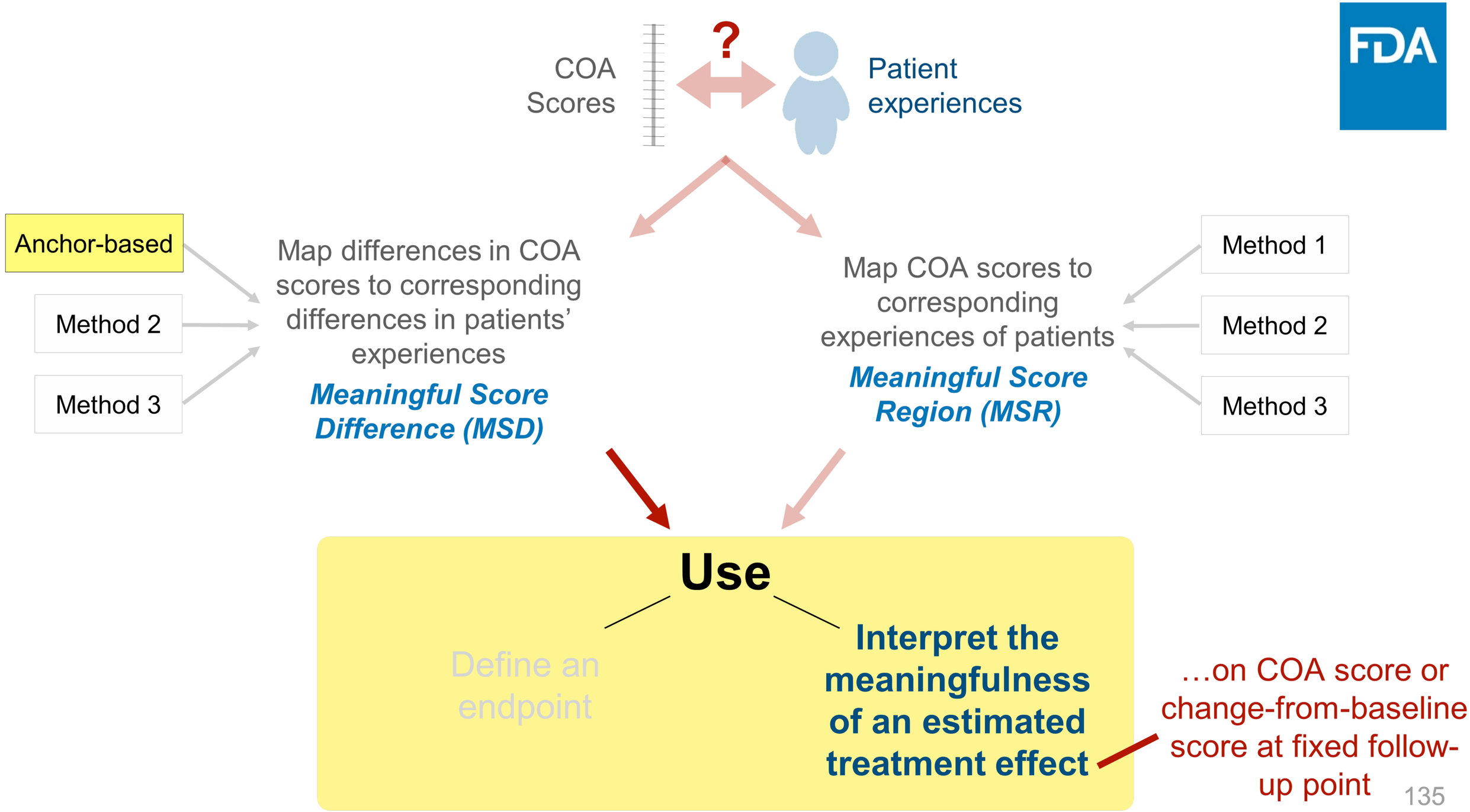
Expected Difference in Scores (or in Change-From-Baseline Scores):

The Horizontal Gap Between Groups



How much better is the average patient likely to feel if they receive Drug rather than Placebo?

[1]Holland PW. Two measures of change in the gaps between the CDFs of test-score distributions. J Edu Behav Stat. 2002;27(1):3-17.

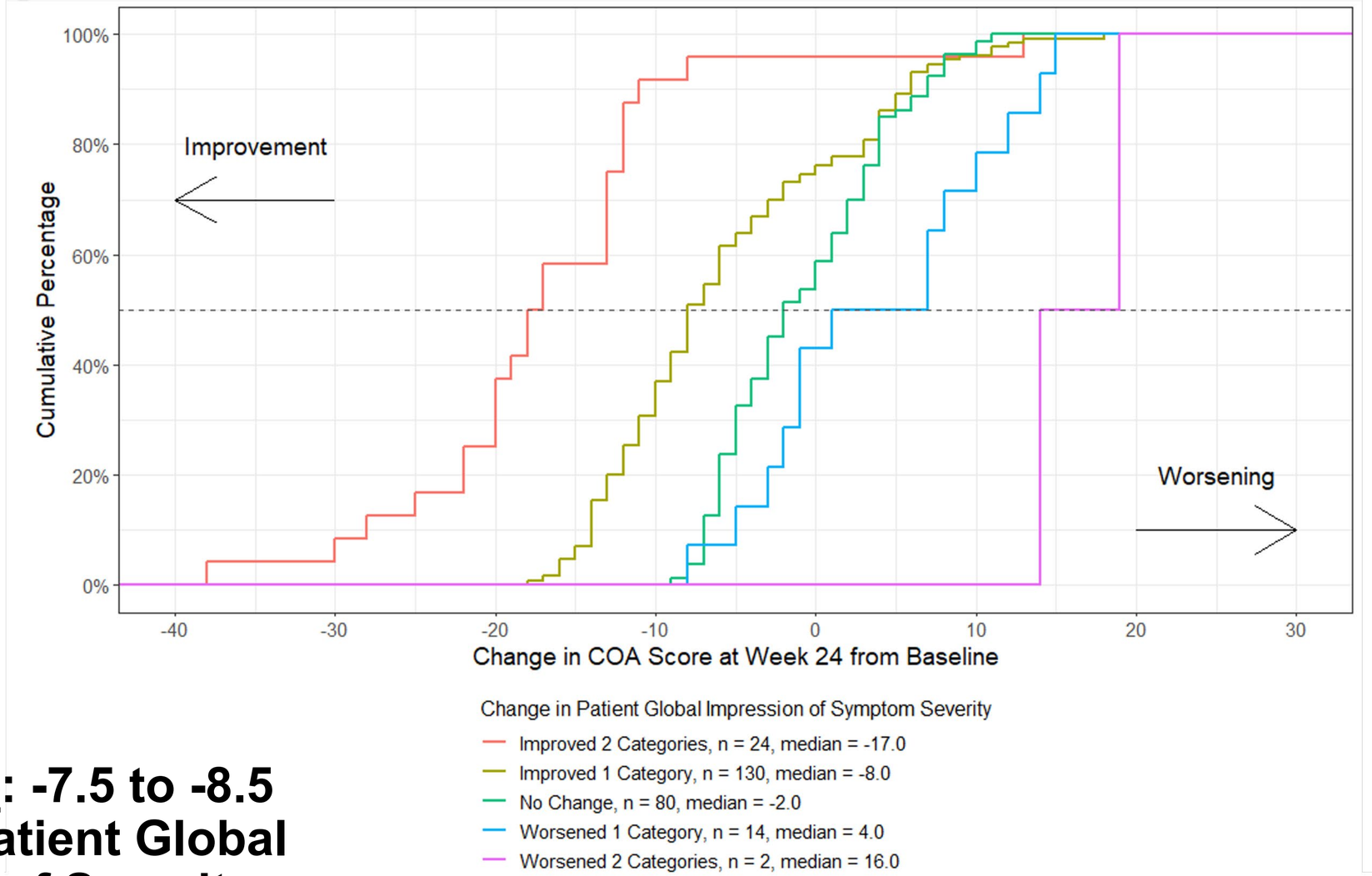

COA Scores
?
Patient experiences
Anchor-based
Method 2
Method 3
Map differences in COA scores to corresponding differences in patients' experiences
Meaningful Score Difference (MSD)
Map COA scores to corresponding experiences of patients
Meaningful Score Region (MSR)
Method 1
Method 2
Method 3
Use
Define an endpoint
Interpret the meaningfulness of an estimated treatment effect
…on COA score or change-from-baseline score at fixed follow-up point

Hypothetical Example: Randomized Clinical Trial

Randomized Clinical Trial of Drug vs Placebo

- Parallel groups design
- N = 250
- Endpoint: Change in *ABC Symptom Index* score from baseline to Week 24
- *ABC Symptom Index* assessed at baseline and 24 weeks post-randomization
  - Score range: 0 to 40
  - Higher score indicate greater symptom severity
- Primary analysis: Comparison of study group mean change-from-baseline scores using analysis of covariance (ANCOVA) with baseline *ABC Symptom Index* score as covariate

Hypothetical Study to Derive Meaningful Score Difference (Anchor-Based)



- **<u>MSD Range</u>: -7.5 to -8.5 based on Patient Global Impression of Severity anchor measure**

Hypothetical Example: Randomized Clinical Trial

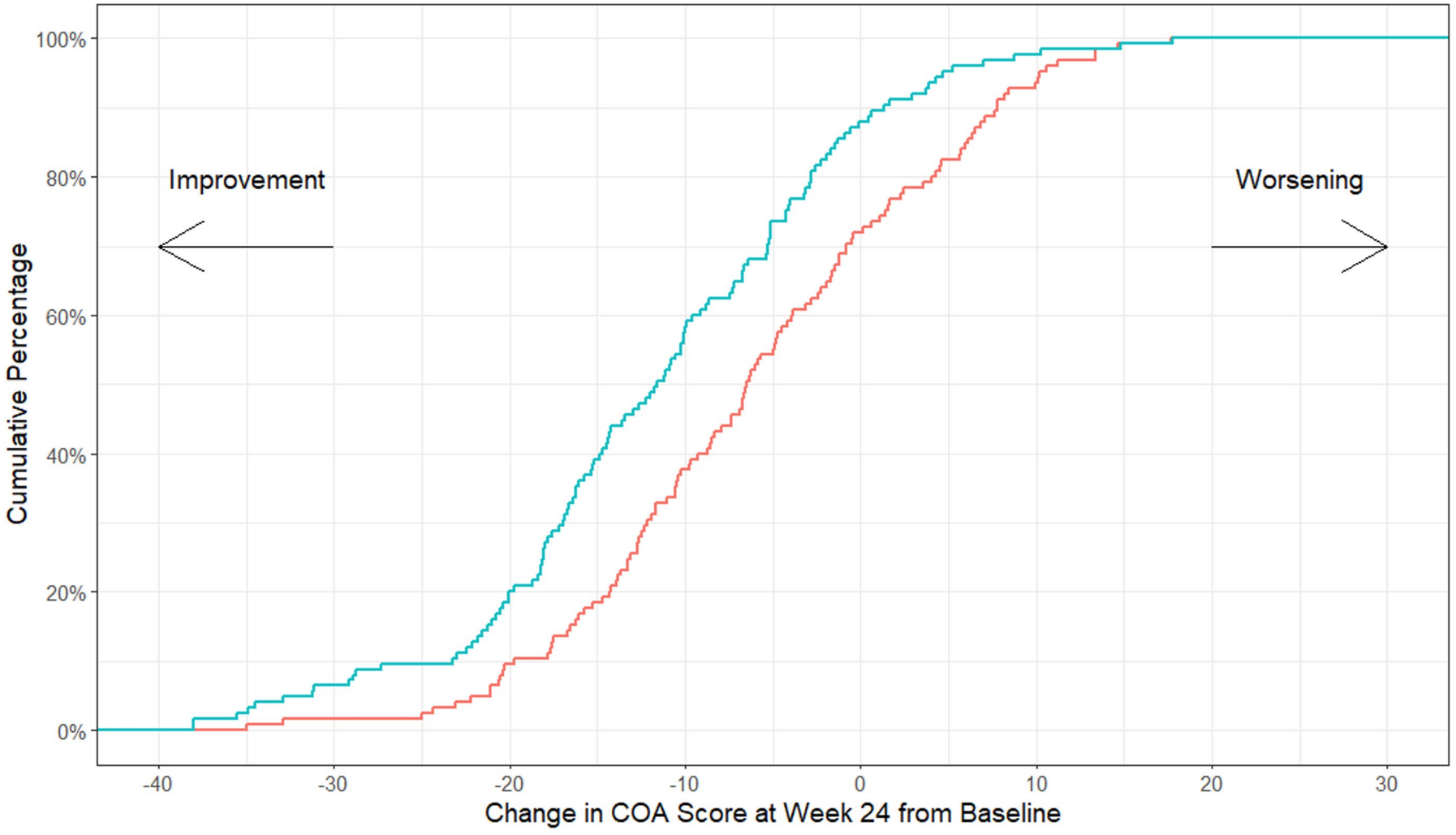
Treatment Group	LS Mean	SE	95% CI
Placebo	28.8	1.21	[26.4, 31.2]
Drug	20.5	1.09	[18.4, 22.6]
LS Mean Difference	-8.3	1.63	[-11.5, -5.1]

Results obtained from an ANCOVA model with covariates treatment arm and baseline COA score.

This is an estimate of the causal effect of treatment for the typical patient in the trial

FDA draft guidance for Industry Adjusting for Covariates in Randomized Clinical Trials for Drugs and Biological Products (May 2021)

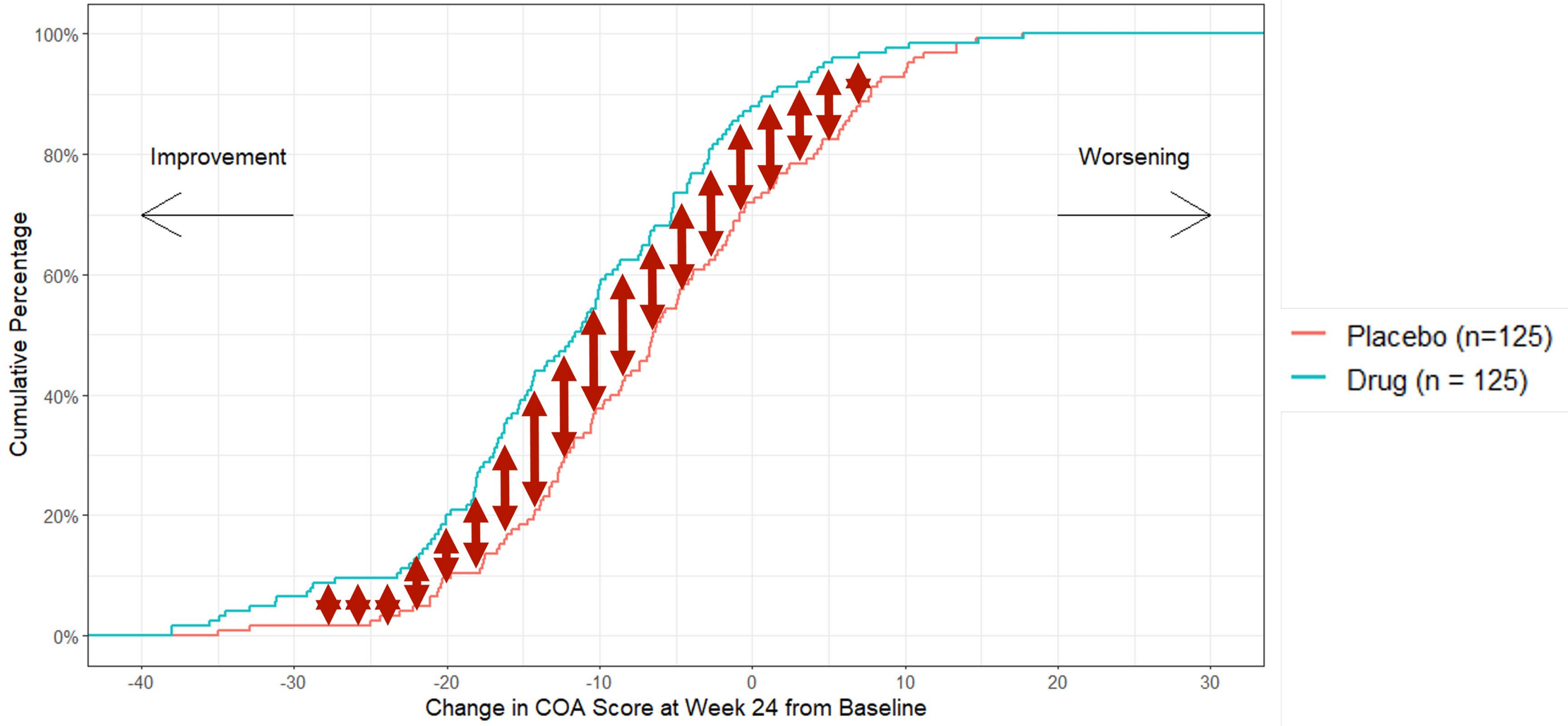
Empirical Cumulative Distribution Function: Change from Baseline to Week 24 in ABC Symptom Index Score by Treatment Group



Expected difference in the probability of exceeding MSDs thresholds (-7.5 to -8.5)

How much more likely are the average patient to experience a meaningful improvement in their ABC Symptoms if given Drug rather than Placebo?

Empirical Cumulative Distribution Function: Change from Baseline to Week 24 in ABC Symptom Index Score by Treatment Group



- **<u>MSD Range</u>: -7.5 to -8.5**

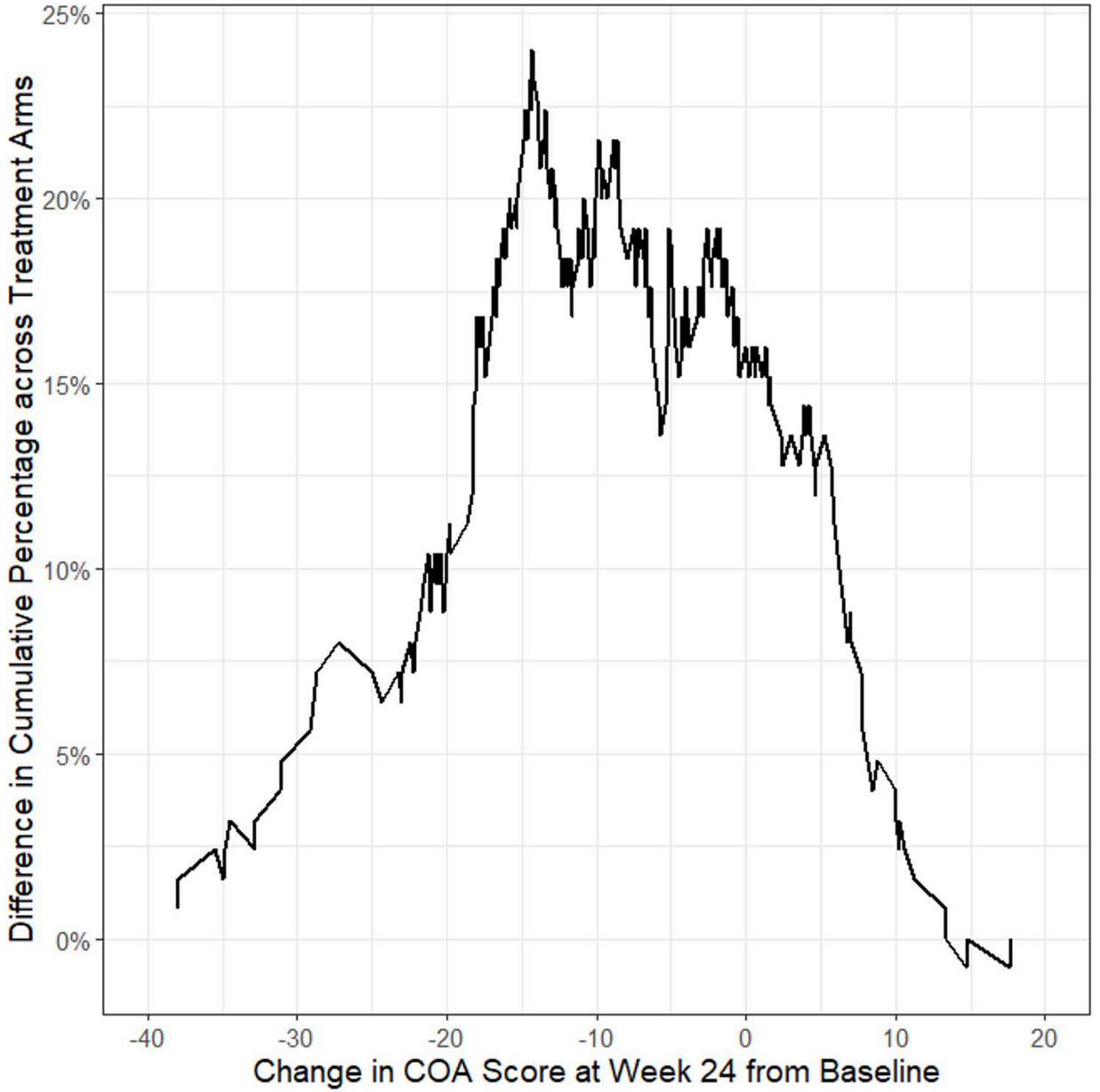
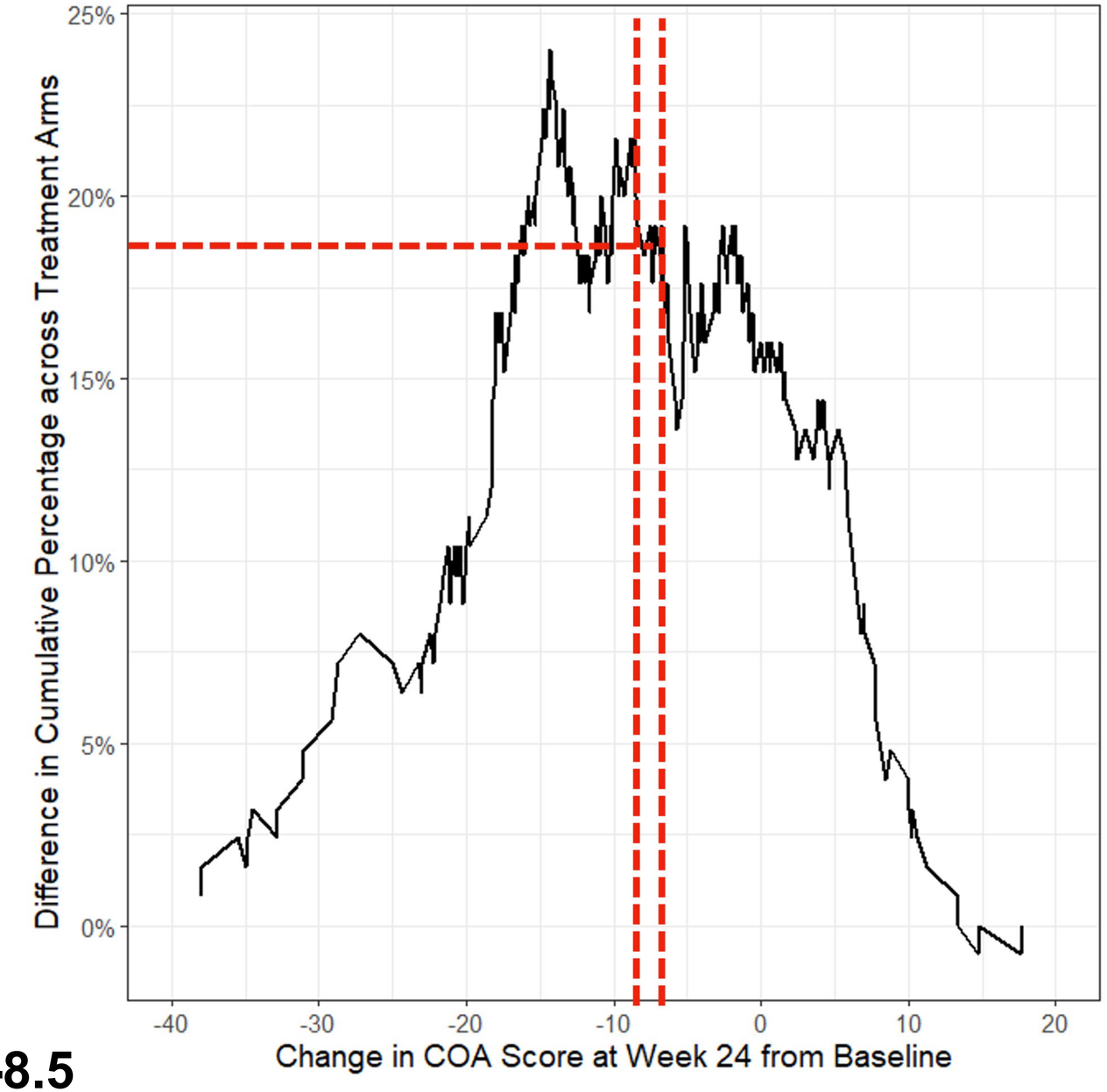


Figure based on Revicki DA, Erickson PA, Sloan JA, et al. Interpreting and Reporting Results Based on Patient-Reported Outcomes. *Value Health*. 2007;10(s2):S116-S124.

Difference = 18 to 20.5%

How much more likely is the average patient to experience a meaningful improvement in their ABC Symptoms if given Drug rather than Placebo?

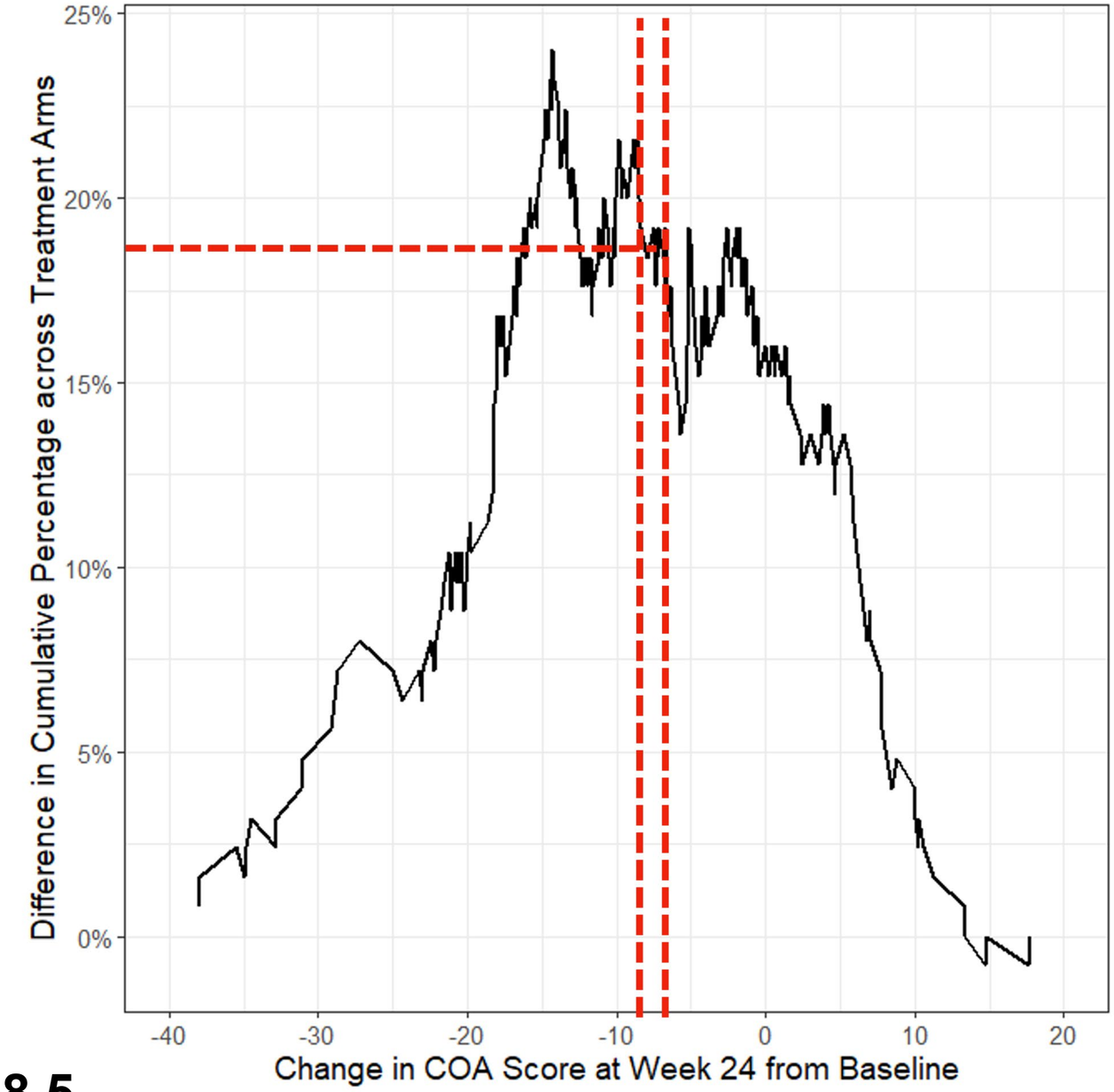


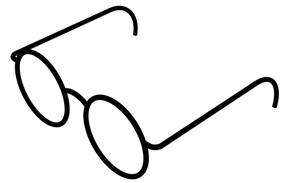
- **<u>MSD Range</u>: -7.5 to -8.5**

Figure based on Revicki DA, Erickson PA, Sloan JA, et al. Interpreting and Reporting Results Based on Patient-Reported Outcomes. *Value Health*. 2007;10(s2):S116-S124.

Difference = 18 to 20.5%

How much more likely is the average patient to experience a meaningful improvement in their ABC Symptoms if given Drug rather than Placebo?





Size of difference at any score threshold is dependent upon separation between group means *and* amount of within-group variance

(Abugov et al., *Pharm Stat* 2023;22(2):312-327)

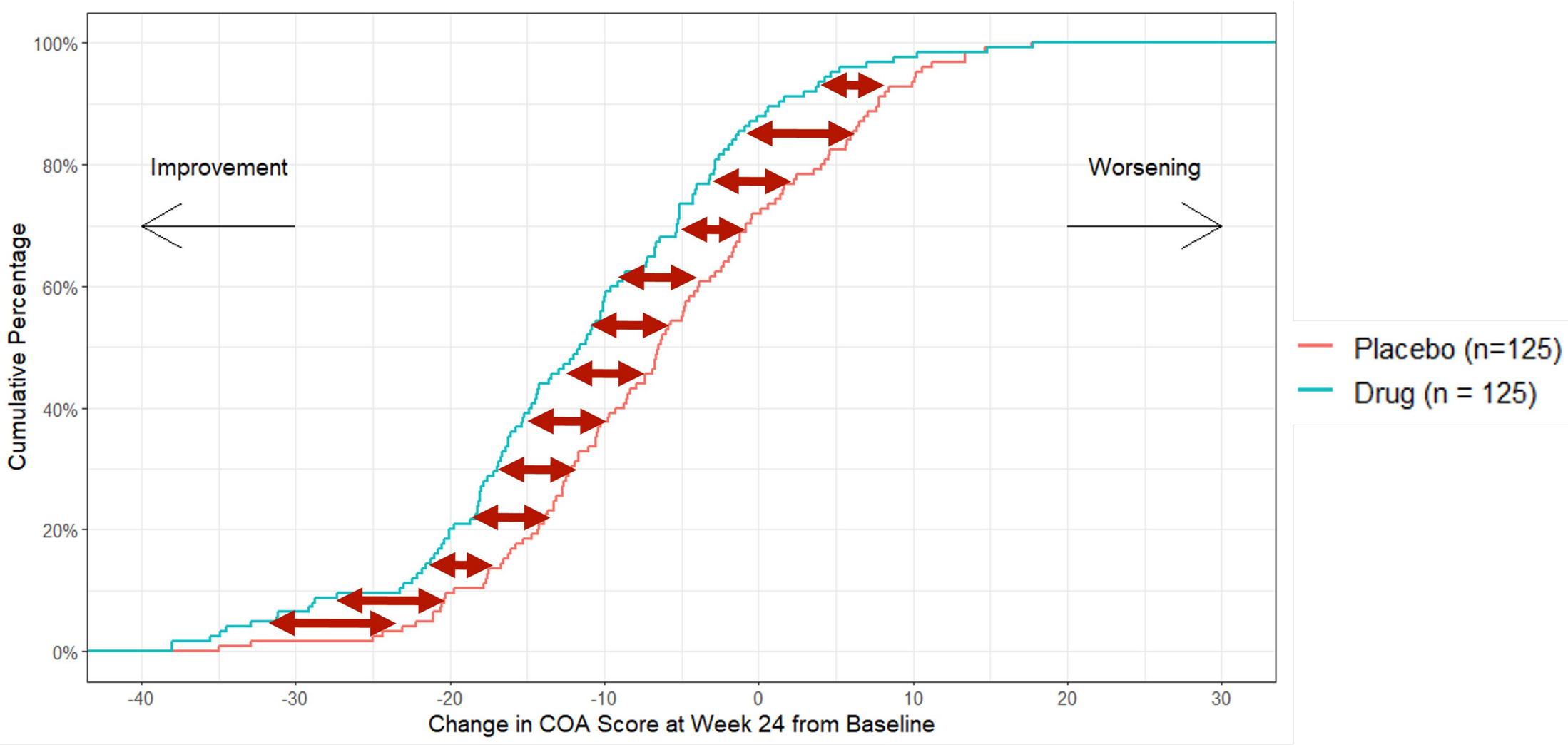
- **<u>MSD Range</u>: -7.5 to -8.5**

Figure based on Revicki DA, Erickson PA, Sloan JA, et al. Interpreting and Reporting Results Based on Patient-Reported Outcomes. *Value Health*. 2007;10(s2):S116-S124.

Expected difference in change from baseline to week 24 ABC Symptom Index scores

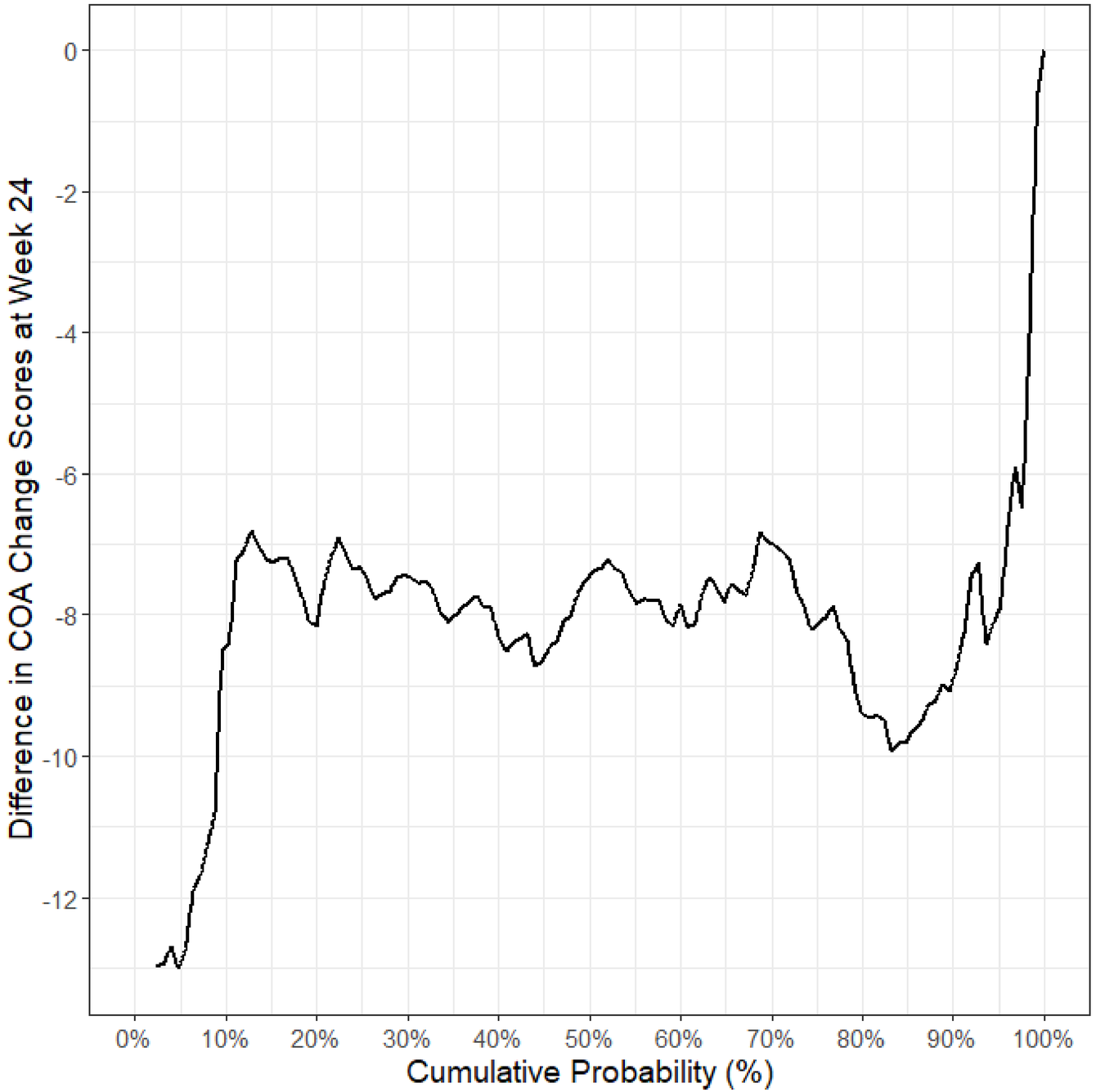
How much better is the average patient's ABC Symptoms likely to be if they receive Drug rather than Placebo?

- Overall estimate of treatment effect (difference between group means) corresponds to the average horizontal gap between eCDFs
- Check to see if it is relatively consistent



Directly plotting the horizontal gap shows it is relatively consistent

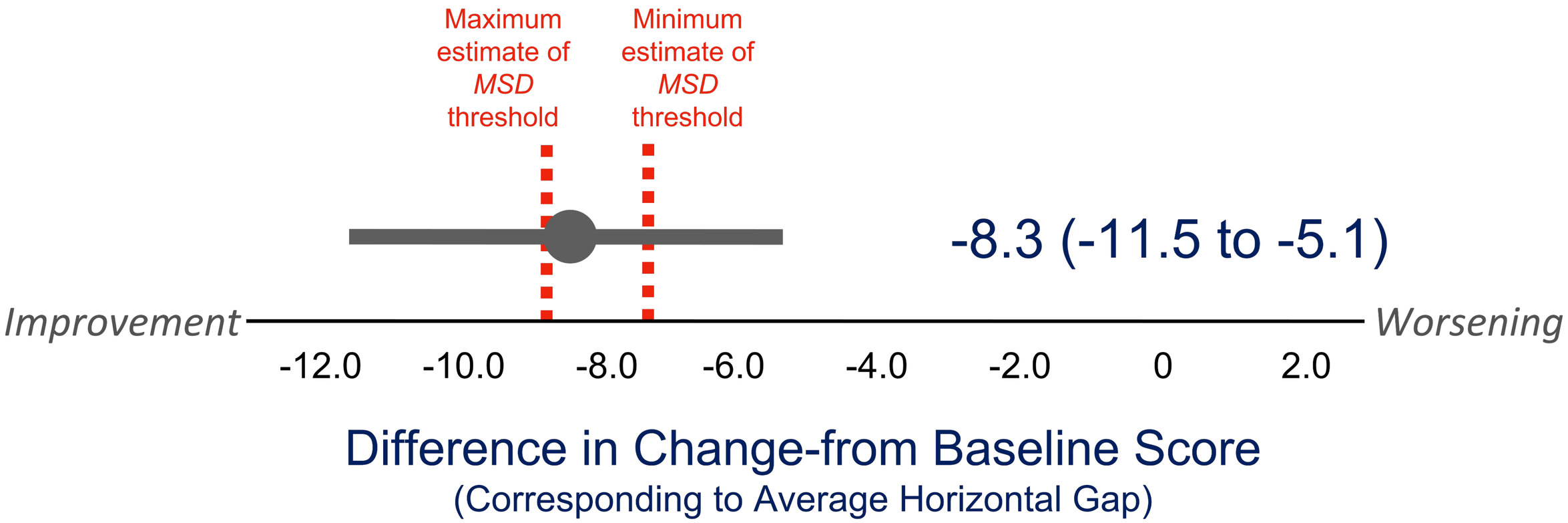
Supports use of difference in group means to estimate size of treatment effect



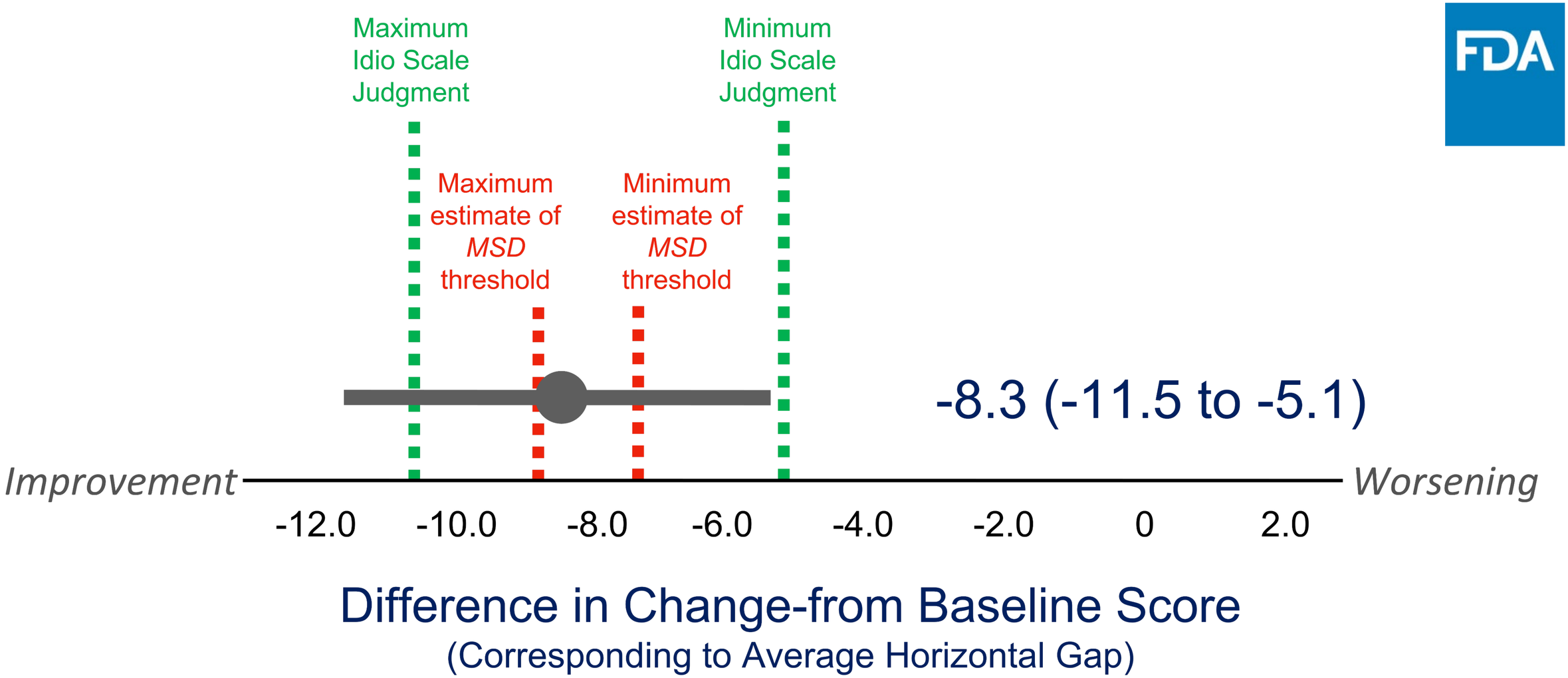
Estimated Difference in Adjusted Means (With 95% Confidence Interval) Between Treatment and Placebo on Change-from Baseline

How much better are the average patient's ABC Symptoms likely be if they receive Drug rather than Placebo?

Estimated Difference in Adjusted Means (With 95% Confidence Interval) Between Treatment and Placebo on Change-from Baseline



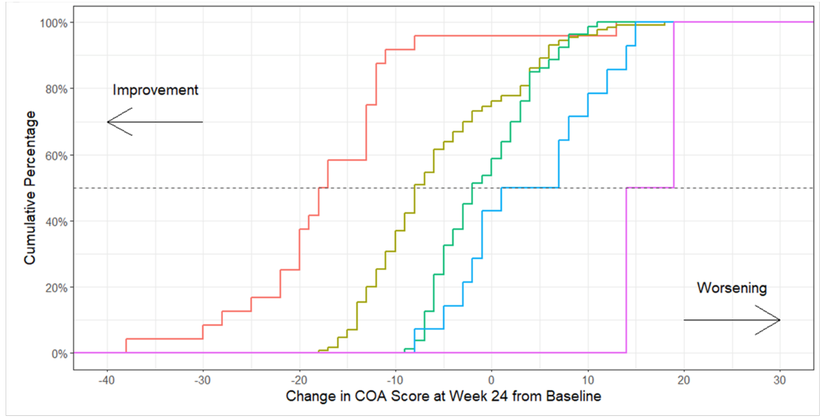
How much better are the average patient's ABC Symptoms likely be if they receive Drug rather than Placebo?



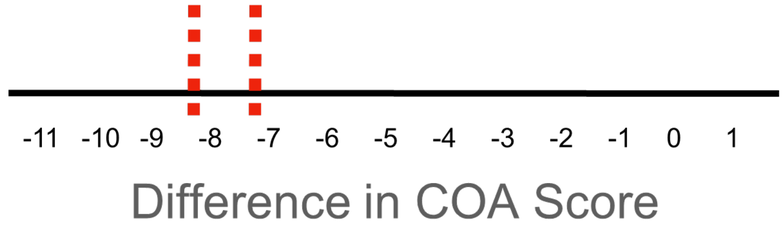
- Not a second level of statistical hypothesis testing
- MSDs are imprecise points of reference that help put the treatment effect in context

Estimating Meaningful Score Difference (*Anchor-based Method*)

Compare distributions from groups of patients

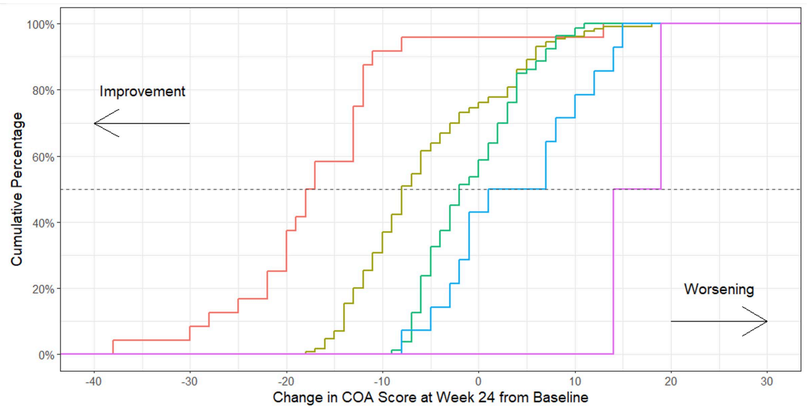


Expected difference in score perceived as meaningful by the average patient

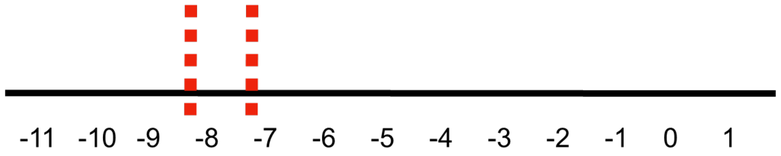


Estimating Meaningful Score Difference (*Anchor-based Method*)

Compare distributions from groups of patients



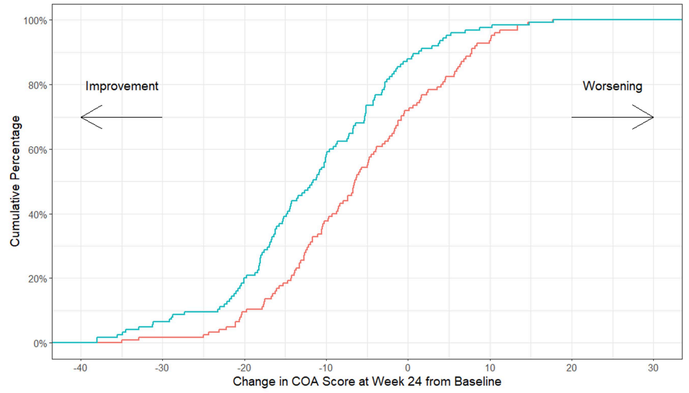
Expected difference in score perceived as meaningful by the average patient



Difference in COA Score

Estimating Treatment Effect (*Parallel Groups Design*)

Compare distributions from groups of patients



Expected difference in probability of exceeding MSD thresholds (-7.5 to –8.5)

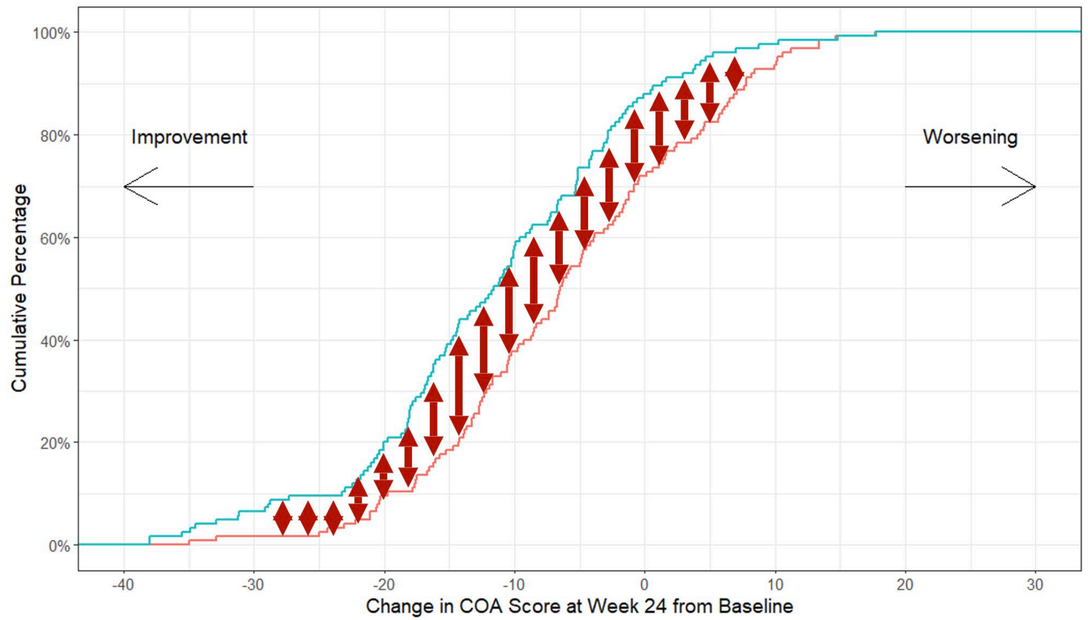
&

Expected difference in change from baseline to Week 24 ABC Symptom Scores

How we examine the meaningfulness of the treatment effect depends on how we look at the group comparison

Expected Difference in the Probability of Exceeding One or More Score Thresholds:

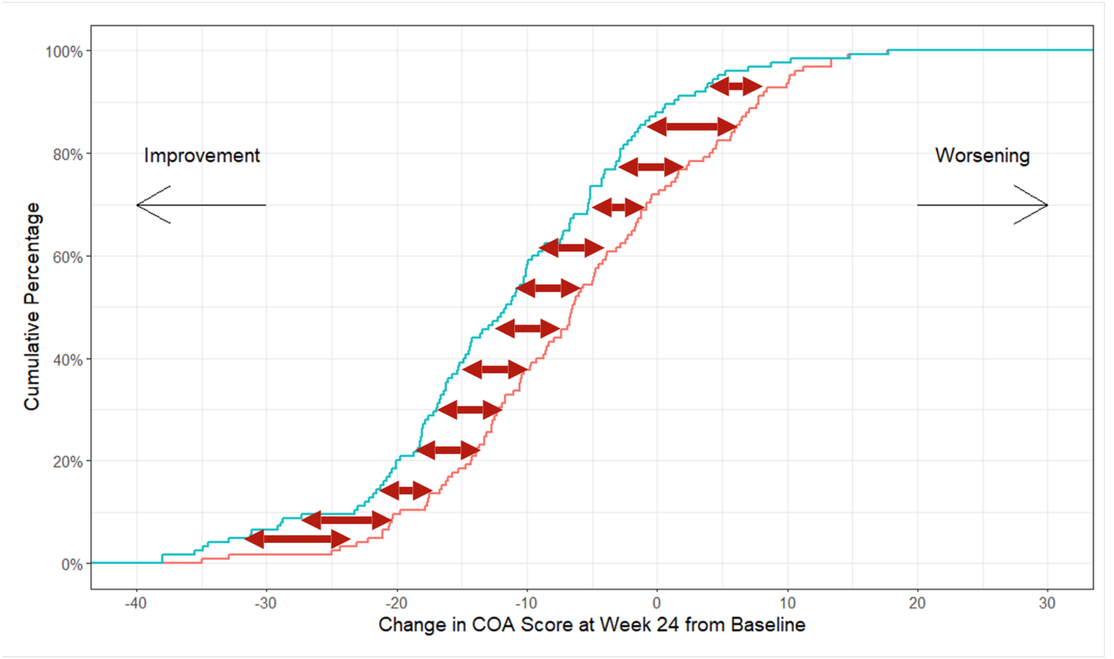
The Vertical Gap Between Groups



How much more likely are the average patient to experience a meaningful improvement in their ABC Symptoms if given Drug rather than Placebo?

Expected Difference in Scores (or Change-From-Baseline Scores):

The Horizontal Gap Between Groups



How much better is the average patient's ABC Symptoms likely to be if they receive Drug rather than Placebo?

Interpreting the Meaningfulness of Treatment Effects

- Using MSDs or MSRs is not like using an algorithm to produce a yes/no answer about meaningfulness
  - It is about creating a richer context in which to view the estimates of treatment effect
- Using MSDs or MSRs to help interpret treatment effects is just one part of assessing the meaningfulness of treatment effects

  Other considerations:

  - Findings using other methodologies or anchors to derive MSDs or MSRs
  - Treatment effects on other endpoints
  - Prespecified sensitivity analyses
  - Analyses to examine heterogeneity of treatment effect
  - Graphical/exploratory analyses to examine analytic assumptions


Patient-Focused
Drug Development
FDA Wants
To Hear
From Patients
FDA
U.S. FOOD & DRUG
ADMINISTRATION

Patient-Focused Drug Development

Question and Answer

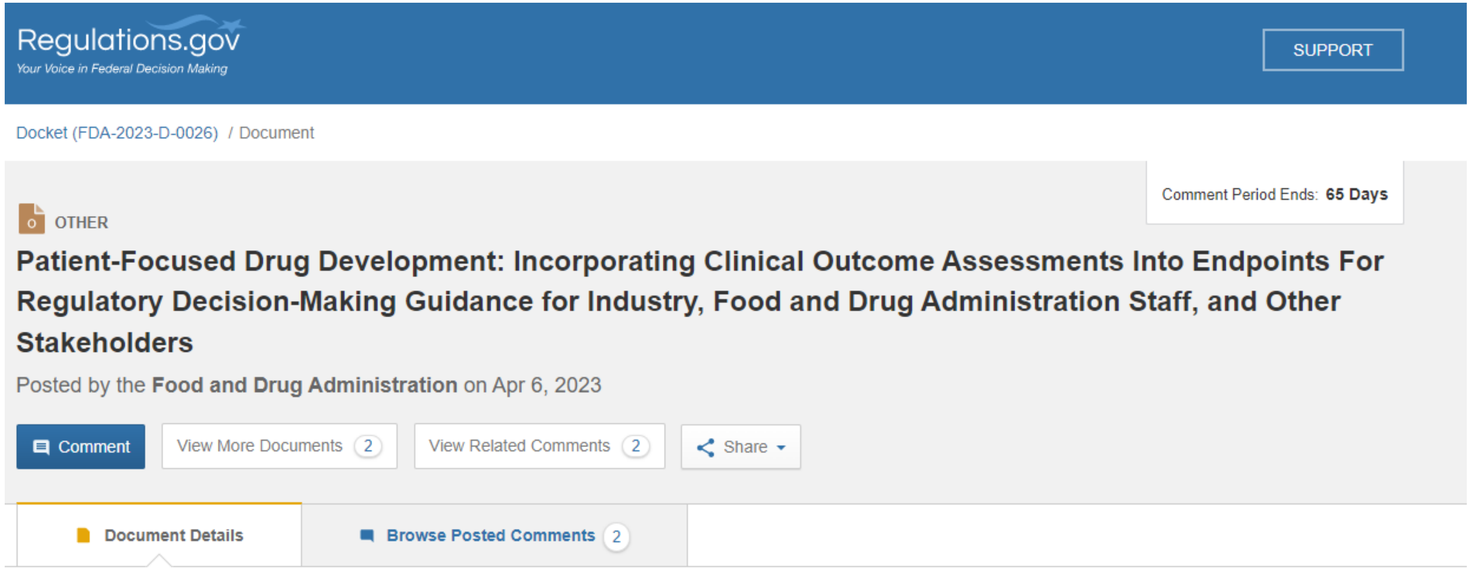
Send us your comments!

Interested stakeholders are invited to submit comments on the draft guidance to the public docket.

The docket will close on July 5, 2023.

How do you submit a comment?

- Please visit: https://www.regulations.gov/docket/FDA-2023-D-0026
- And **Click Comment**



U.S. FOOD & DRUG ADMINISTRATION

Thank you!