FDA | U.S. FOOD & DRUG ADMINISTRATION

Public Workshop

Patient-Focused Drug Development: Workshop #2 to Discuss Methodologic and Other Challenges Related to Patient Experience Data

September 18-19, 2025

Welcome

Ethan Gabbour, MS

Public Health Analyst, Patient-Focused Drug Development Program

Office of the Center Director

Center for Drug Evaluation and Research

US Food and Drug Administration

Agenda – Day 1

FDA

12:30	Welcome & Opening Remarks
12:40	Understanding the Intersection of Meaningful Aspect(s) of Health and Concept of Interest
1:40	Understanding Patient Baseline Severity and How it May Impact a COA's Ability to Detect Change
2:40	Break
2:55	Challenges and Opportunities for Deriving and Applying Meaningful Score Regions
4:55	Day 1 Wrap-Up

Agenda – Day 2

FDA

12:30	Welcome
12:35	Information on a COA Instrument Reviewed by the FDA: Updates to, and Learnings from, the 2009 Outline of the PRO Dossier
2:35	Break
2:50	Clinical Outcome Assessment Data: Interaction of Data Standardization and Statistical Programming
3:50	Closing Remarks

FDA U.S. FOOD & DRUG ADMINISTRATION

Patient-Focused Drug Development: Workshop #2 to Discuss Methodologic and Other Challenges Related to Patient Experience Data

September 18, 2025

Theresa M Mullin, PhD

Associate Center Director | Strategic Initiatives

US FDA Center for Drug Evaluation and Research

A primary goal for FDA is to hear patients' perspectives on living with their disease

Patients' Perspectives

- Which **symptoms have the most significant impact** on your daily life?... On your ability to do specific activities?

- How well does your **current treatment regimen** treat the most significant symptoms of your disease?

- What **factors do you take into account when making decisions** about using treatments?

Helps inform regulatory decision making on products to treat that disease

Integrating patient input into medical product development and decision making

FDA

What impacts (burden of disease and burden of treatment) matter most to patients and how do we measure them?

What aspects of clinical trials can be better tailored to meet the needs of patients who (might) participate in the trial?

How do we better integrate patient reported outcome data or elicited patient preferences into Benefit-Risk (BR) assessments?

How do we best communicate information to patients and prescribers?

Translational **Clinical Trials** **Pre-market review** **Post-market**

Need to build in patient input starting in the translational phase

PFDD Meetings

FDA

Designed to engage patients and elicit their perspectives on two topic areas:

(1) the most significant symptoms of their condition and the impact of the condition on daily life;

(2) their current approaches to treatment.

FDA has conducted over 30 PFDD meetings

Recently:

- Systemic Sclerosis
- Stimulant Use Disorder
- Vitiligo
- Long COVID (English and Spanish)

There have been more than 110 Externally-Led PFDD Meetings conducted by Patient Groups

https://www.fda.gov/industry/prescription-drug-user-fee-amendments/fda-led-patient-focused-drug-development-pfdd-public-meetings
https://www.fda.gov/industry/prescription-drug-user-fee-amendments/externally-led-patient-focused-drug-development-meetings

Methodological Guidance to enable stakeholders to go beyond powerful narrative and collect data that can serve as study endpoints and be used as a basis for marketing decisions

Dimension	Evidence and Uncertainties	Conclusions and Reasons
Analysis of Condition	PFDD Meetings and Reports provide powerful narrative that gives regulators insights about clinical context and what matters to patients	
Current Treatment Options		
Benefit	Using measures & tools (COAs) to systematically capture what matters most during clinical trials can turn narrative into evidence for regulatory decision making	
Risk and Risk Management		
Benefit-Risk Summary and Assessment		

Methodologic Guidance Documents

Collecting Comprehensive and Representative Input

Methods to Identify What is Important to Patients

Selecting, Developing or Modifying Fit-for-Purpose Clinical Outcome Assessments

Incorporating Clinical Outcome Assessments into Endpoints for Regulatory Decision Making

Attending to quality will enhance usability in decision making

- The quality of collected patient experience data will determine the extent to which it can be used to inform regulatory decision making

- **Today's workshop is focusing on specific areas of methodologic challenge identified by FDA and stakeholders including medicine developers**

Thank you for joining us today!

Understanding the Intersection of Meaningful Aspect(s) of Health and Concept of Interest

Overview of Session

FDA

Presentation (approx. 10 minutes)

Understanding the Intersection of Meaningful Aspect(s) of Health and Concept(s) of Interest: Building Bridges from Patient Voice to Regulatory Evidence

- Ji Li - Senior Staff Fellow, Division of Clinical Outcome Assessment, Center for Drug Evaluation and Research, US Food and Drug Administration

Panel Discussion (approx. 35 minutes)

- Bill Byrom - Vice President of Product Intelligence and Positioning, and Principal eCOA Science, Signant Health

- Erin O'Brien - Psychometrician, Division of Patient-Centered Development, Center for Devices and Radiological Health

- Elektra Papadopoulos - Director, Patient Experience Data & Strategy in Immunology and Oncology, AbbVie

- Poornima Sharma - Physician, Malignant Hematology Branch, Office of Clinical Evaluation, Center for Biologics Evaluation and Research, US Food and Drug Administration

Audience Questions and Answers (approx. 15 minutes)

Understanding the Intersection of Meaningful Aspect(s) of Health and Concept(s) of Interest:

Building Bridges from Patient Voice to Regulatory Evidence

Patient-Focused Drug Development: Workshop #2 to Discuss Methodologic and Other Challenges Related to Patient Experience Data

September 18, 2025

Overview

- A roadmap to patient-focused outcome measurement
- Evidence to support the fitness-for-purpose of selected clinical outcome assessments (COAs)
- Understanding the patient's perspective of meaningful aspects of health (MAHs)
- Interpretation of concepts of interest (COIs) from the patient experience
- Connecting MAHs to COIs
- Additional considerations
- Summary

Using Patient Experience Data to Inform Patient-Focused Drug Development

- **Ensuring clinical trial outcomes reflect what truly matters to patients** - The path from patient voice to regulatory evidence requires careful navigation to capture meaningful endpoints.

- **Two central elements enable this transformation** - MAHs and COIs serve as the foundation for systematic patient experience integration for COAs.

- **The ultimate goal is transforming patient experiences into measurable outcomes** - This process converts qualitative patient insights into quantitative data that can support regulatory decision-making and drug development.

Roadmap to Patient-Focused Outcome Measurement in Clinical Trials

Understanding the Disease or Condition

- Patient/caregiver perspectives
- Natural history of the disease or condition
- Patient subpopulations
- Health care environment
- Other expert input (healthcare providers, payers, regulators)

Conceptualizing Clinical Benefits and Risks

- Identify concept(s) of interest (COI), i.e. how a patient feels, functions, or survives
- Define context of use (COU) for clinical trial

Selecting/Developing the Outcome Measure

Select clinical outcome assessment (COA) type: PRO, ObsRO, ClinRO, or PerfO measure

→ Search for existing COA measuring concept of interest in context of use

- COA exists for COI, can be used unmodified for COU → Use existing COA
- COA exists for COI, but might need to be modified for COU → Collect evidence and modify COA as necessary
- No COA exists for COI and COU → Develop new COA and empirically evaluate

Fit-for-Purpose COA

A. COI and COU clearly described
B. Clear rationale
C. Sufficient evidence to justify rationale

Source: Patient-Focused Drug Development (PFDD) draft guidance, "Selecting, Developing, or Modifying Fit-for-Purpose Clinical Outcome Assessments", https://www.fda.gov/media/159500/download

Eight Components Comprising an Evidence-Based Rationale for Proposing a COA as Fit-for-Purpose

A	The concept of interest should be assessed by [*COA type*] because . . .
B	The COA selected captures all the important parts of the concept of interest.
C	The COA is administered appropriately.
D	Respondents understand the instructions and items/tasks of the measure as intended by the measure developer.
E	The method of scoring responses to the COA is appropriate for assessing the concept of interest.
F	Scores from the COA are not overly influenced by processes/concepts that are not part of the concept of interest.
G	Scores from the COA are not overly influenced by measurement error.
H	Scores from the COA correspond to the meaningful aspect of health related to the concept of interest.

Understanding of MAHs

- MAHs reflect dimensions of health that patients identify as important in their daily lives.

- **Key characteristics**:
  - Not laboratory or clinical biomarkers
  - Focus on some aspect of an individual's feeling or functioning
  - Based on patient priorities and lived experiences
  - Can be established through various pathways depending on the development context and existing knowledge base

Interpretation of COIs

- COI is what is specifically measured by a COA to help understand how a medical product affects a MAH.

- The COI may be identical to or related to MAH (direct or indirect).

- **Important characteristics:**
    - Reflects something important to patients
    - Can reasonably be modified through treatment
    - Can be measured reliably within trial timeframe

Connecting MAHs to COIs

From broad understanding to specified measurement:
- Begin with burden assessment (disease or treatment)
- Identify MAHs based on patient priorities and target of medical product
- Select measurable COIs for trial contexts

Framework construction:
- Conceptual framework depicting COA-based endpoint approach (includes **MAH**, **COI**, COA type, name of COA, COA endpoint)
- Measurement model showing how COA generates interpretable scores

Understanding the Disease/Condition

Conceptualizing Clinical Benefits & Risks

Selecting/Developing the Outcome Measure

Fit-for-Purpose COA

Illustration of Conceptual Framework for COIs Assessed by Different Types of COAs

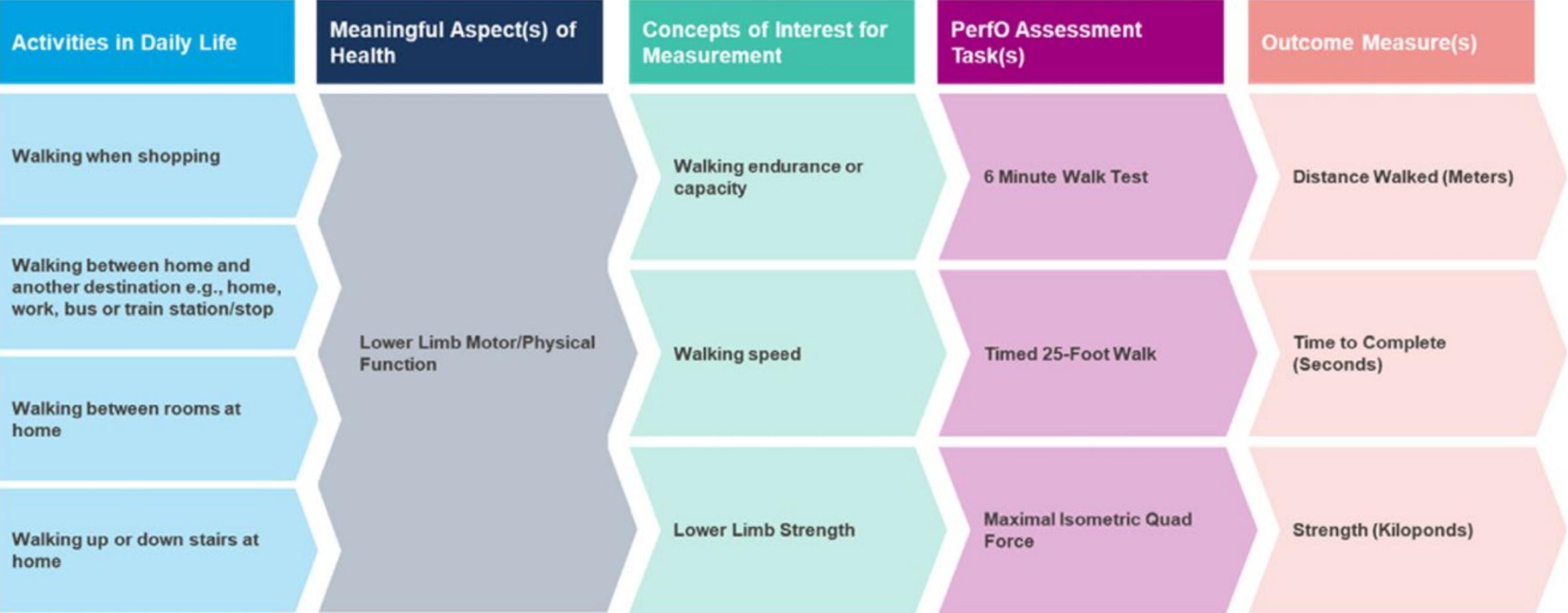Source: Patient-Focused Drug Development (PFDD) draft guidance, "Selecting, Developing, or Modifying Fit-for-Purpose Clinical Outcome Assessments", https://www.fda.gov/media/159500/download
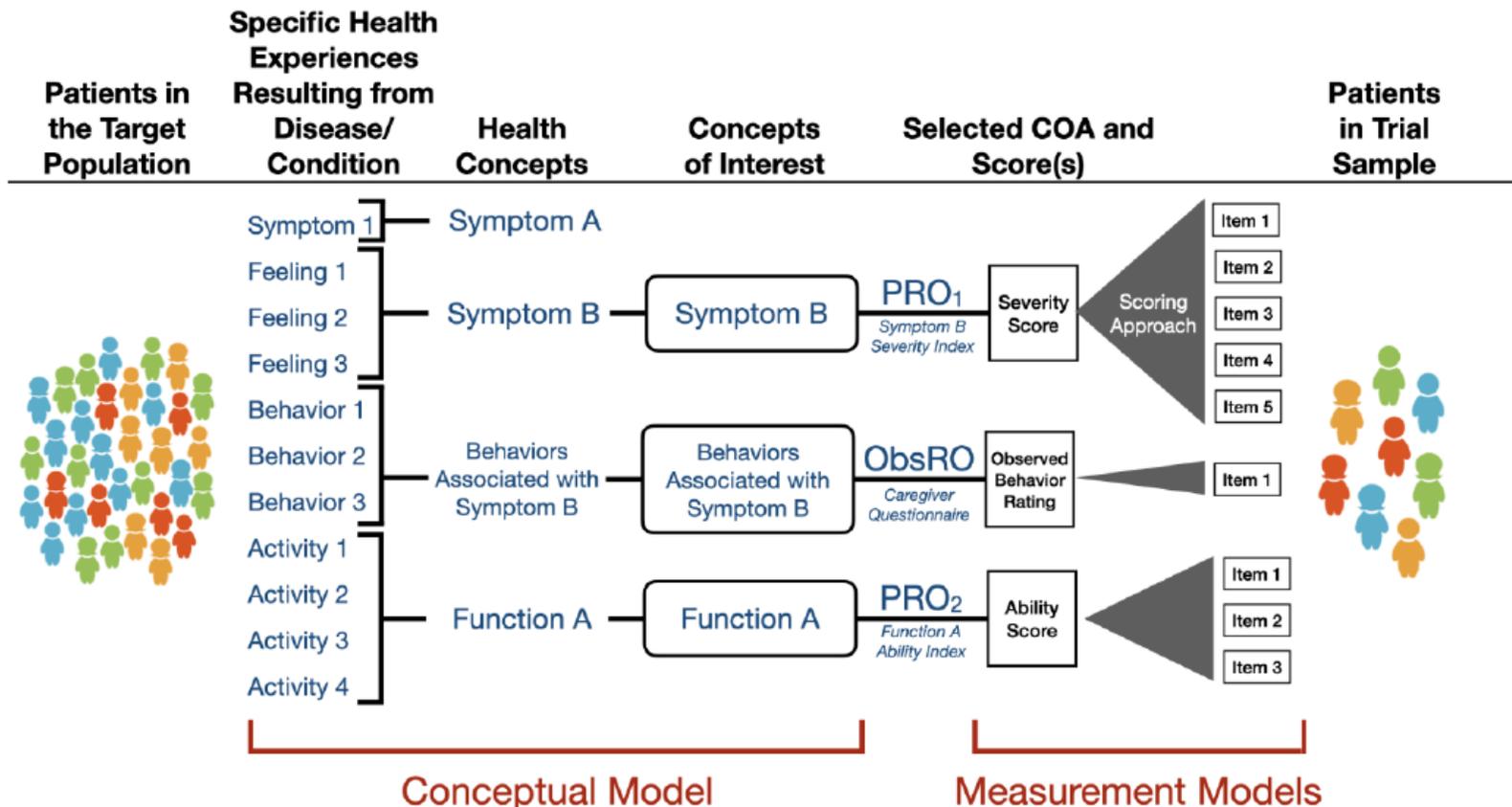
Examples:
- Crohn's disease
  - MAH: Fatigue
  - COI: Fatigue severity
  - COA: PRO
  - Endpoint: Change in PRO score

- Progressive familial intrahepatic cholestasis
  - MAH: Pruritus (itch)
  - COI: Itching intensity and scratching intensity
  - COA: PRO (itching) and ObsRO (scratching)
  - Endpoint: Change in itching intensity score, Change in scratching intensity score

Simple vs. Complex MAHs

- MAHs may represent:
  - Primary manifestations of a disease or condition (i.e., proximal or core aspects of a disease or condition)
    - Example: Assessment of pain intensity in endometriosis (menstrual and non-menstrual)
  - Downstream (i.e., distal) impacts of core aspects of a disease or condition
    - Example: Assessment of the impact of endometriosis pain intensity on daily activities
  - Multiple COIs

Direct vs. Indirect Reflections of MAHs

- A COI may directly or indirectly reflect the MAH.

- Direct measures (e.g., PRO)
  - Correspond closely to patients' lived experience(s)
  - Example: Assessment of pain intensity in patients with endometriosis who report pain as MAH

- Indirect measures (e.g., ObsRO, PerfO)
  - Assess concepts related to, but not identical to, patients' lived experience(s)
  - Used when MAH can not be validly and reliably assessed using direct measurement

FDA

Example of Multiple COIs Indirectly Reflecting the MAH

Activities in Daily Life	Meaningful Aspect(s) of Health	Concepts of Interest for Measurement	PerfO Assessment Task(s)	Outcome Measure(s)
Walking when shopping	Lower Limb Motor/Physical Function	Walking endurance or capacity	6 Minute Walk Test	Distance Walked (Meters)
Walking between home and another destination e.g., home, work, bus or train station/stop		Walking speed	Timed 25-Foot Walk	Time to Complete (Seconds)
Walking between rooms at home				
Walking up or down stairs at home		Lower Limb Strength	Maximal Isometric Quad Force	Strength (Kiloponds)

Additional Considerations

- Additional considerations
  - Collection of social media data
  - Application of digital health technology (DHT)
  - Assessment timepoints
  - COA score interpretability

Summary

- Early engagement: Involve patients to identify MAH and COI for COA measurement strategy

- Innovative methodologies: Different approaches and methods may be used to incorporate patient input into identifying MAH and COI

- Collaborative frameworks: Foster dialogue between sponsors, regulators, and patients

Panel Discussion
followed by
Q+A

Understanding Patient Baseline Severity and How it May Impact a COA's Ability to Detect Change

Overview of Session

Presentation (approx. 10 minutes)

Introduction to Common Challenges with Baseline COA Scores

- David Reasner - Division Director, Division of Clinical Outcome Assessment, Center for Drug Evaluation and Research, US Food and Drug Administration

Panel Discussion (approx. 40 minutes)

- Stacie Hudgens - Chief Strategy and Scientific Officer, Clinical Outcomes Solutions
- Marian Strazzeri - Mathematical Statistician, Division of Biometrics III, Office of Biostatistics, Center for Drug Evaluation and Research, US Food and Drug Administration
- Andrew Trigg - Statistical Innovation Lead, Bayer
- Therri Usher - Senior Mathematical Statistician, Division of Biometrics IV, Office of Biostatistics, Center for Drug Evaluation and Research, US Food and Drug Administration

Audience Questions and Answers (approx. 10 minutes)

Introduction to common challenges with baseline COA scores

Understanding Patient Baseline Severity and How it May Impact a COA's Ability to Detect Change
September 18th, 2025 | 1:40 pm – 2:40 pm

David Reasner, PhD
Division Director, Division of Clinical Outcome Assessment
FDA|CDER|OND|ODES|DCOA

Introduction to the Introduction

- This introduction presents **common patterns in baseline data that present challenges** in the interpretation of clinical outcome assessments (COAs) for clinical trials intended for regulatory decision-making.

Background

- COA scores may reflect an underlying construct (i.e., an instrument or scale) or simply tally the presence or absence of an event or activity (e.g., patient diary).

- Whether the way a patient feels and functions is expected to worsen during the study period is important when constructing endpoints (i.e., disease trajectory).

- We are often interested in measuring improvement at the item, domain, or instrument level, but also stabilization (i.e., no change) or worsening.

Measure what you don't expect to change!

Outline

- When Populations Differ

- When A Subpopulation Differs

- When There Is A Subset with "None" or a Floor Effect at Baseline

- When Misclassification Presents a Challenge

Simplifying Assumptions Suggest that Treatment May Reduce Symptoms Across the Population

Two Baseline Samples Centered at 100 units
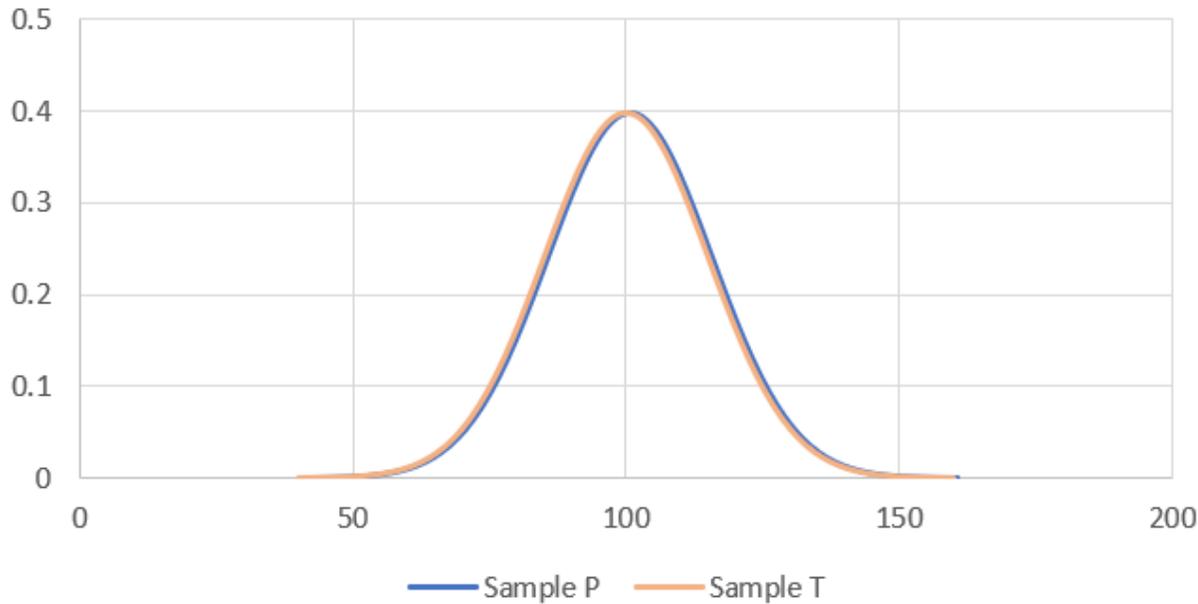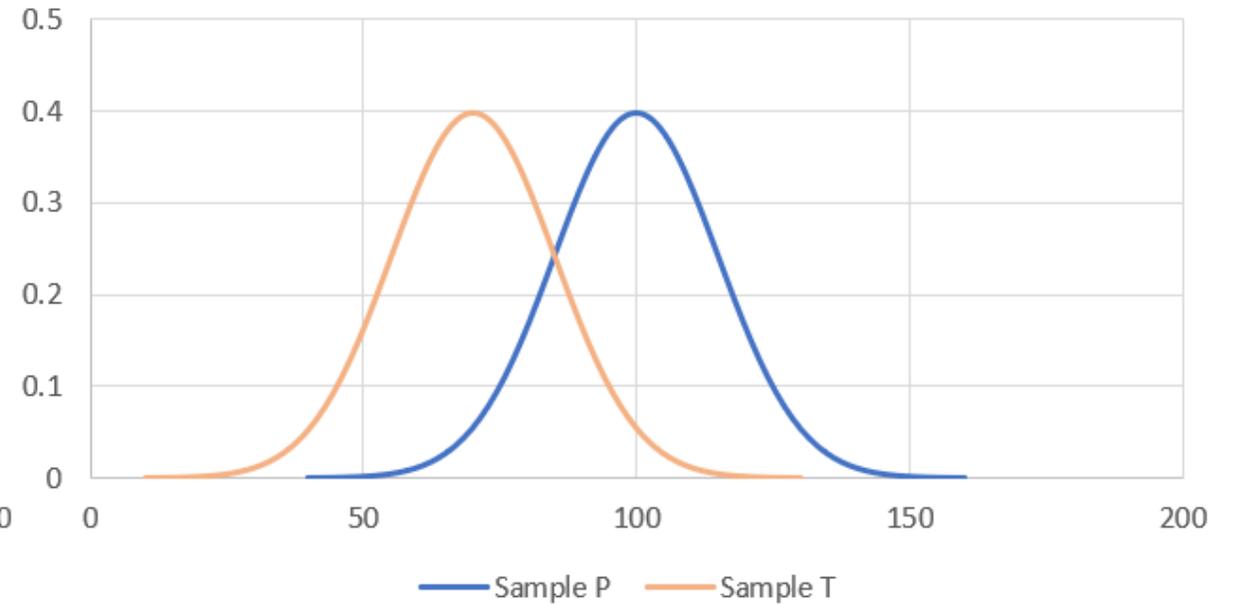
Two Samples After Treatment - Populations Differ

There is clear separation between the two treatments (e.g., sample means, $\bar{x}$, differ)

Two Baseline Samples Centered at 100 units

Two Samples After Treatment - Populations Differ

70   100

Sample P   Sample T
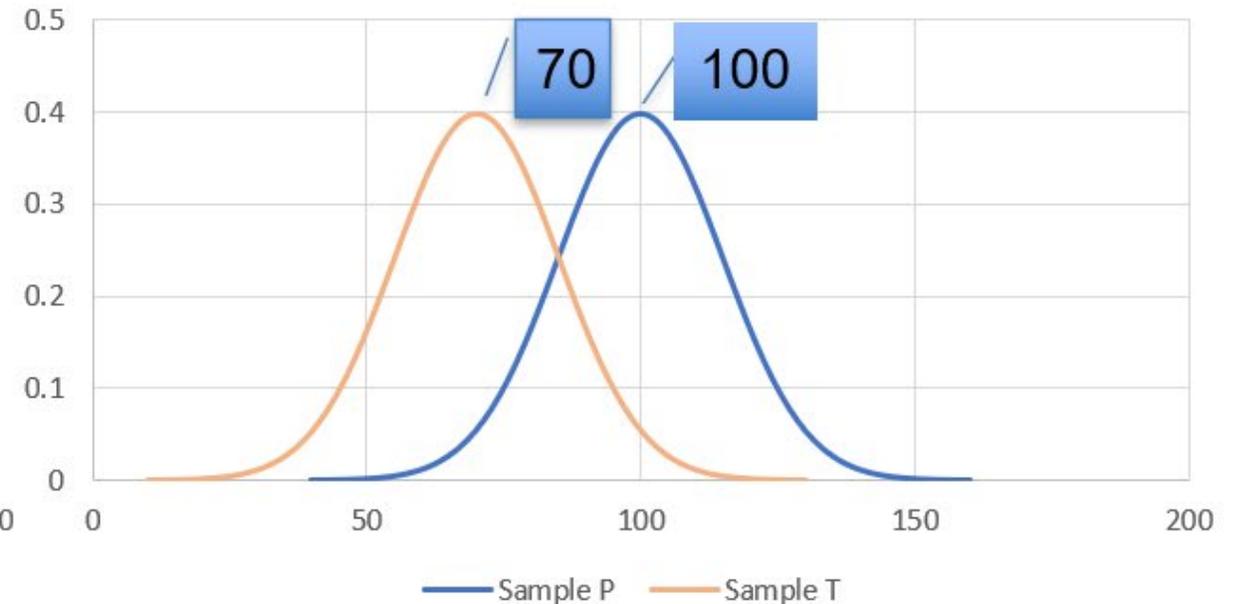
There is clear separation between the two treatments
(e.g., proportion of patients, $\hat{p}$, in a meaningful range differ)

Two Baseline Samples Centered at 100 units

Two Samples After Treatment - Populations Differ

0.52

0.02
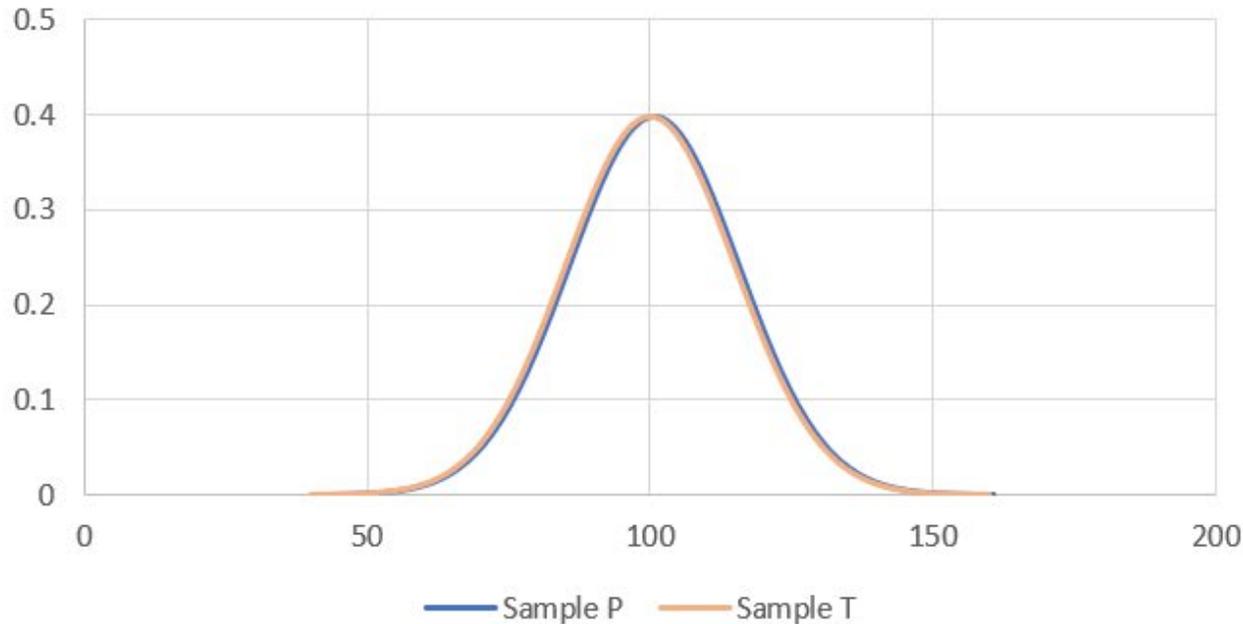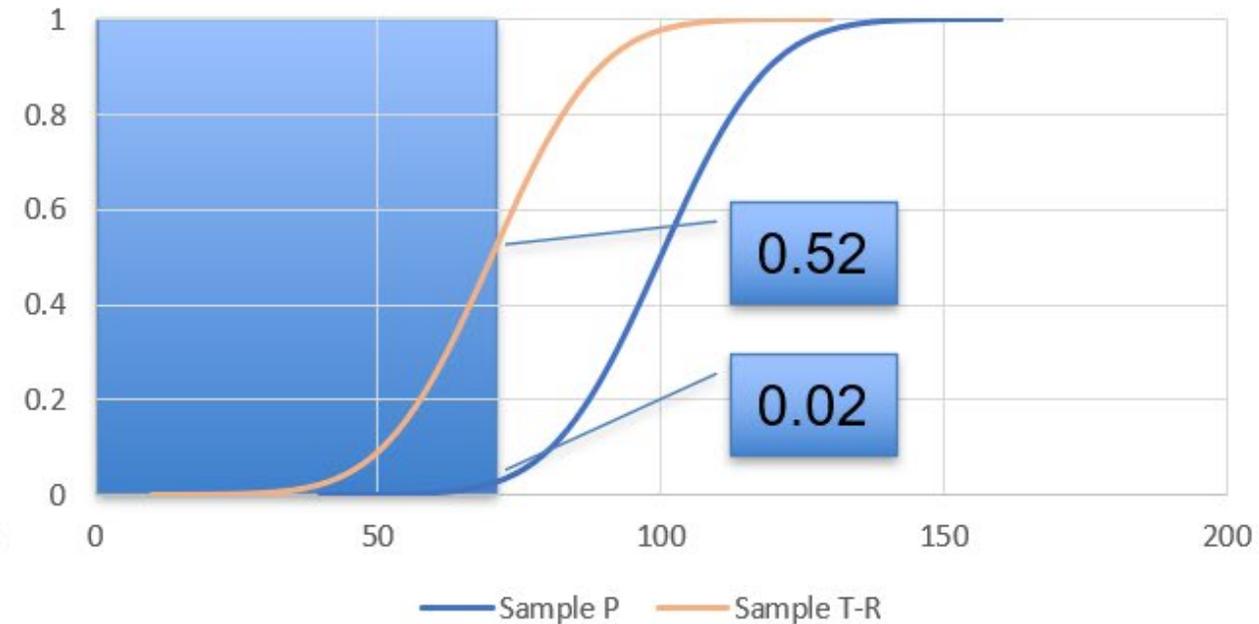
Scoring below 70 is meaningful to patients

Alternatively, Treatment May Reduce Symptoms in a Subset
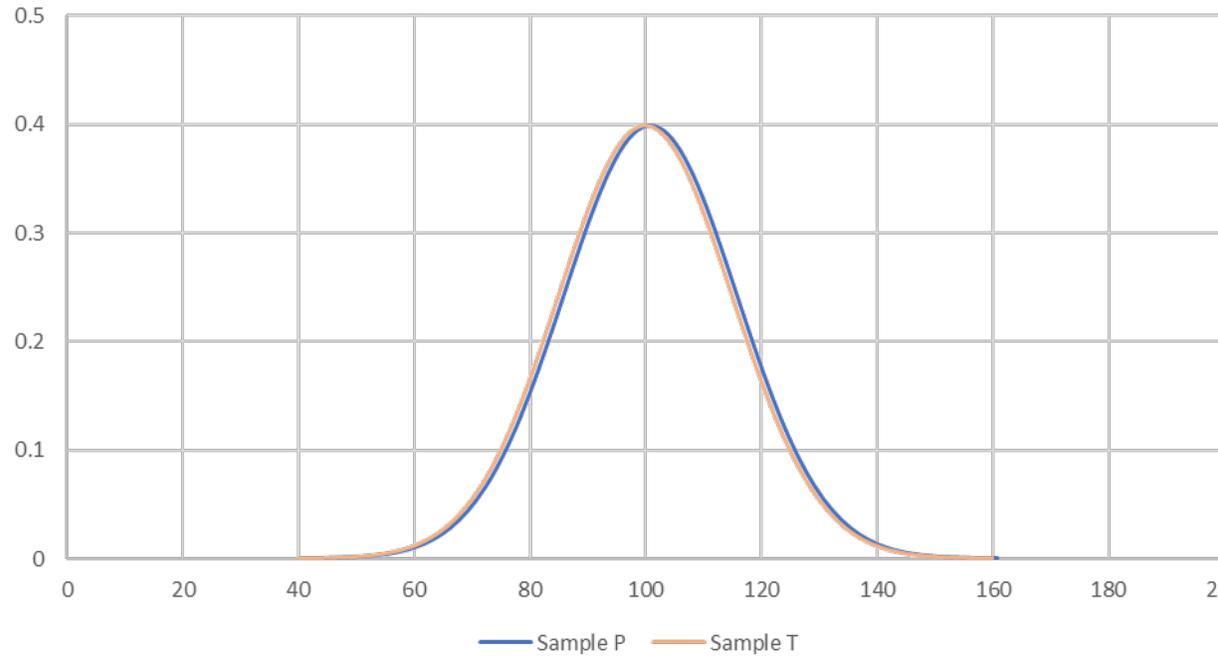

Two Baseline Samples Centered at 100 units — Sample P, Sample T

Two Samples After Treatment - A Subset Differs — Sample P, Sample T-R, Sample T-NR

Alternatively, Treatment May Reduce Symptoms in a Subset

Two Baseline Samples Centered at 100 units
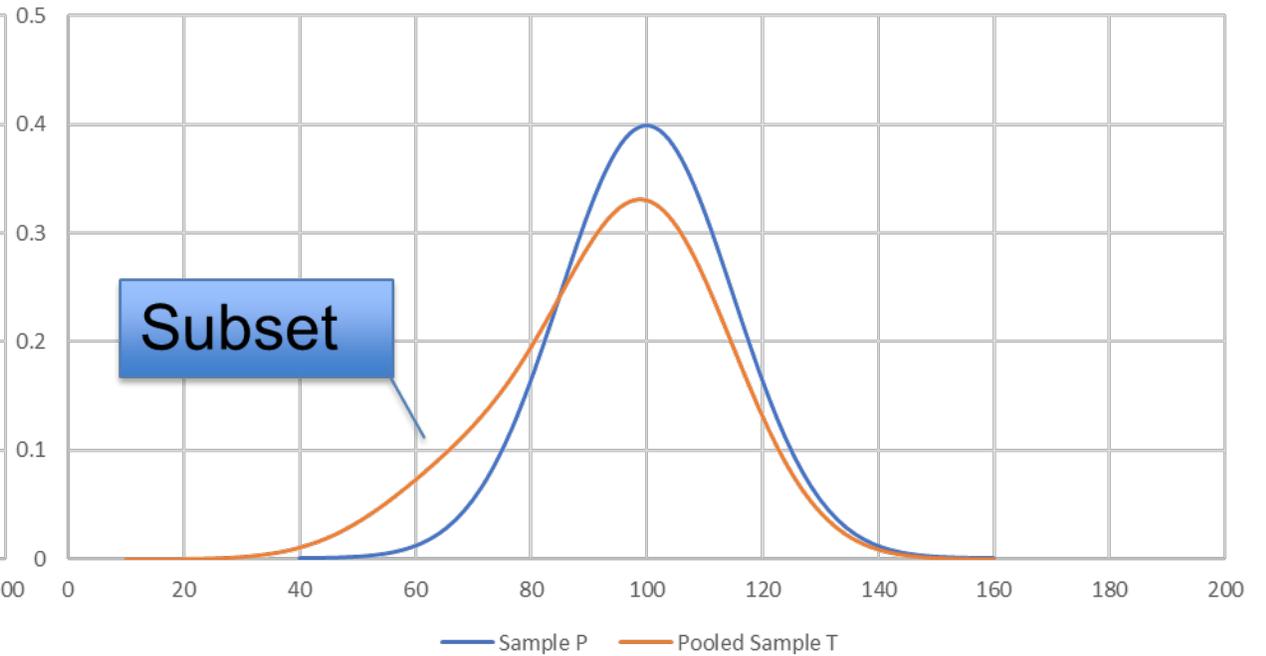
Two Samples After Treatment - A Subset Differs

Subset

Sample P   Sample T
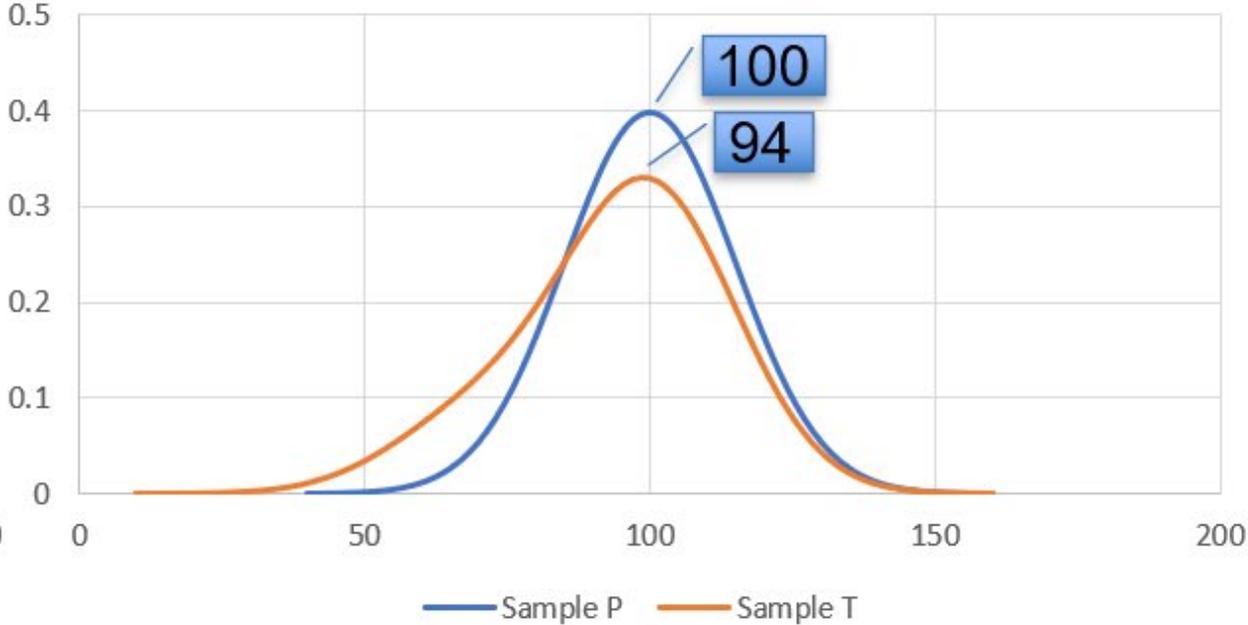
Sample P   Pooled Sample T
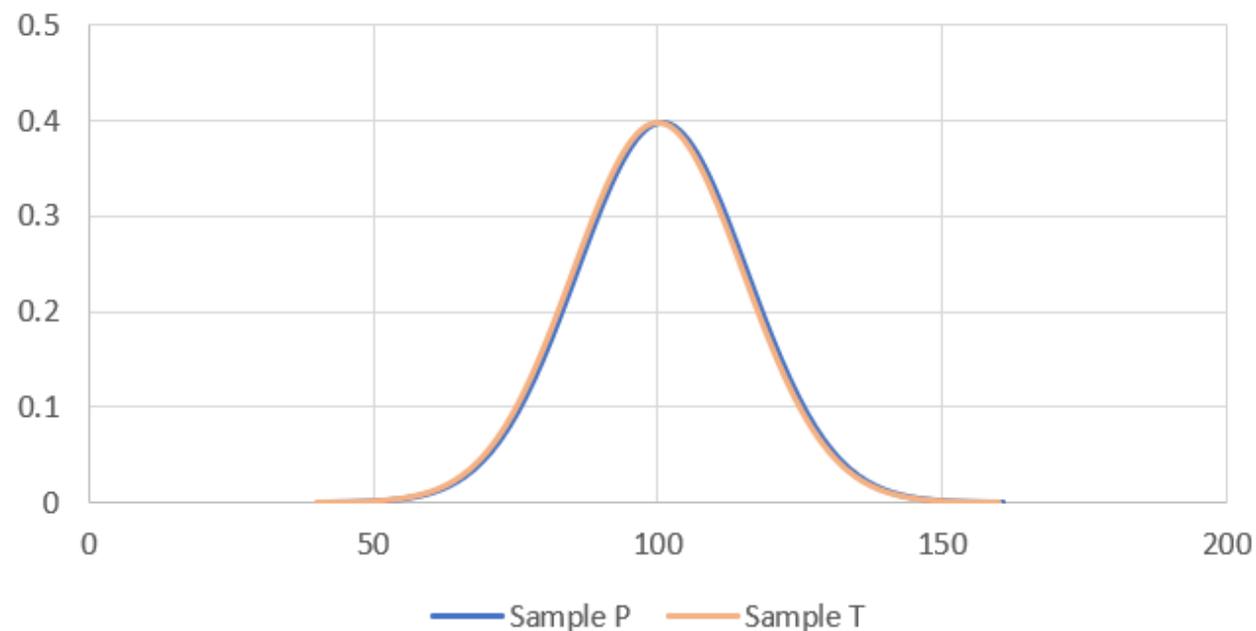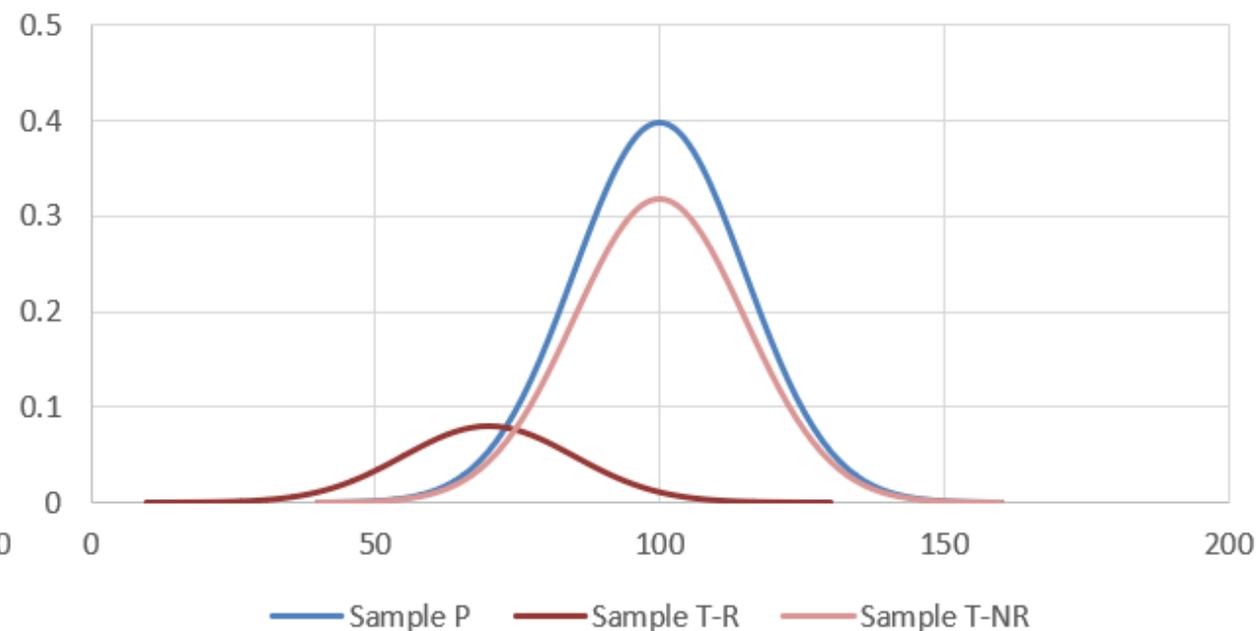
There is apparent separation between the two treatments (e.g., sample means, $\bar{x}$, differ)

Two Baseline Samples Centered at 100 units
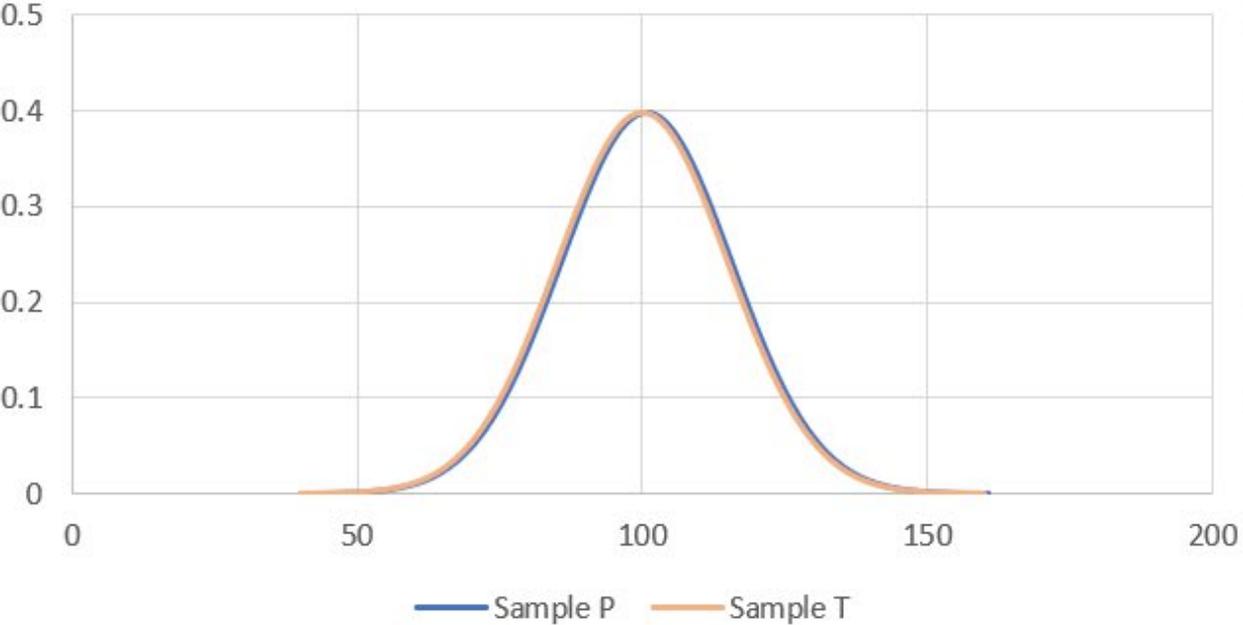
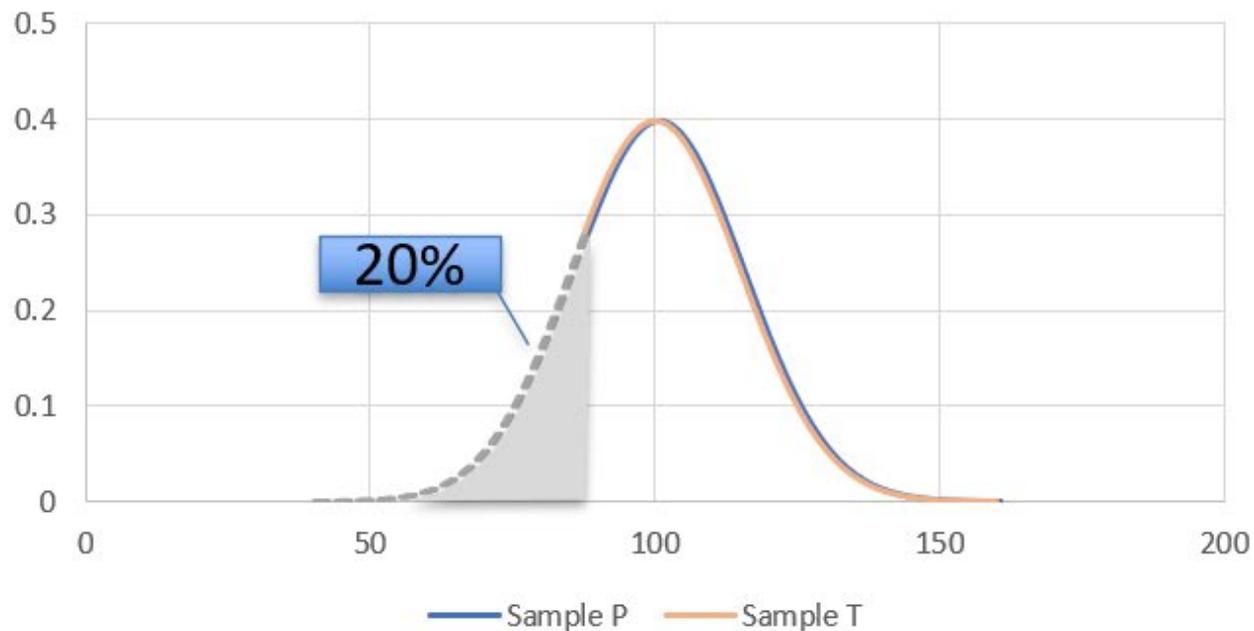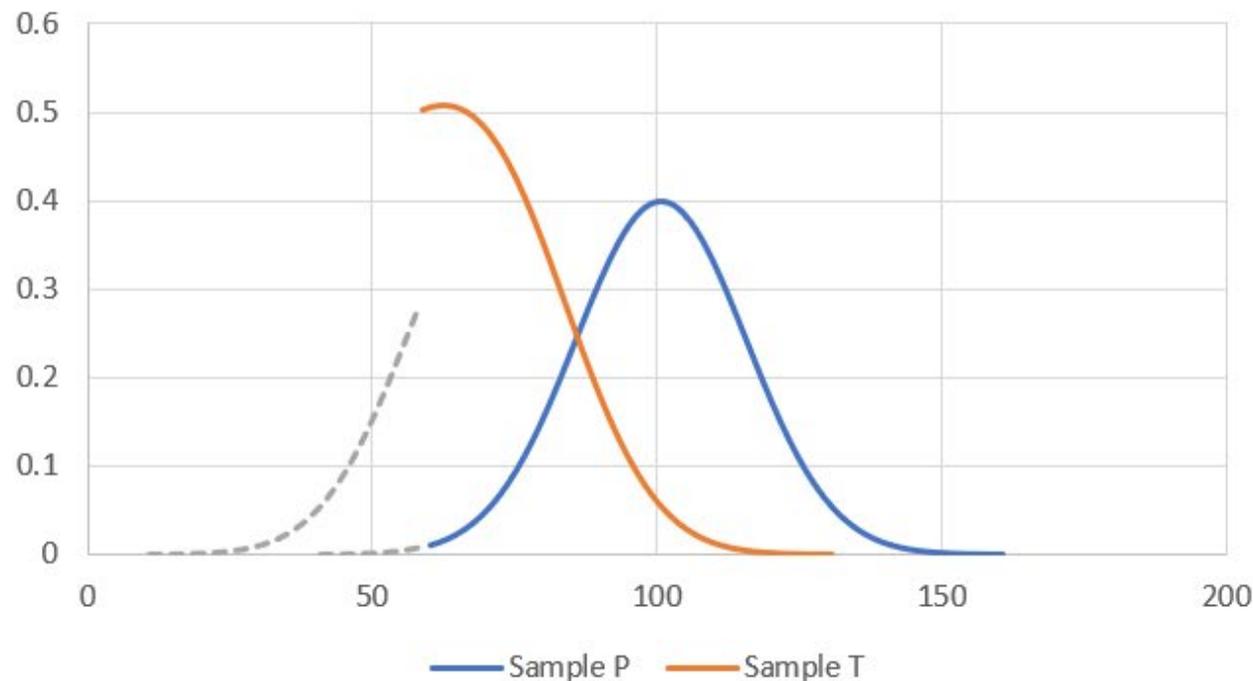Two Samples After Treatment - A Subset Differs

A Subset with "None" or A Subset with a Floor Effect

# Ensuring Patients are Truly Symptomatic at Baseline

- Careful selection of enrollment criteria - multifactorial

- Emphasize qualification at screening

- Consider a "run-in" period

- Randomly assign only symptomatic patients to treatment

- Avoid qualifying patients with the primary assessment

- Avoid selecting the most extreme baseline scores increases misclassification and apparent improvement.
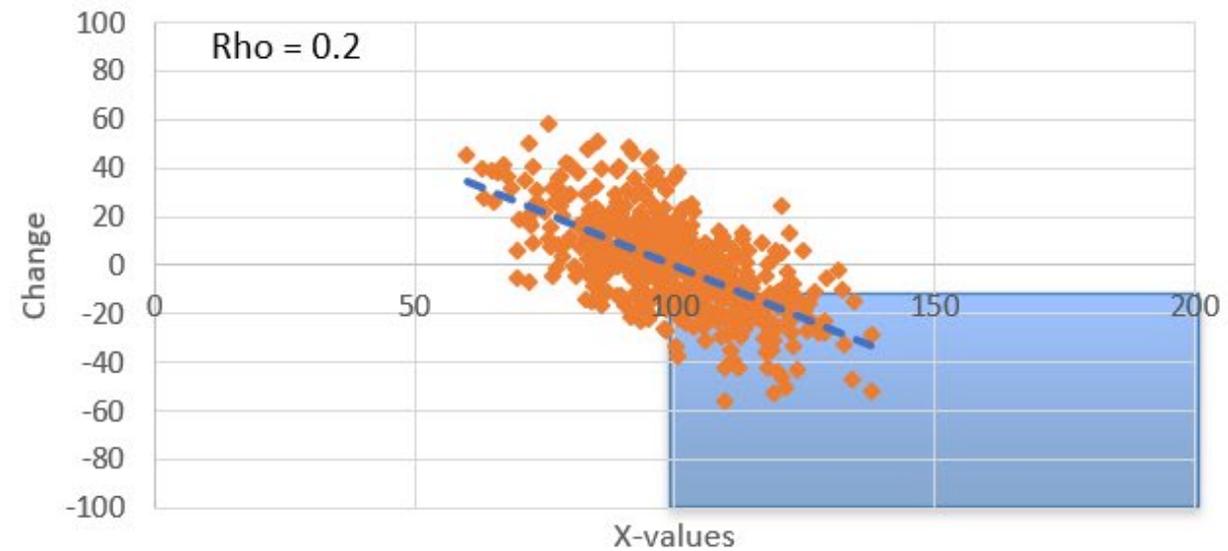
Selecting the most extreme baseline scores increases misclassification and apparent improvement

A Single Sample Before and After Treatement - No Change

Rho = 0.2

A Single Sample Before and After Treatement - No Change

Rho = 0.2
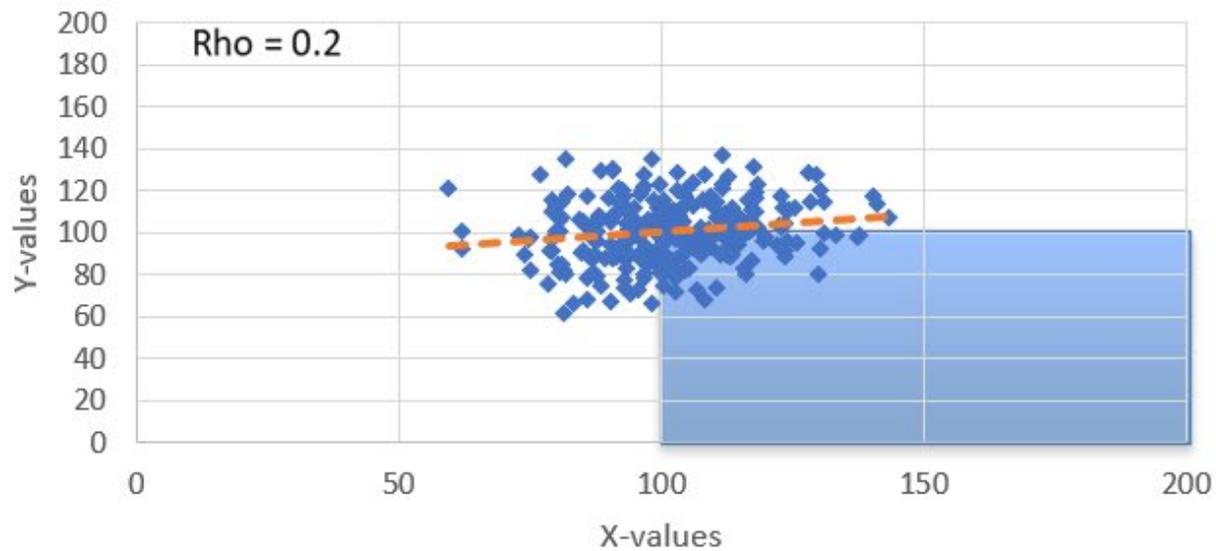
Key Takeaways

- **Floor and ceiling effects can occur at the individual or group-level :**
  - **A patient's COA score is *numerically* at the floor or ceiling when the score is at the lower or upper limit of the possible score range**
  - A patient may be *effectively* at a floor or ceiling if their COA score does not change due to no change in the measured construct
  - *Floor and ceiling effects may occur at the group-level* when a notable proportion of patients have baseline COA *scores at the lower or upper limit of the possible score range*
- **Floor and ceiling effects can hinder quantification of treatment differences by:**
  - **Restricting the range of COA scores** observed in a trial, which can make it harder to detect a *true* treatment effect quantified by endpoint based on those COA scores
  - **Interacting with a notable imbalance between groups at baseline** in the proportion of patients at the floor (or ceiling) which can make the investigational therapy appear *more* <u>or</u> *less* efficacious

45

Thank you for your attention!

What can sponsors do to help set themselves up for success?

FDA

As you design your clinical trial:

Think about *expected and potential <u>interactions</u>* among the following:

- Target patient population
- Investigational therapy's mechanism of action
- Measured construct(s)
- Measures (COAs) and COA score properties (granularity, validity, reliability)
- Disease trajectory over the study period, with respect to measured construct(s) as operationalized by the study COAs
  - With no treatment
  - On standard of care (SOC)
  - On the investigational therapy
- Sample sizes
- Imbalance between arms in sample size and baseline characteristics

For example, you might ask yourself:

- What does the anticipated disease progression look like over time with respect to a given construct of interest?
  - Think about the *timespan* over which change is expected to occur and the anticipated *pace* and *trajectory* of change over that time.
  - If there is an approved therapy or SOC for the disease, how does the approved therapy affect the disease progression with respect to a given construct?
- Do the patients you plan to enroll in your trial have the possibility to numerically *and* physiologically improve or decline over the course of the treatment period in your trial—either on or off treatment?
- Is the drug's anticipated treatment effect similar across patients in the target population, or does it vary across clinical subgroups? Can the subgroups be defined using baseline characteristics?
- Is there any a priori knowledge of how the target patient population for your clinical trial performs on the study COAs?

Panel Discussion
followed by
Q+A

Break

Please return at 3:00 p.m. ET

Challenges and Opportunities for Deriving and Applying Meaningful Score Regions

Overview of Session

Presentation (approx. 50 minutes)

Challenges and Opportunities for Deriving and Applying Meaningful Score Regions
- Monica Morell - Senior Statistician, Division of Biometrics III, Office of Biostatistics, Center for Drug Evaluation and Research, US Food and Drug Administration

Panel Discussion (approx. 50 minutes)
- Jennifer Clark - Lead Mathematical Statistician, Division of Biometrics II, Office of Biostatistics, Center for Drug Evaluation and Research, US Food and Drug Administration
- Ebony Dashiell-Aje - Executive Director, Head, Patient Centered Outcomes Science, BioMarin Pharmaceutical Inc.
- Rikki Mangrum - Distinguished Scientist, Vector Psychometric Group
- Lori McLeod - Vice President of Patient-Centered Outcomes Assessment, RTI Health Solutions
- Jammbe Musoro - Senior Statistician, Statistics and Quality of Life Departments, European Organisation for Research and Treatment

Audience Questions and Answers (approx. 20 minutes)

FDA's patient-focused drug development (PFDD) Guidance 4 Incorporating Clinical Outcome Assessments into Endpoints for Regulatory Decision Making

Putting COA Scores in the Context of Patients' Lives: Score Interpretation Metric

Meaningful Score Regions (MSRs)

What meaningful patient experiences correspond to ranges of COA scores?

COA Scores
- 40 — Can walk up stairs unassisted
- Can walk up stairs with assistance
- 0 — Cannot walk up stairs

Meaningful Score Differences (MSDs)

What size difference on the COA score metric corresponds to a meaningful difference in the patient's experience?

Difference in COA scores
- +10 — A lot better
- A little better
- 0 — No change
- A little worse
- -10 — A lot worse
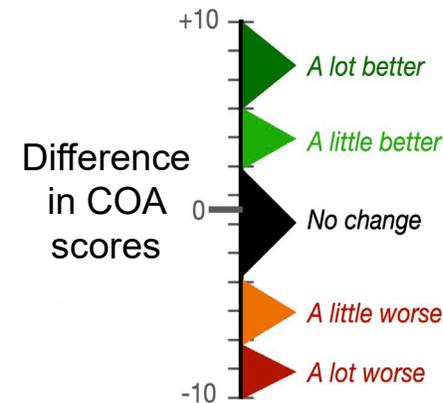
Neither MSR nor MSD approaches are necessarily better than the other
Select approach that best suits the COA, endpoint, and context of use

Creating Score Interpretation Metrics: MSRs and MSDs

- Methods directly consider **patient voice**

    - Quantitative approaches (e.g., anchor-based methods)
    - Qualitative approaches (e.g., cognitive interviews)
    - Mixed-Method approaches (e.g., bookmarking)

- Use of multiple methods

- Distribution-based approaches (e.g., effect size) are <u>not sufficient as the primary method</u>

PFDD Draft Guidance 4

01 Create **score interpretation metric**: meaningful score regions (**MSRs**) & meaningful score differences (**MSDs**)

Anchor-based Methods

Bookmarking

...

02 Apply score interpretation metric: **MSRs & MSDs**

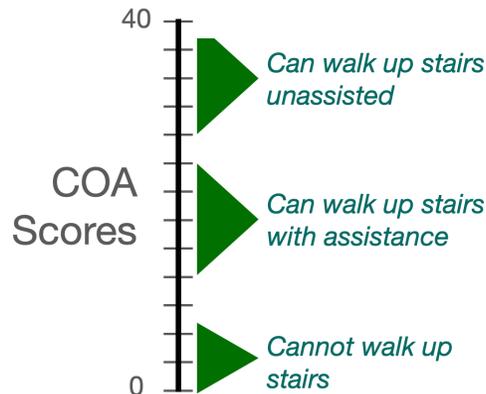To define an endpoint (e.g., a responder endpoint or a time-to-event endpoint)

To interpret a continuous endpoint: "vertical approach" "horizontal approach"

Interpreting the Treatment Effect using the Horizontal Approach and Vertical Approach

Empirical Cumulative Distribution Function (eCDF): COA at Month 12 by Treatment Group

Interpreting the Treatment Effect using the Horizontal Approach and Vertical Approach

Empirical Cumulative Distribution Function (eCDF): COA at Month 12 by Treatment Group

Better

Worse

COA Score at Month 12
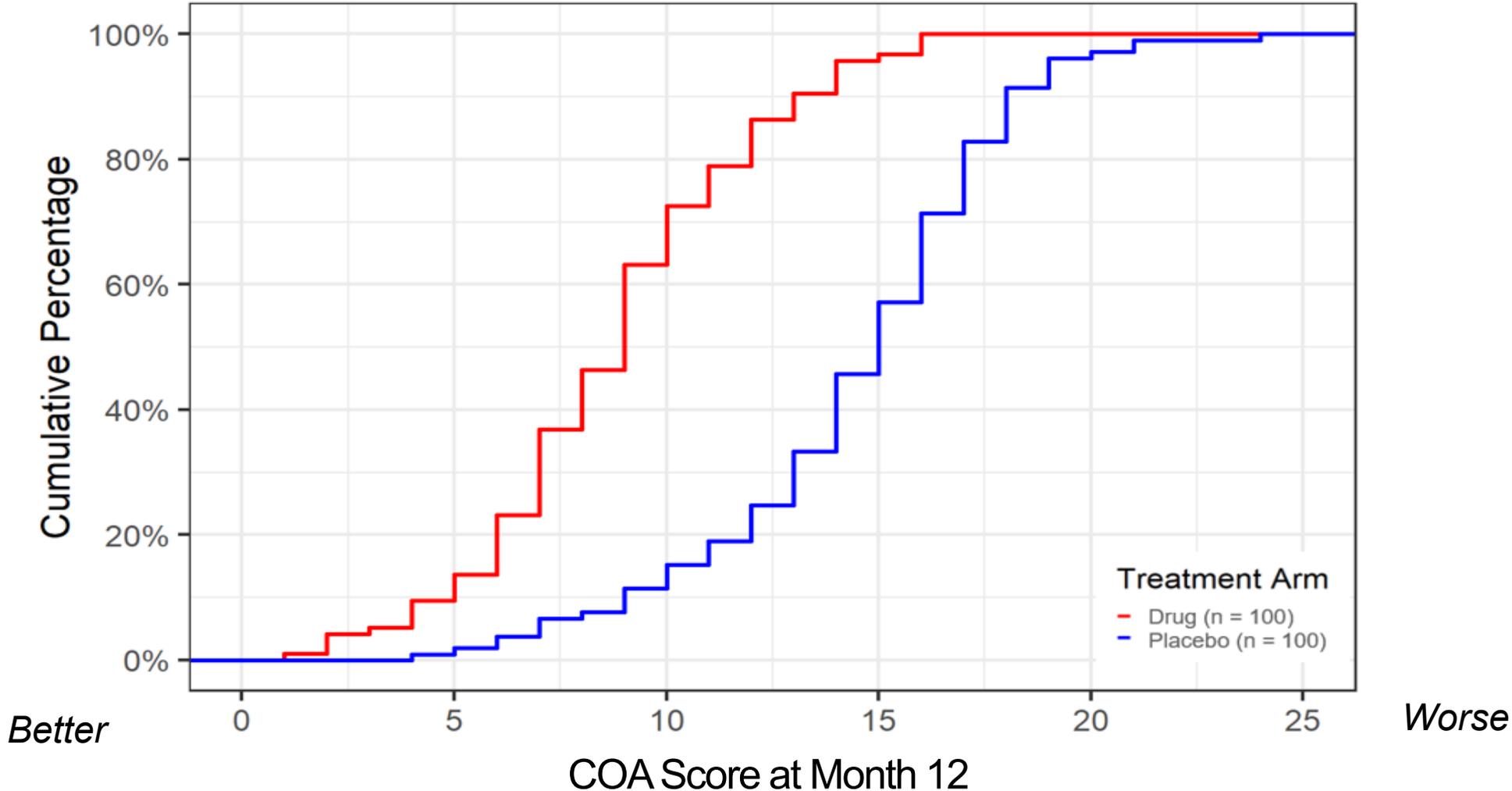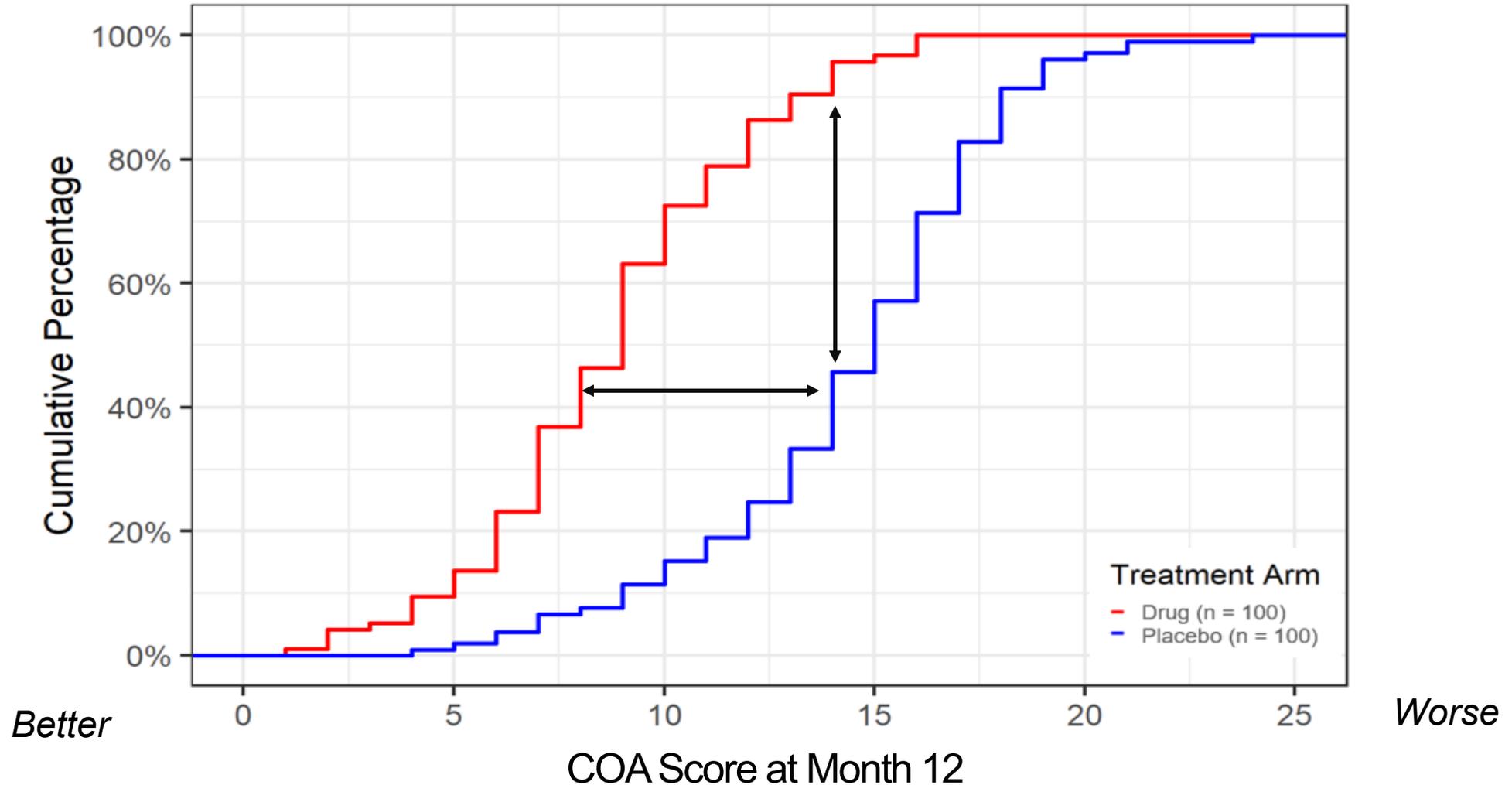
Focus of Comparing Study Groups' eCDFs When Interpreting Treatment Effect Using MSRs

Vertical Gap

Expected difference in probability of achieving COA-based endpoint scores across treatment arms

Helps us answer the following question:

❑ **How much more likely is a patient to experience a meaningful benefit if receiving treatment rather than placebo?**

❑ e.g., Patients are about 35% more likely to have "none" or "mild" symptoms if given treatment rather than placebo.

COA Score at Month 12

Focus of Comparing Study Groups' eCDFs When Interpreting Treatment Effect Using MSRs

FDA

COA Score at Month 12

Average Horizontal Gap
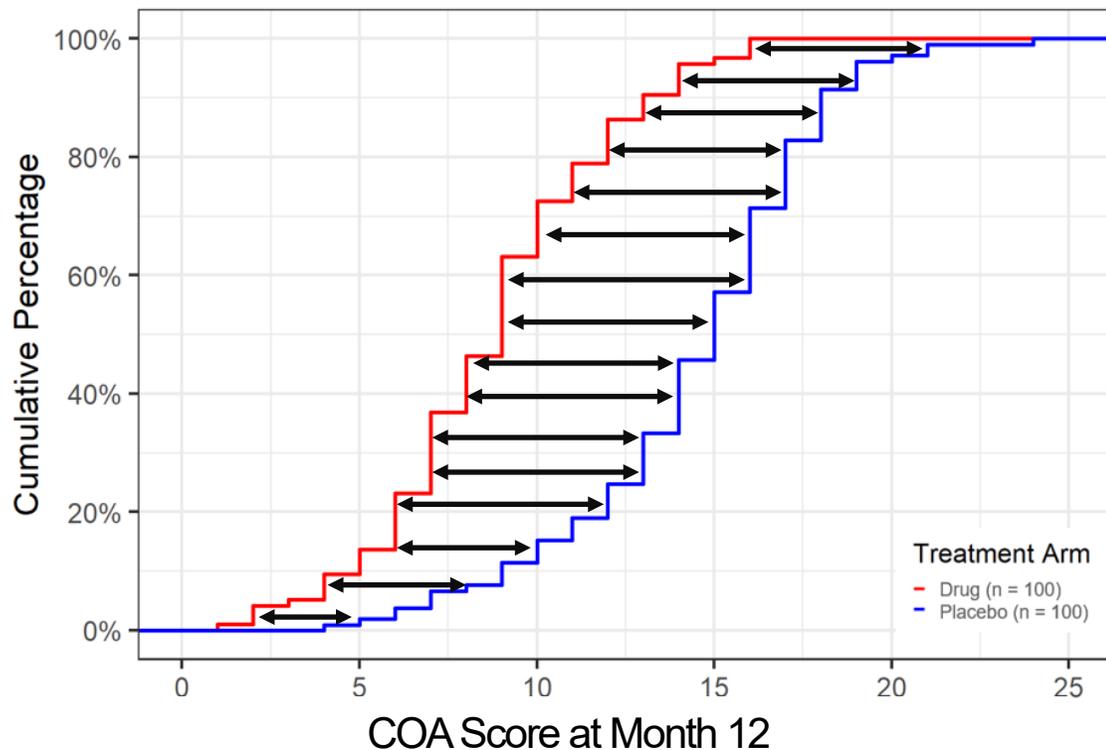
Expected difference in COA-based endpoint score across treatment arms
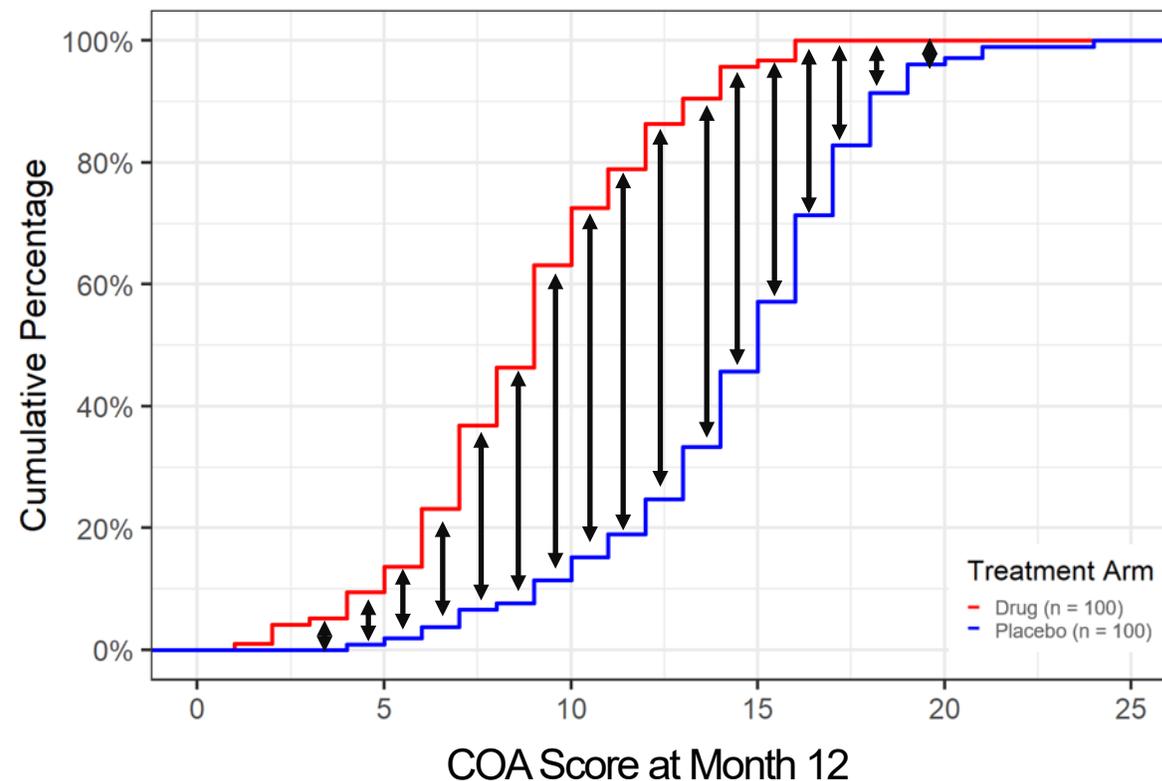
Helps us answer the following question:

- ❑ **How much better is a patient expected to feel if given treatment rather than control?**
- ❑ e.g., On average, patients' symptoms are likely to be "none" on treatment and "moderate" on placebo.

Two Hypothetical Examples

Hypothetical Example 1

- **MSR approach**: Bookmarking
- **Trial Endpoint**: COA score at follow-up timepoint
- **Interpreting trial endpoint results using MSRs**:
  - Vertical Approach
  - Horizontal Approach

Hypothetical Example 2

- **MSR approach**: Anchor-based
- **Trial Endpoint**: Change from Baseline in COA score at follow-up timepoint
- **Interpreting trial endpoint results using MSRs**:
  - Vertical Approach
  - Horizontal Approach

Analyses and results that follow are meant to illustrate key points and do not include all of the analyses that might need to be done for a real submission.

Hypothetical Example # 1 (Bookmarking Method)

Hypothetical Example 1

FAKE DATA

FDA

Meaningful Score Regions Derivation Study

- N = 80

- COA: Patient-reported *XYZ Symptom Index*

  - Scores 0 (better) to 24 (worst)

- Method: Bookmarking (mixed-methods approach)

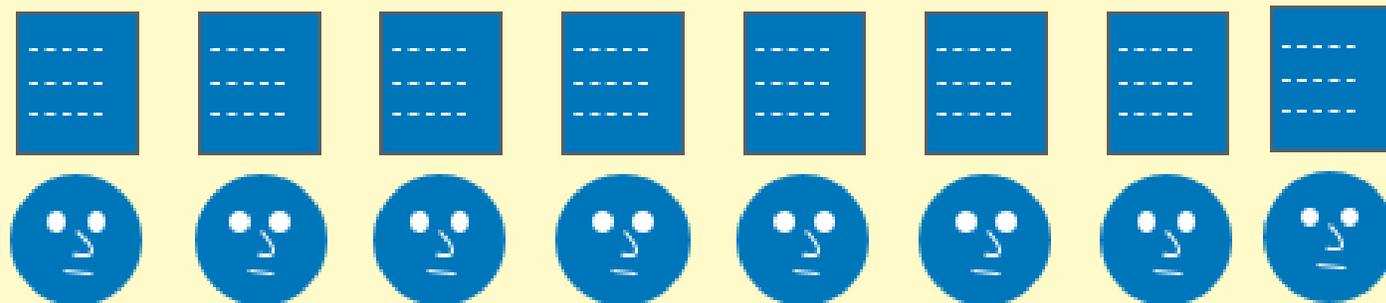- XYZ Symptom Index has a 7-day recall period

Meaningful Score Regions Approach
Bookmarking Method

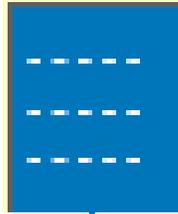COA Score — *Better* | | | | | | | | | | | | | | | | | | | | | | | *Worse*

Patient Experience

Meaningful Score Regions Approach
Bookmarking Method

In the last 7 days, Anne never got tired easily and never had trouble starting things because she was too tired. She felt that she was never so tired that it was hard for her to focus on work. She never found it hard to get out of bed in the morning because she was too tired and she never felt too tired to do sports or exercise.

In summary, Anne reports:

- Never getting tired easily.
- She never had trouble starting things because she was too tired.
- Never feeling so tired that it was hard for her to focus on work.
- It was never hard to get out of bed in the morning because of being too tired.
- Never feeling too tired to do sports or exercise.
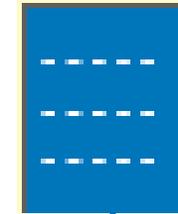
In the last 7 days, Jessica sometimes was too tired to do things outside. Sometimes she had trouble finishing things because she was too tired, and being tired sometimes kept her from having fun. She was also sometimes too tired to read. She never felt too tired to eat.

In summary, Jessica reports:

- Sometimes she was too tired to do things outside.
- Sometimes she had trouble finishing things because she was too tired.
- Being tired sometimes kept her from having fun.
- Sometimes she felt too tired to read.
- She never felt too tired to eat.

Example from Morgan, E. M., Mara, C. A., Huang, B., Barnett, K., Carle, A. C., Farrell, J. E., & Cook, K. F. (2017). Establishing clinical meaning and defining important differences for Patient-Reported Outcomes Measurement Information System (PROMIS®) measures in juvenile idiopathic arthritis using standard setting with patients, parents, and providers. *Quality of Life Research : an international journal of quality of life aspects of treatment, care and rehabilitation*, 26(3), 565–586. https://doi.org/10.1007/s11136-016-1468-2

Meaningful Score Regions Approach
Bookmarking Method

COA Score — Better — Worse

Patient Experience

Patient Judgment of Experience

None · **Mild** · **Moderate** · **Severe**

Approximate Meaningful Score Regions

None | Mild | Moderate | Severe

Hypothetical Example: Approximate MSRs

Approximate Meaningful Score Regions | None | Mild | Moderate | Severe

0 2 4 6 8 10 12 14 16 18 20 22 24

XYZ Symptom Index Score
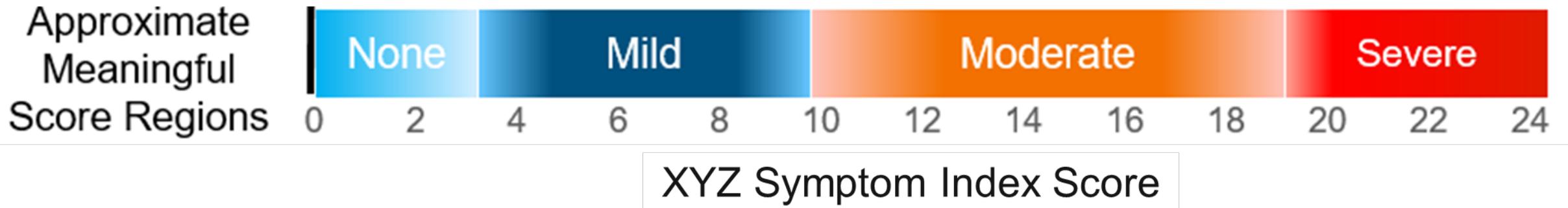
Hypothetical Example 1 (continued)

FAKE DATA

FDA

Randomized Clinical Trial (RCT) of Drug vs Placebo

- Parallel groups design

- N = 250

- XYZ Symptom Index assessed at baseline and 12 months post-randomization

- Endpoint: *XYZ Symptom Index score* at Month 12

- Primary analysis: Comparison of study group mean *XYZ Symptom Index* scores at *Month 12* using analysis of covariance (ANCOVA) with baseline *XYZ Symptom Index* score as covariate

Qualitative Data from Patients

Approximate Meaningful Score Regions: None | Mild | Moderate | Severe
0 2 4 6 8 10 12 14 16 18 20 22 24

XYZ Symptom Index Score

Meaningful benefit?

Two scenarios:
- Scenario A
- Scenario B

Qualitative Data from Patients: Scenario A

Approximate Meaningful Score Regions

| None | Mild | Moderate | Severe |

0 2 4 6 8 10 12 14 16 18 20 22 24

XYZ Symptom Index Score
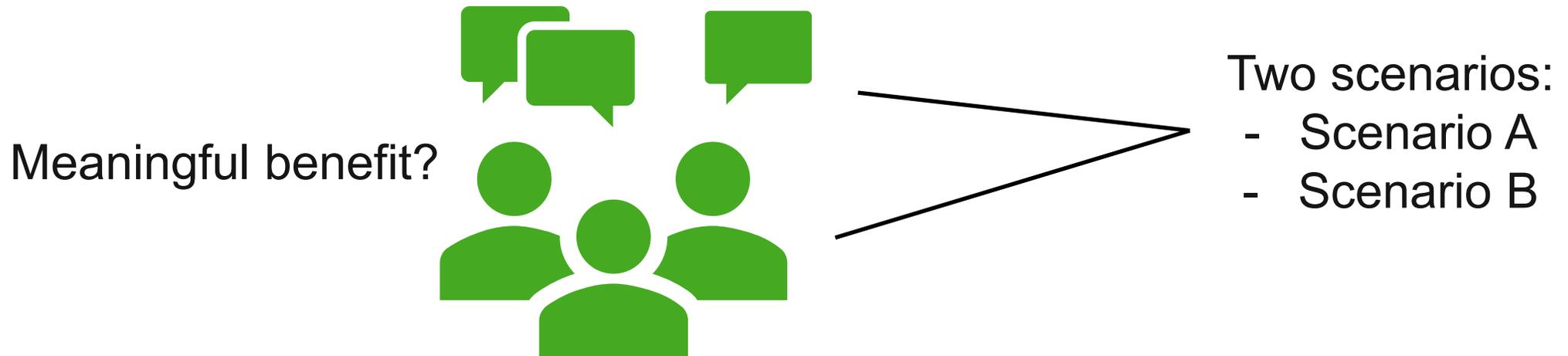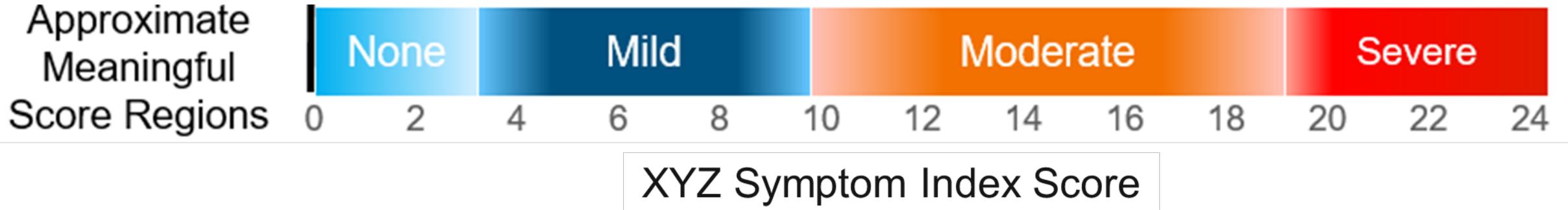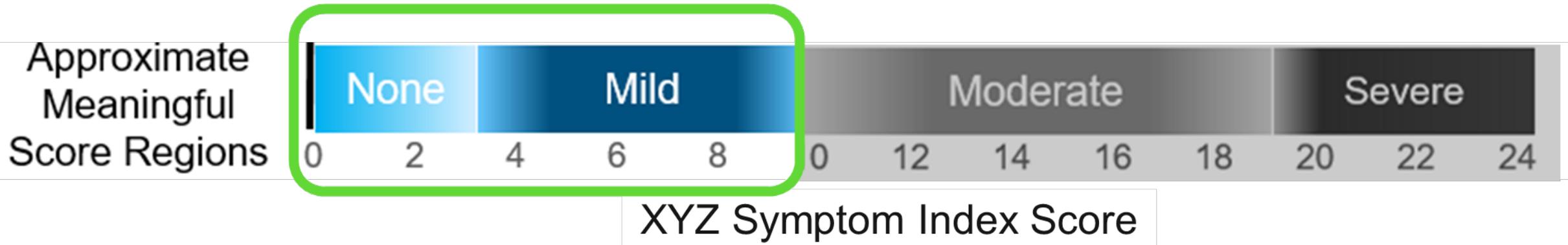
- Patients report having *no* or *mild* XYZ Symptoms* at Month 12 would be a meaningful benefit

*These are meaningful in this hypothetical example. In some cases, some other region(s) may be important.

Qualitative Data from Patients: Scenario B
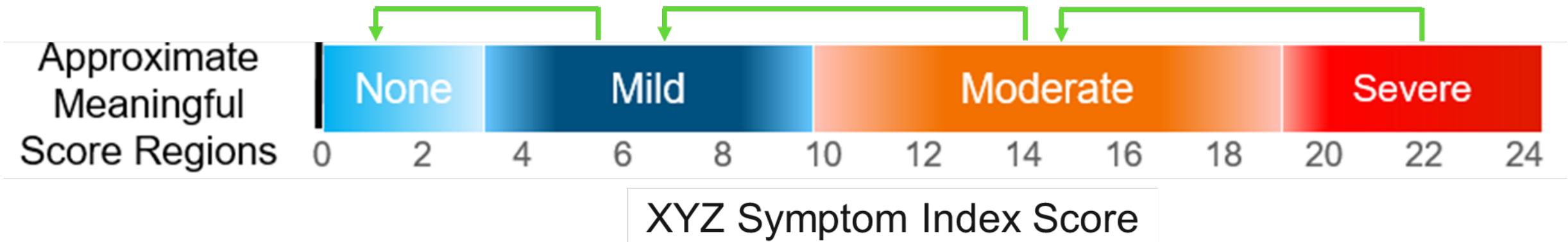
- Patients report at least a 1-MSR* improvement (e.g., *moderate* to *mild*) in XYZ Symptom Index Scores would be a meaningful benefit

*These are meaningful in this hypothetical example. In some cases, some difference(s) may be important.

Hypothetical Example 1: Efficacy Results

FDA

- **Endpoint**: *XYZ Symptom Index score* at Month 12

Treatment Group	LS Mean	SE	95% CI
Placebo	14.4	0.17	[14.0, 14.7]
Drug	8.8	0.18	[8.4, 9.1]
LS Mean Difference	5.6	0.24	[5.2, 6.1]

Results obtained from an ANCOVA model with covariates treatment arm and baseline COA score.
Outcome variable is Month 12 COA score.

Interpreting Trial Results:
Scenario A

Expected difference in the probability of having Month 12 XYZ Symptom Index score in the regions of *<u>none</u>* or *<u>mild</u>*

Difference in Cumulative Percentage between Study Groups as a Function of XYZ Symptom Index Score at Month 12
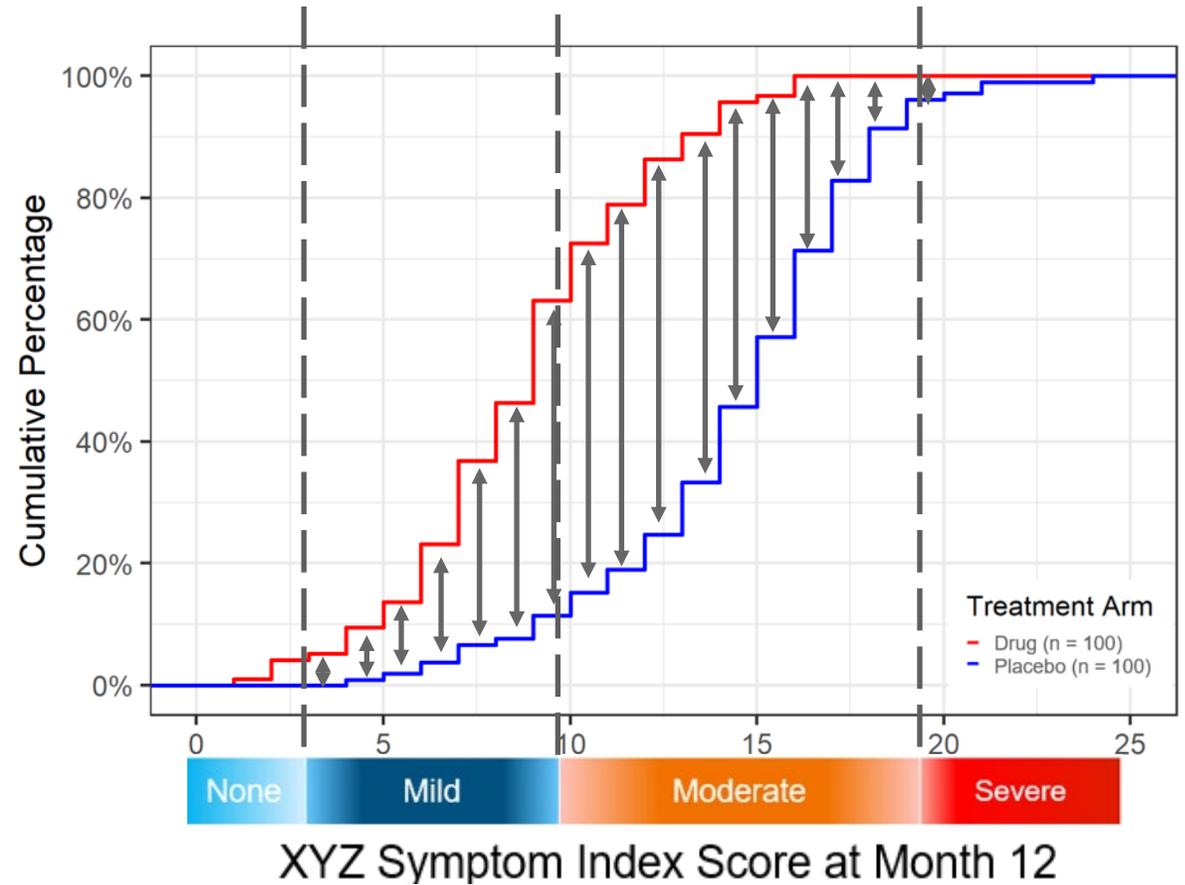
Probability of Patients' XYZ Symptom Index Scores Falling in each MSR at Month 12

Based on qualitative interviews with patients, having **none** or **mild** symptoms* at Month 12 represents a meaningful benefit to patients. *Scenario A*

Meaningful Score Region	Treatment Arm	
	Treatment n (p)	Placebo n (p)
None	6 (.06)	0
Mild	64 (.67)	16 (.15)
Moderate	26 (.27)	85 (.81)
Severe	0	4 (.04)

- The probability of patients having XYZ Symptom Index scores in the *none* or *mild* regions at Month 12 on treatment is 0.73
- The probability of patients having XYZ Symptom Index scores in the none or mild regions at Month 12 on placebo is 0.15

*These are the regions that are meaningful in this hypothetical example. In other cases, it may be different region(s).

Interpreting Trial Results:
Scenario B

Expected difference in Month 12 XYZ Symptom Index scores (i.e., MSR) across treatment arms

Difference between Study Groups in Expected XYZ Symptom Index Score at Month 12
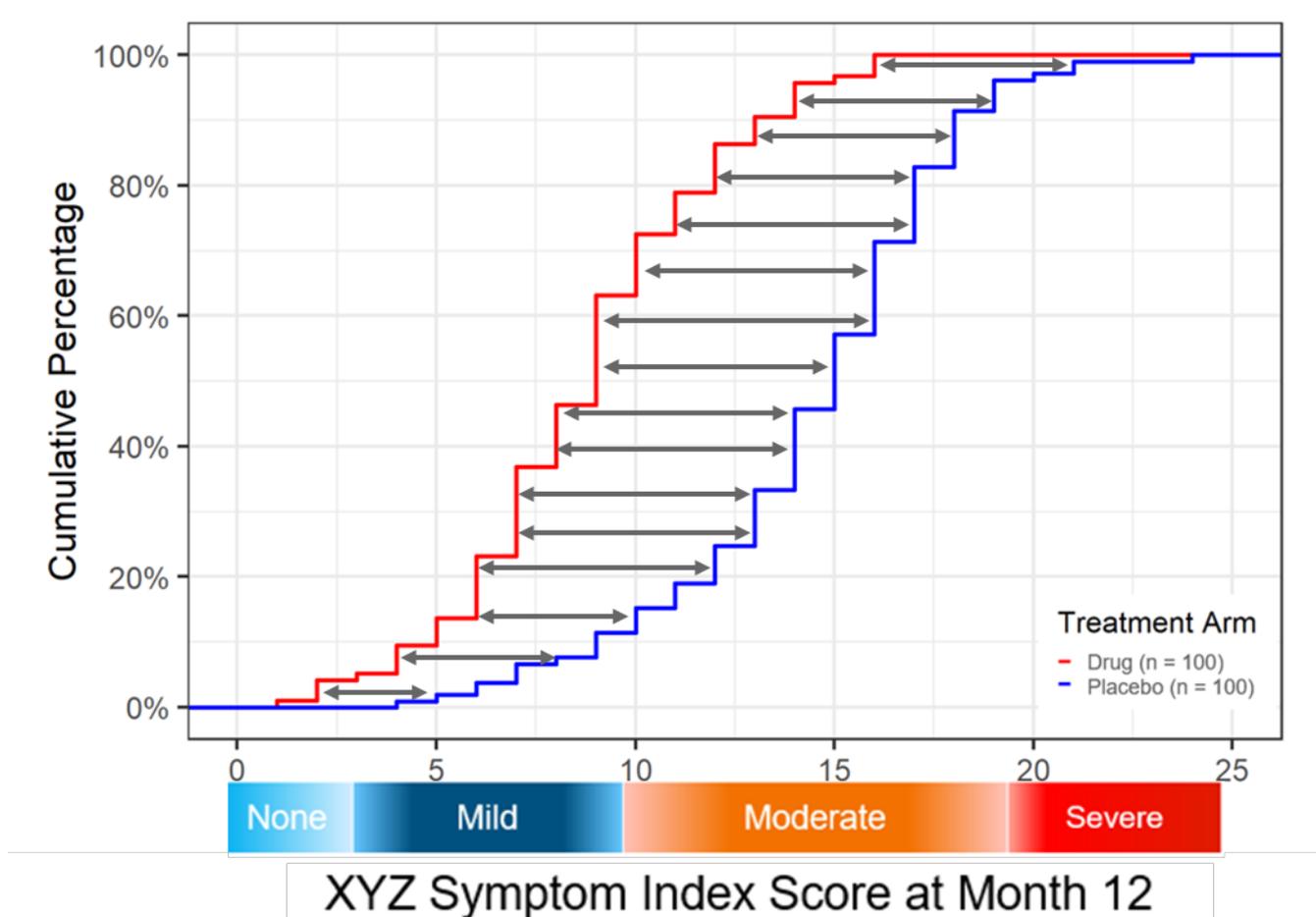
Expected Difference in XYZ Symptom Index Score at Month 12
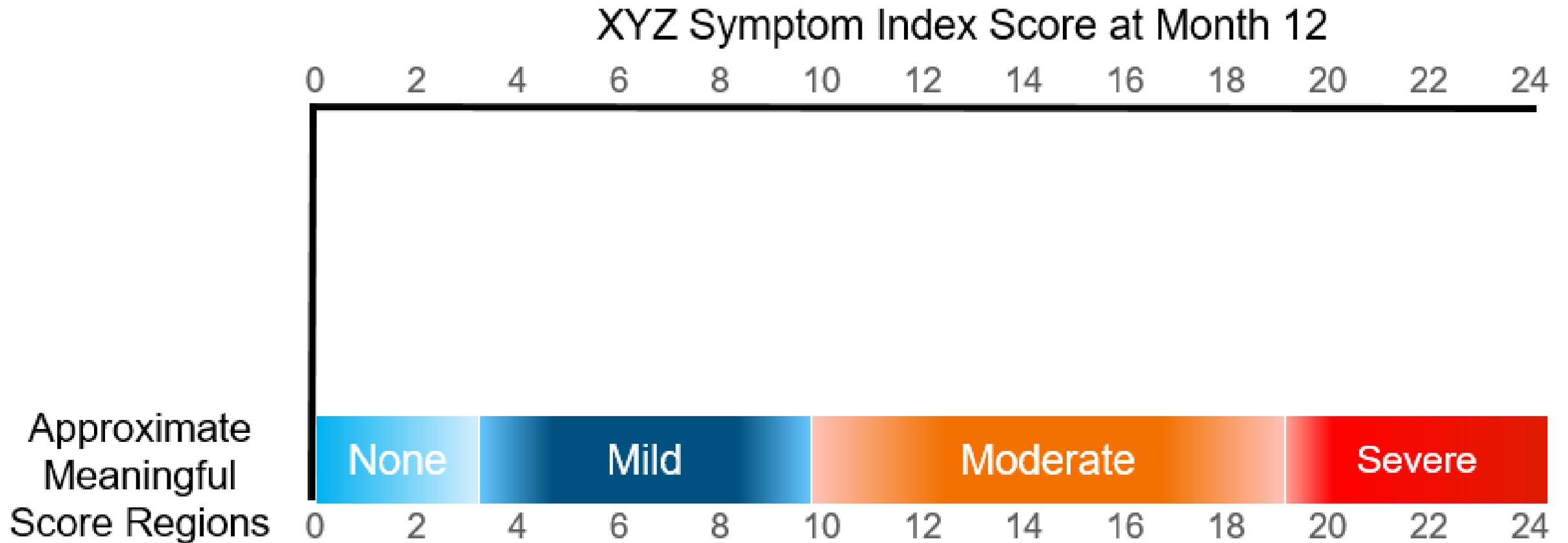
XYZ Symptom Index Score at Month 12

0 2 4 6 8 10 12 14 16 18 20 22 24

Approximate Meaningful Score Regions

None | Mild | Moderate | Severe

0 2 4 6 8 10 12 14 16 18 20 22 24

*This is the difference that is meaningful in this hypothetical example; in some cases, a 2-region or other difference may be important.

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

XYZ Symptom Index Score at Month 12

Treatment effect = 5.6 (95% CI: 5.2, 6.1)

LS Means (95% CIs) at 12 Months

Drug
Placebo

Approximate Meaningful Score Regions

None | Mild | Moderate | Severe

Based on qualitative interviews with patients, a difference of at least 1-region* represents a meaningful difference in severity. *Scenario B*

*This is the difference that is meaningful in this hypothetical example; in some cases, a 2-region or other difference may be important.

82
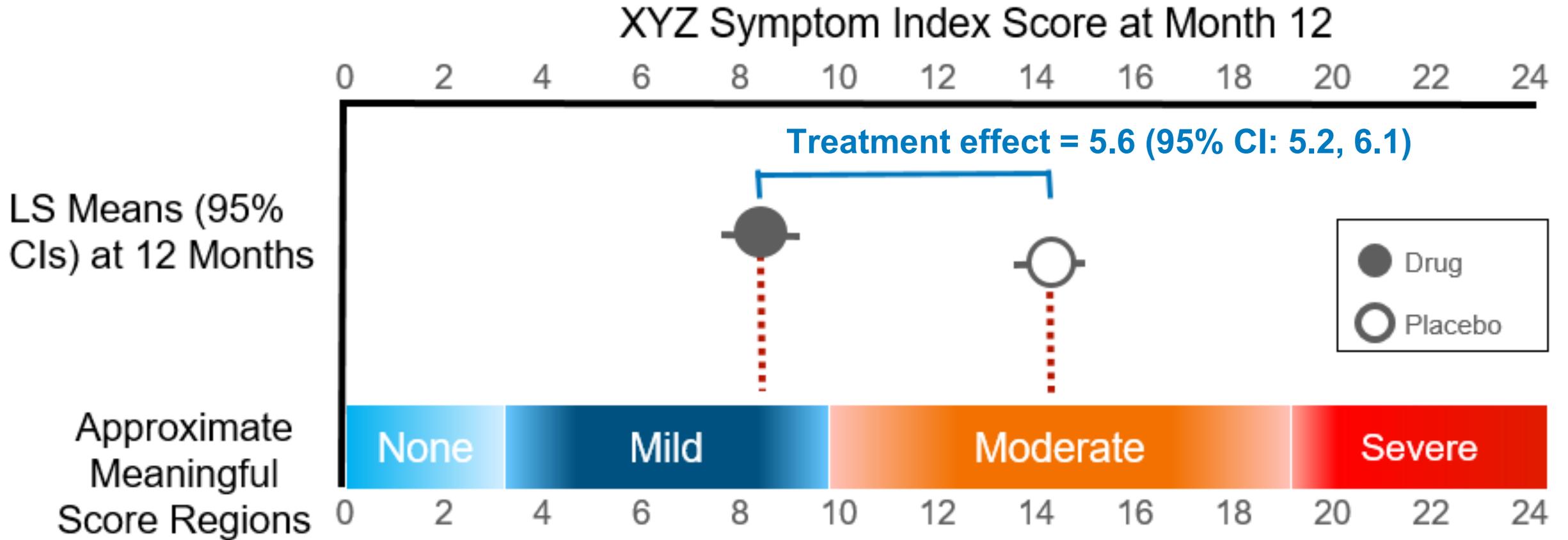
Hypothetical Example 1: Summary

Endpoint: *XYZ Symptom Index score* at Month 12

- MSR approach: Bookmarking

- LS Mean Difference: 5.6 (95% CI: 5.2, 6.1)

Hypothetical Example 1: Summary

FDA

> **Vertical Approach**
>
> **Expected difference in probability of having Month 12 XYZ Symptom Index scores in MSRs of *none* or *mild***

*These are meaningful in this hypothetical example. In some cases, some other region(s) or difference may be important.

Hypothetical Example 1: Summary

FDA

Vertical Approach
Expected difference in probability of having Month 12 XYZ Symptom Index scores in MSRs of *none* or *mild*
Scenario A: Qualitative data supports having ***none*** or ***mild*** XYZ symptoms* at Month 12 is meaningful to patients

*These are meaningful in this hypothetical example. In some cases, some other region(s) or difference may be important.

Hypothetical Example 1: Summary

Vertical Approach **Expected difference in probability of having Month 12 XYZ Symptom Index scores in MSRs of *none* or *mild***
Scenario A: Qualitative data supports having *none* or *mild* XYZ symptoms***** at Month 12 is meaningful to patients
Probability that patients in the treatment group had XYZ Symptom Index scores at Month 12 corresponding to an experience of *none* or *mild* was .**73** Probability that patients in the placebo group had XYZ Symptom Index scores at Month 12 corresponding to an experience of *none* or *mild* was **.15**

*****These are meaningful in this hypothetical example. In some cases, some other region(s) or difference may be important.

Hypothetical Example 1: Summary

FDA

Vertical Approach
Expected difference in probability of having Month 12 XYZ Symptom Index scores in MSRs of *none* or *mild*
Scenario A: Qualitative data supports having ***none*** or ***mild*** XYZ symptoms* at Month 12 is meaningful to patients
Probability that patients in the treatment group had XYZ Symptom Index scores at Month 12 corresponding to an experience of *none* or *mild* was .**73** Probability that patients in the placebo group had XYZ Symptom Index scores at Month 12 corresponding to an experience of *none* or *mild* was **.15**
Patients' XYZ Symptoms are more likely to be *none* or *mild* (reported as meaningful by patients) on treatment than on placebo

*These are meaningful in this hypothetical example. In some cases, some other region(s) or difference may be important.

Hypothetical Example 1: Summary

Vertical Approach	Horizontal Approach
Expected difference in probability of having Month 12 XYZ Symptom Index scores in MSRs of *none* or *mild*	**Expected difference in Month 12 XYZ Symptom Index Scores (i.e., MSR)**
Scenario A: Qualitative data supports having *none* or *mild* XYZ symptoms* at Month 12 is meaningful to patients	
Probability that patients in the treatment group had XYZ Symptom Index scores at Month 12 corresponding to an experience of *none* or *mild* was .**73** Probability that patients in the placebo group had XYZ Symptom Index scores at Month 12 corresponding to an experience of *none* or *mild* was .**15**	
Patients' XYZ Symptoms are more likely to be *none* or *mild* (reported as meaningful by patients) on treatment than on placebo	

*These are meaningful in this hypothetical example. In some cases, some other region(s) or difference may be important.

Hypothetical Example 1: Summary

Vertical Approach	Horizontal Approach
Expected difference in probability of having Month 12 XYZ Symptom Index scores in MSRs of *none* or *mild*	**Expected difference in Month 12 XYZ Symptom Index Scores (i.e., MSR)**
Scenario A: Qualitative data supports having ***none*** or ***mild*** XYZ symptoms* at Month 12 is meaningful to patients	**Scenario B**: Qualitative data supports at least ***1-MSR difference**** is meaningful to patients
Probability that patients in the treatment group had XYZ Symptom Index scores at Month 12 corresponding to an experience of *none* or *mild* was *.73* Probability that patients in the placebo group had XYZ Symptom Index scores at Month 12 corresponding to an experience of *none* or *mild* was *.15*	
Patients' XYZ Symptoms are more likely to be *none* or *mild* (reported as meaningful by patients) on treatment than on placebo	

*These are meaningful in this hypothetical example. In some cases, some other region(s) or difference may be important.

Hypothetical Example 1: Summary

Vertical Approach	Horizontal Approach
Expected difference in probability of having Month 12 XYZ Symptom Index scores in MSRs of *none* or *mild*	**Expected difference in Month 12 XYZ Symptom Index Scores (i.e., MSR)**
Scenario A: Qualitative data supports having *none* or *mild* XYZ symptoms* at Month 12 is meaningful to patients	**Scenario B**: Qualitative data supports at least *1-MSR difference** is meaningful to patients
Probability that patients in the treatment group had XYZ Symptom Index scores at Month 12 corresponding to an experience of *none* or *mild* was *.73* Probability that patients in the placebo group had XYZ Symptom Index scores at Month 12 corresponding to an experience of *none* or *mild* was *.15*	On average, a patient in the treatment group is expected to achieve an XYZ Symptom Index score at Month 12 corresponding to an experience of *mild* On average, a patient in the placebo group is expected to achieve an XYZ Symptom Index score at Month 12 corresponding to an experience of *moderate*
Patients' XYZ Symptoms are more likely to be *none* or *mild* (reported as meaningful by patients) on treatment than on placebo	

*These are meaningful in this hypothetical example. In some cases, some other region(s) or difference may be important.

Hypothetical Example 1: Summary

FDA

Vertical Approach	Horizontal Approach
Expected difference in probability of having Month 12 XYZ Symptom Index scores in MSRs of *none* or *mild*	**Expected difference in Month 12 XYZ Symptom Index Scores (i.e., MSR)**
Scenario A: Qualitative data supports having *none* or *mild* XYZ symptoms* at Month 12 is meaningful to patients	**Scenario B**: Qualitative data supports at least *1-MSR difference** is meaningful to patients
Probability that patients in the treatment group had XYZ Symptom Index scores at Month 12 corresponding to an experience of *none* or *mild* was *.73* Probability that patients in the placebo group had XYZ Symptom Index scores at Month 12 corresponding to an experience of *none* or *mild* was *.15*	On average, a patient in the treatment group is expected to achieve an XYZ Symptom Index score at Month 12 corresponding to an experience of *mild* On average, a patient in the placebo group is expected to achieve an XYZ Symptom Index score at Month 12 corresponding to an experience of *moderate*
Patients' XYZ Symptoms are more likely to be *none* or *mild* (reported as meaningful by patients) on treatment than on placebo	Patients' XYZ Symptoms are likely to be one MSR better (reported as meaningful by patients) on treatment than on placebo

*These are meaningful in this hypothetical example. In some cases, some other region(s) or difference may be important.

Hypothetical Example # 2 (Anchor-Based Method)

Hypothetical Example # 2

FAKE DATA

Meaningful Score Regions Derivation Study

- N = 180

- COA: Patient-reported *XYZ Symptom Index*

  - Scores 0 (better) to 24 (worst)

- Method: Anchor-based

- *XYZ Symptom Index* has a 14-day recall period

- Patient Global Impression of Severity (PGIS) has a 14-day recall period

  - *None, mild, moderate, severe*

- Administered at baseline and at Month 12

Distribution of XYZ Symptom Index Scores by PGIS at Month 12
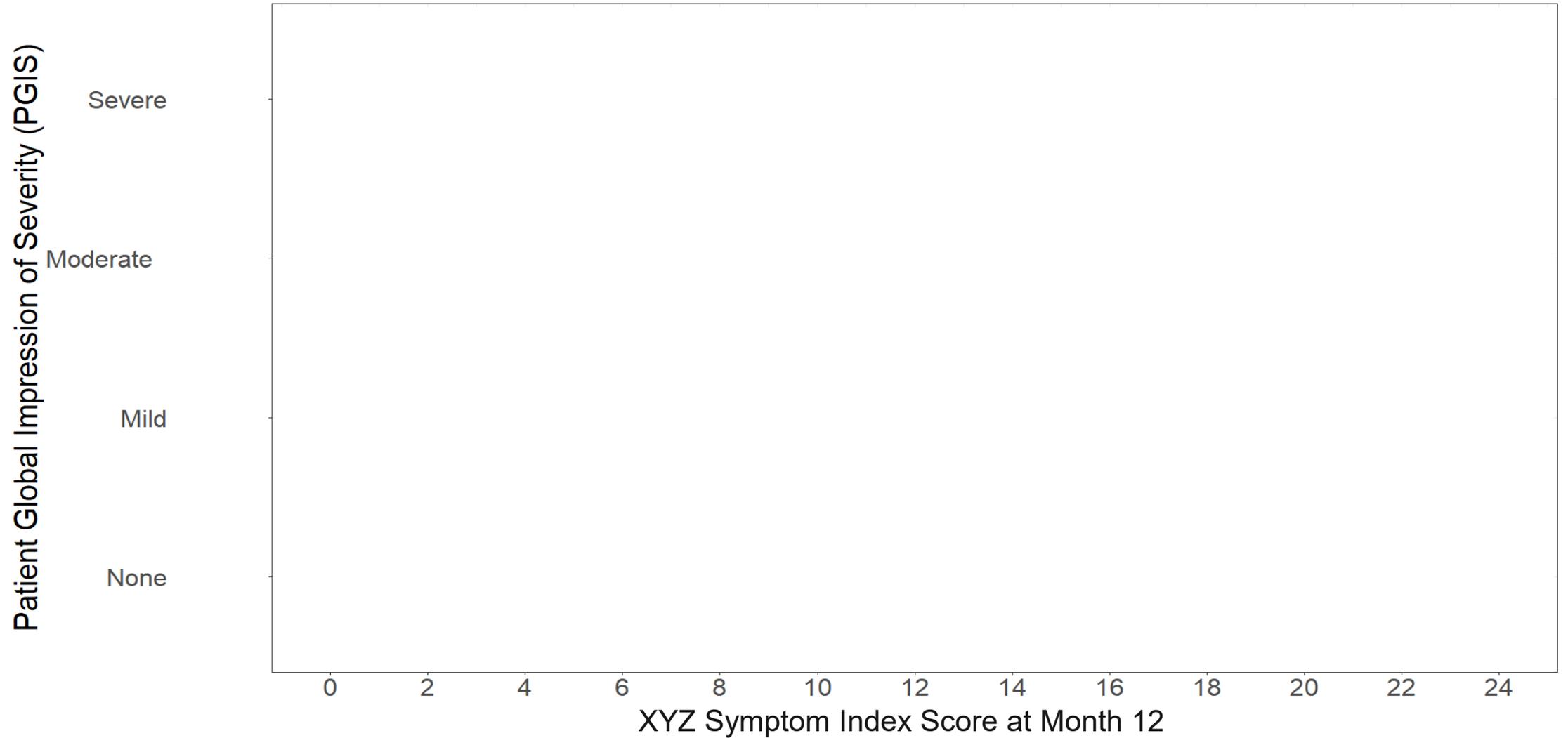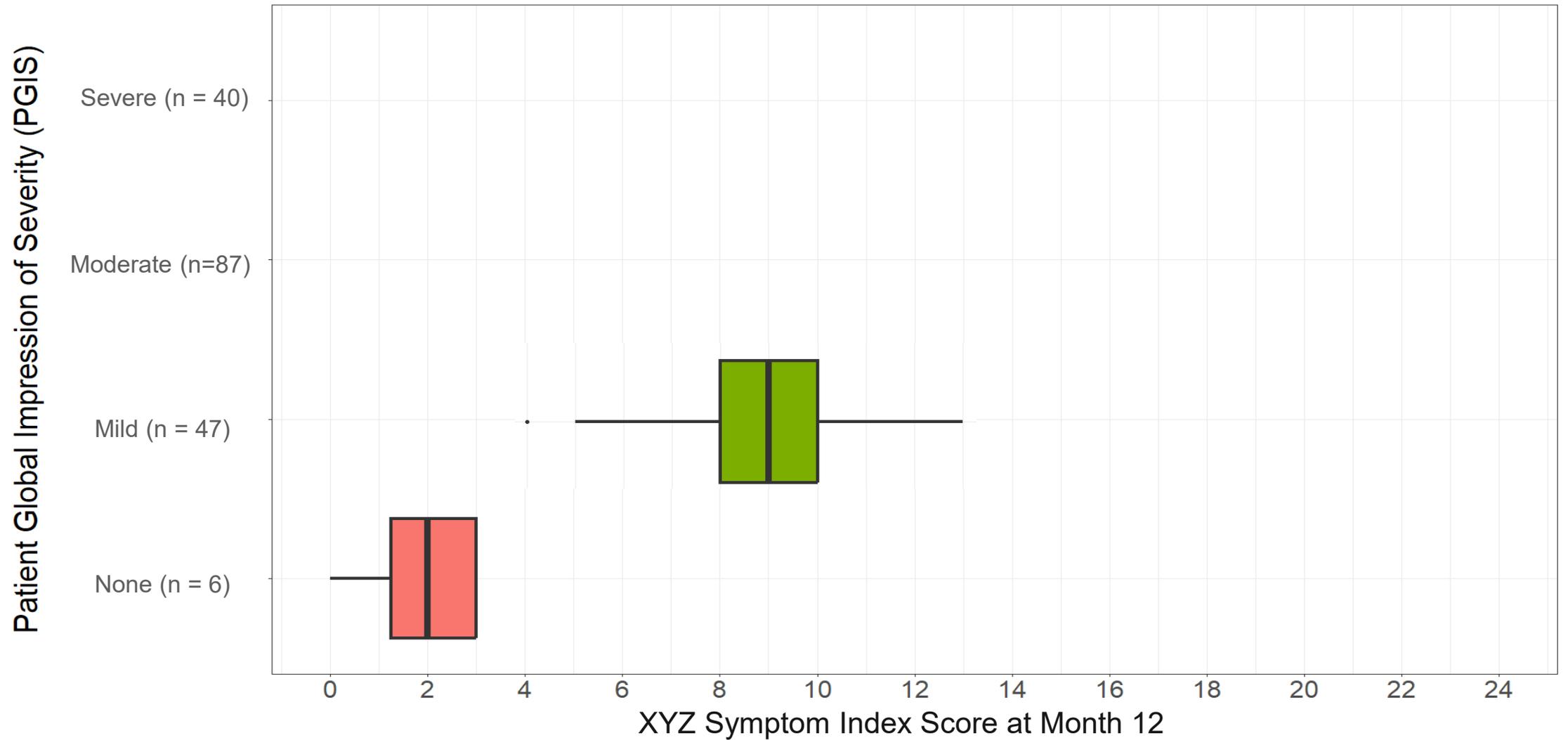
Distribution of XYZ Symptom Index Scores by PGIS at Month 12

Distribution of XYZ Symptom Index Scores by PGIS at Month 12
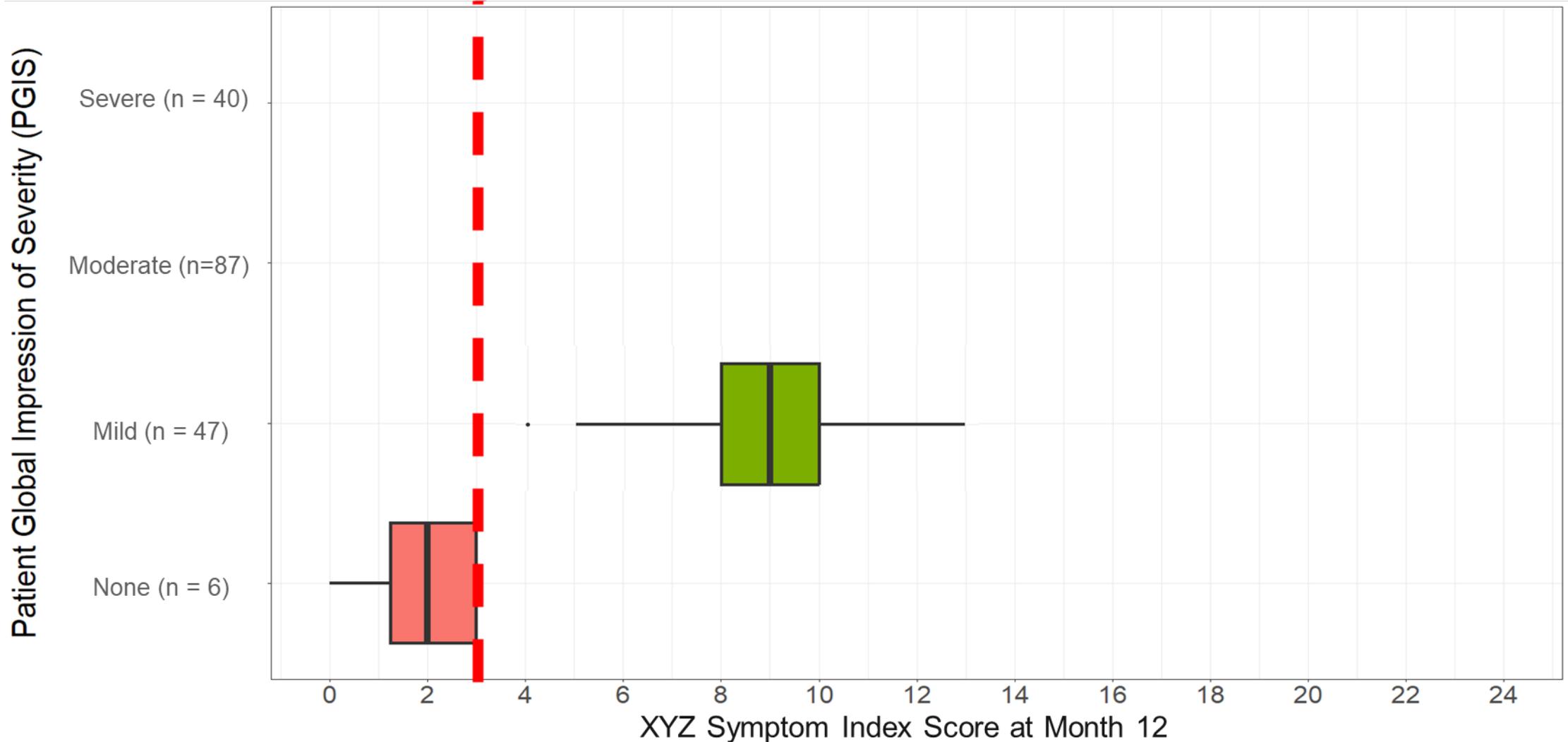
Distribution of XYZ Symptom Index Scores by PGIS at Month 12
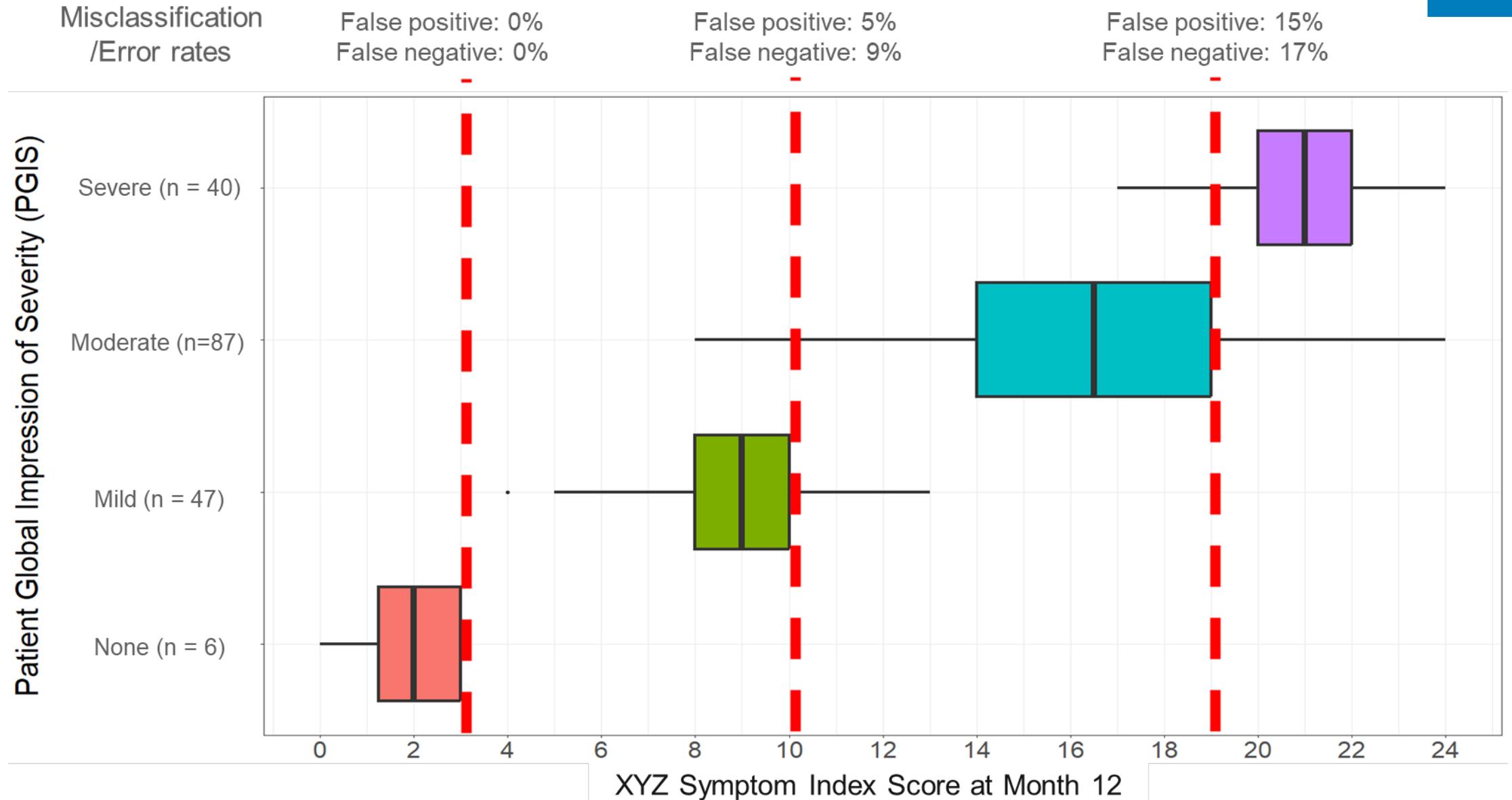
Hypothetical Example #2: Approximate MSRs

Hypothetical Example # 2 (continued)

FAKE DATA

Randomized Clinical Trial of Drug vs Placebo

- Parallel groups design

- N = 200

- XYZ Symptom Index assessed at baseline and 12 months post-randomization

- Endpoint: Change from baseline to Month 12 in *XYZ Symptom Index score*

- Primary analysis: Comparison of study group mean *change from baseline to Month 12 in XYZ Symptom Index* scores using analysis of covariance (ANCOVA) with baseline *XYZ Symptom Index* score as covariate

Qualitative Data from Patients: Scenario B

Approximate Meaningful Score Regions

| None | Mild | Moderate | Severe |

0 2 4 6 8 10 12 14 16 18 20 22 24

XYZ Symptom Index Score
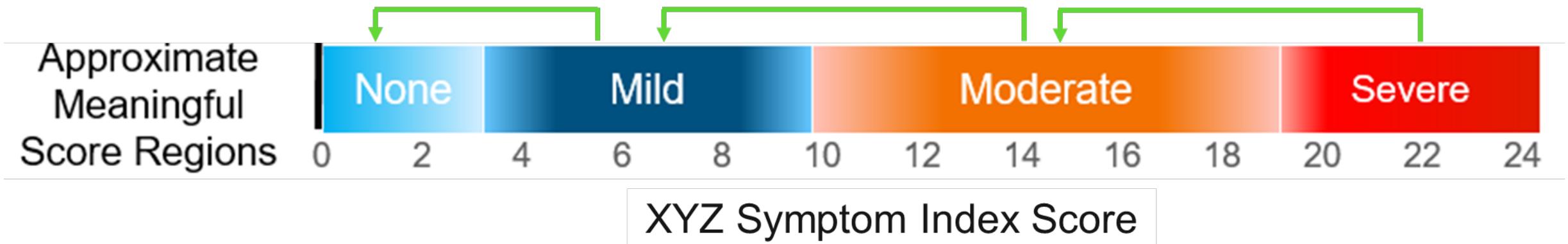
- Patients report at least a 1-MSR* improvement (e.g., *moderate* to *mild*) in XYZ Symptom Index Scores from baseline to month 12 would be a meaningful benefit: Scenario B

*These are meaningful in this hypothetical example. In some cases, other difference(s) may be important.

Hypothetical Example #2: Efficacy Results

- **Trial Endpoint**: Change from baseline to Month 12 in *XYZ Symptom Index score*

Treatment Group	LS Mean	SE	95% CI
Placebo	- 0.6	0.17	[0.2, 0.9]
Drug	5.1	0.18	[4.7, 5.4]
LS Mean Difference	5.6	0.24	[5.2, 6.1]

Results obtained from an ANCOVA model with covariates treatment arm and baseline COA score; Outcome variable is change from baseline to Month 12 in COA score.

Interpreting Trial Results:
Scenario B

Expected difference in the probability of achieving an improvement of ≥ 1 MSR between Baseline and Month 12

Probability of Patients having XYZ Symptom Index Scores Improved by 1 MSR or more from Baseline to Month 12

Overall	Treatment n (p)	Placebo n (p)
	59 (.59)	12 (.12)
Baseline MSR	Treatment n (p)	Placebo n (p)
None	0 (0)	0 (0)
Mild	5 (.14)	1 (.05)
Moderate	45 (.71)	5 (.08)
Severe	9 (1.00)	6 (.60)

- Probability of patients in treatment group improved by at least one MSR is 0.59 compared to 0.12 for patients in placebo group

Interpreting Trial Results:
Scenario B

Expected difference in Month 12 XYZ Symptom Index scores (i.e., MSR) across treatment arms

Expected difference in Month 12 XYZ Symptom Index scores (i.e., MSR) across treatment arms

- Hypothetical Example 1:
  - **Endpoint**: *XYZ Symptom Index score* at Month 12

- Hypothetical Example 2:
  - **Endpoint**: Change from Baseline to Month 12 in *XYZ Symptom Index score*

Hypothetical Example 2

- **Trial Endpoint**: Change from baseline to Month 12 in *XYZ Symptom Index score*

Treatment Group	LS Mean	SE	95% CI
Placebo	0.6	0.17	[0.2, 0.9]
Drug	5.1	0.18	[4.7, 5.4]
LS Mean Difference	5.6	0.24	[5.2, 6.1]

Results obtained from an ANCOVA model with covariates treatment arm and baseline COA score;
Outcome variable is change from baseline to Month 12 in COA score.

- **To facilitate interpretation**: *XYZ Symptom Index score* at Month 12
  - **Placebo**:   LS Mean = 14.4, 95% CI [14.0, 14.7]
  - **Drug**:   LS Mean = 8.8, 95% CI[8.4, 9.1]

Expected Difference in XYZ Symptom Index Score at Month 12
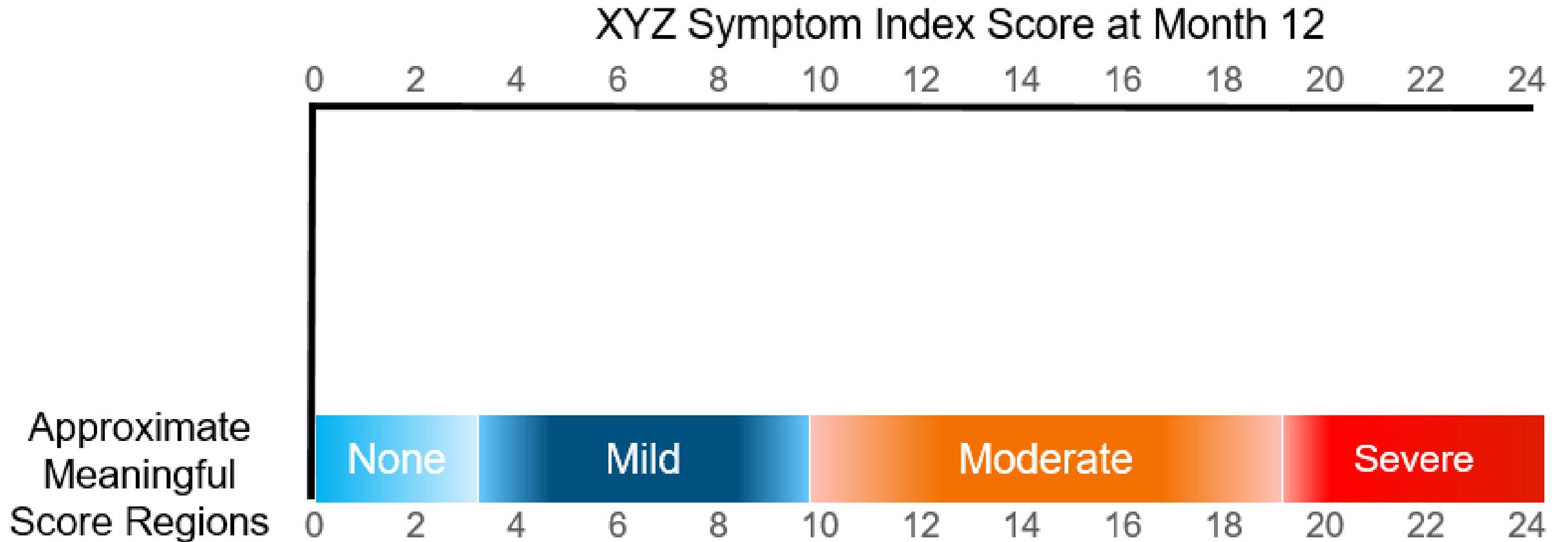
How much better is the average patient's XYZ Symptoms likely to be if they receive Drug rather than Placebo?

XYZ Symptom Index Score at Month 12

0 2 4 6 8 10 12 14 16 18 20 22 24

Treatment effect = 5.6 (95% CI: 5.2, 6.1)

LS Means (95% CIs) at 12 Months

● Drug
○ Placebo

Approximate Meaningful Score Regions

| None | Mild | Moderate | Severe |

0 2 4 6 8 10 12 14 16 18 20 22 24

Based on qualitative interviews with patients, a difference of at least 1-region* represents a meaningful difference in severity. *Scenario B*

*This is the difference that is meaningful in this hypothetical example; in some cases, a 2-region or other difference may be important.

Hypothetical Example 2: Summary

Endpoint: Change from Baseline to Month 12 in *XYZ Symptom Index score*

- MSR approach: Anchor-based
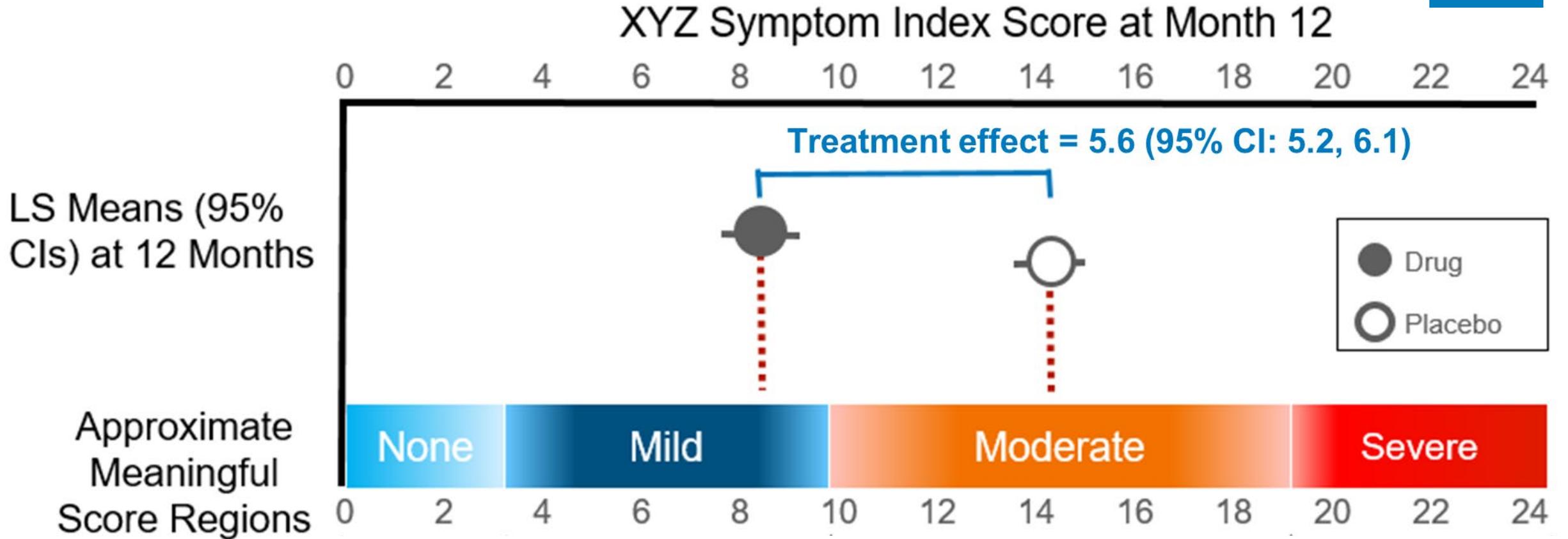
- LS Mean Difference: 5.6 (95% CI: 5.2, 6.1)

Hypothetical Example 2: Summary

FDA

Vertical Approach

Expected difference in probability of having XYZ Symptom Index scores improve by ≥ 1 MSRs from Baseline to Month 12

*These are meaningful in this hypothetical example. In some cases, some other region(s) or difference may be important.

Hypothetical Example 2: Summary

FDA

Vertical Approach
Expected difference in probability of having XYZ Symptom Index scores improve by ≥ 1 MSRs from Baseline to Month 12
Scenario B: Qualitative data supports *1-MSR difference** or more is meaningful to patients

*These are meaningful in this hypothetical example. In some cases, some other region(s) or difference may be important.

Hypothetical Example 2: Summary

FDA

Vertical Approach
Expected difference in probability of having XYZ Symptom Index scores improve by ≥ 1 MSRs from Baseline to Month 12
Scenario B: Qualitative data supports *1-MSR difference*** or more is meaningful to patients
Probability that patients in the treatment group improved in XYZ Symptom Index scores by 1 MSR or more from Baseline to Month 12 is **0.59** Probability that patients in the placebo group improved in XYZ Symptom Index scores by 1 MSR or more from Baseline to Month 12 is **0.12**

*These are meaningful in this hypothetical example. In some cases, some other region(s) or difference may be important.

Hypothetical Example 2: Summary

Vertical Approach
Expected difference in probability of having XYZ Symptom Index scores improve by ≥ 1 MSRs from Baseline to Month 12
Scenario B: Qualitative data supports *1-MSR difference** or more is meaningful to patients
Probability that patients in the treatment group improved in XYZ Symptom Index scores by 1 MSR or more from Baseline to Month 12 is **.59**   Probability that patients in the placebo group improved in XYZ Symptom Index scores by 1 MSR or more from Baseline to Month 12 is **.12**
Patient's XYZ Symptoms are more likely improve by at least one MSR (reported as meaningful by patients) on treatment than on placebo

*These are meaningful in this hypothetical example. In some cases, some other region(s) or difference may be important.

Hypothetical Example 2: Summary

Vertical Approach	Horizontal Approach
Expected difference in probability of having XYZ Symptom Index scores improve by ≥ 1 MSRs from Baseline to Month 12	**Expected difference in Month 12 XYZ Symptom Index Scores**
Scenario B: Qualitative data supports *1-MSR difference** or more is meaningful to patients	
Probability that patients in the treatment group improved in XYZ Symptom Index scores by 1 MSR or more from Baseline to Month 12 is **.59** Probability that patients in the placebo group improved in XYZ Symptom Index scores by 1 MSR or more from Baseline to Month 12 is **.12**	
Patient's XYZ Symptoms are more likely improve by at least one MSR (reported as meaningful by patients) on treatment than on placebo	

*These are meaningful in this hypothetical example. In some cases, some other region(s) or difference may be important.

Hypothetical Example 2: Summary

Vertical Approach	Horizontal Approach
Expected difference in probability of having XYZ Symptom Index scores improve by ≥ 1 MSRs from Baseline to Month 12	**Expected difference in Month 12 XYZ Symptom Index Scores**
Scenario B: Qualitative data supports *1-MSR difference** or more is meaningful to patients	**Scenario B:** Qualitative data supports *1-MSR difference** or more is meaningful to patients
Probability that patients in the treatment group improved in XYZ Symptom Index scores by 1 MSR or more from Baseline to Month 12 is **.59** Probability that patients in the placebo group improved in XYZ Symptom Index scores by 1 MSR or more from Baseline to Month 12 is **.12**	
Patient's XYZ Symptoms are more likely improve by at least one MSR (reported as meaningful by patients) on treatment than on placebo	

*These are meaningful in this hypothetical example. In some cases, some other region(s) or difference may be important.

Hypothetical Example 2: Summary

<u>Vertical Approach</u> **Expected difference in probability of having XYZ Symptom Index scores improve by ≥ 1 MSRs from Baseline to Month 12**	<u>Horizontal Approach</u> **Expected difference in Month 12 XYZ Symptom Index Scores**
Scenario B: Qualitative data supports *1-MSR difference** or more is meaningful to patients	**Scenario B:** Qualitative data supports *1-MSR difference** or more is meaningful to patients
Probability that patients in the treatment group improved in XYZ Symptom Index scores by 1 MSR or more from Baseline to Month 12 is **.59** Probability that patients in the placebo group improved in XYZ Symptom Index scores by 1 MSR or more from Baseline to Month 12 is **.12**	On average, a patient in the treatment group is expected to achieve an XYZ Symptom Index score at Month 12 corresponding to an experience of *mild* On average, a patient in the placebo group is expected to achieve an XYZ Symptom Index score at Month 12 corresponding to an experience of *moderate*
Patient's XYZ Symptoms are more likely improve by at least one MSR (reported as meaningful by patients) on treatment than on placebo	

*These are meaningful in this hypothetical example. In some cases, some other region(s) or difference may be important.

Hypothetical Example 2: Summary

Vertical Approach	Horizontal Approach
Expected difference in probability of having XYZ Symptom Index scores improve by ≥ 1 MSRs from Baseline to Month 12	**Expected difference in Month 12 XYZ Symptom Index Scores**
Scenario B: Qualitative data supports *1-MSR difference** or more is meaningful to patients	**Scenario B:** Qualitative data supports *1-MSR difference** or more is meaningful to patients
Probability that patients in the treatment group improved in XYZ Symptom Index scores by 1 MSR or more from Baseline to Month 12 is **.59** Probability that patients in the placebo group improved in XYZ Symptom Index scores by 1 MSR or more from Baseline to Month 12 is **.12**	On average, a patient in the treatment group is expected to achieve an XYZ Symptom Index score at Month 12 corresponding to an experience of *mild* On average, a patient in the placebo group is expected to achieve an XYZ Symptom Index score at Month 12 corresponding to an experience of *moderate*
Patient's XYZ Symptoms are more likely improve by at least one MSR (reported as meaningful by patients) on treatment than on placebo	Patient's XYZ Symptoms are likely to be one MSR better (reported as meaningful by patients) on treatment than on placebo

*These are meaningful in this hypothetical example. In some cases, some other region(s) or difference may be important.

Final Thoughts

Important Reminders

- MSRs are <u>approximate</u> points of reference that help put the treatment effect in context

- Aim to have multiple estimates of MSRs using different methods (hypothetical examples had only one each)

  - Data can be messy

  - Talk to patients

  - Quality COAs (including anchors)

- Flexibility in PFDD Draft G4 framework

- The use of score interpretation metrics is just one part of assessing the meaningfulness of treatment effects.

Panel Discussion
followed by
Q+A

Day 1 Wrap-Up

Robyn Bent, RN, MS

Director, Patient-Focused Drug Development Program

Office of the Center Director

Center for Drug Evaluation and Research

U.S. Food and Drug Administration