Collecting and Applying Evidence to Interpret Clinical Outcome Assessment (COA)-based Endpoints in Clinical Trials

Weimeng Wang, PhD
Patient-Focused Statistical Scientists (PFSS)
Division of Biometrics III
CDER/OTS/Office of Biostatistics

Outline

- Background
- Score interpretation metrics: Putting COA scores in the context of patients' lives
  - Meaningful Score Regions (MSRs)
  - Meaningful Score Differences (MSDs)

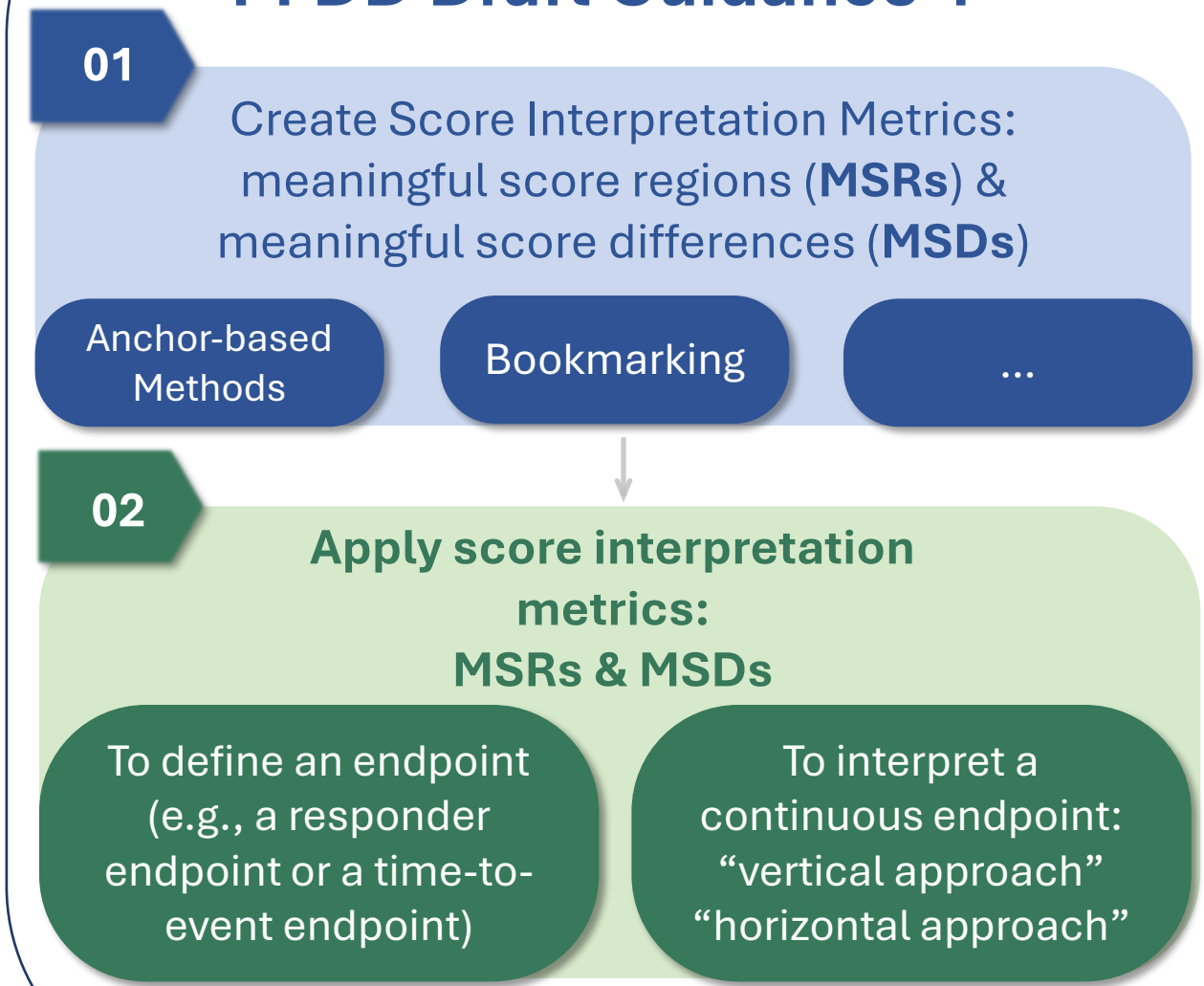
- Hypothetical example: Deriving MSDs and interpreting the meaningfulness of an estimated treatment effect on a COA-based endpoint
  - "Vertical Approach"
  - "Horizontal Approach"

Background: Importance of Interpretating COA scores

- **Why do we care?**
  - Statistical significance can be achieved for small differences between comparator groups
    - Does not indicate whether patients experienced **meaningful clinical benefit**
  - Need to assess improvement and deterioration from the **patients' perspectives**
  - Part of ensuring benefit of treatment outweighs the risk

- How does a **patient's status** on a COA-based endpoint correspond to the way they **feel and/or function in their daily lives**?

PFDD Draft Guidance 4

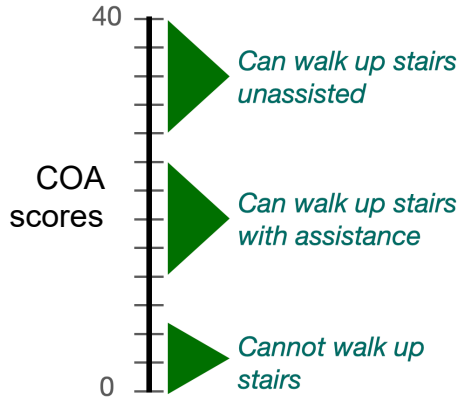


FDA's patient-focused drug development (PFDD) Guidance 4 Incorporating Clinical Outcome Assessments into Endpoints for Regulatory Decision Making

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

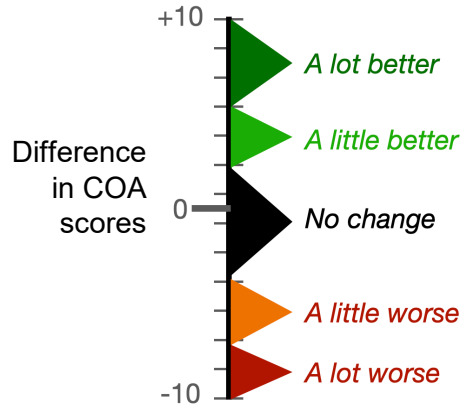
Meaningful Score Regions (MSRs)

What meaningful patient experiences correspond to ranges of COA scores?



Meaningful Score Differences (MSDs)

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



Neither MSR nor MSD approaches are necessarily better than the other

Select approach that best suits the PRO, endpoint, and context of use

Creating Score Interpretation Metrics: MSRs and MSDs

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>

Creating Score Interpretation Metrics: Anchor-based Methods

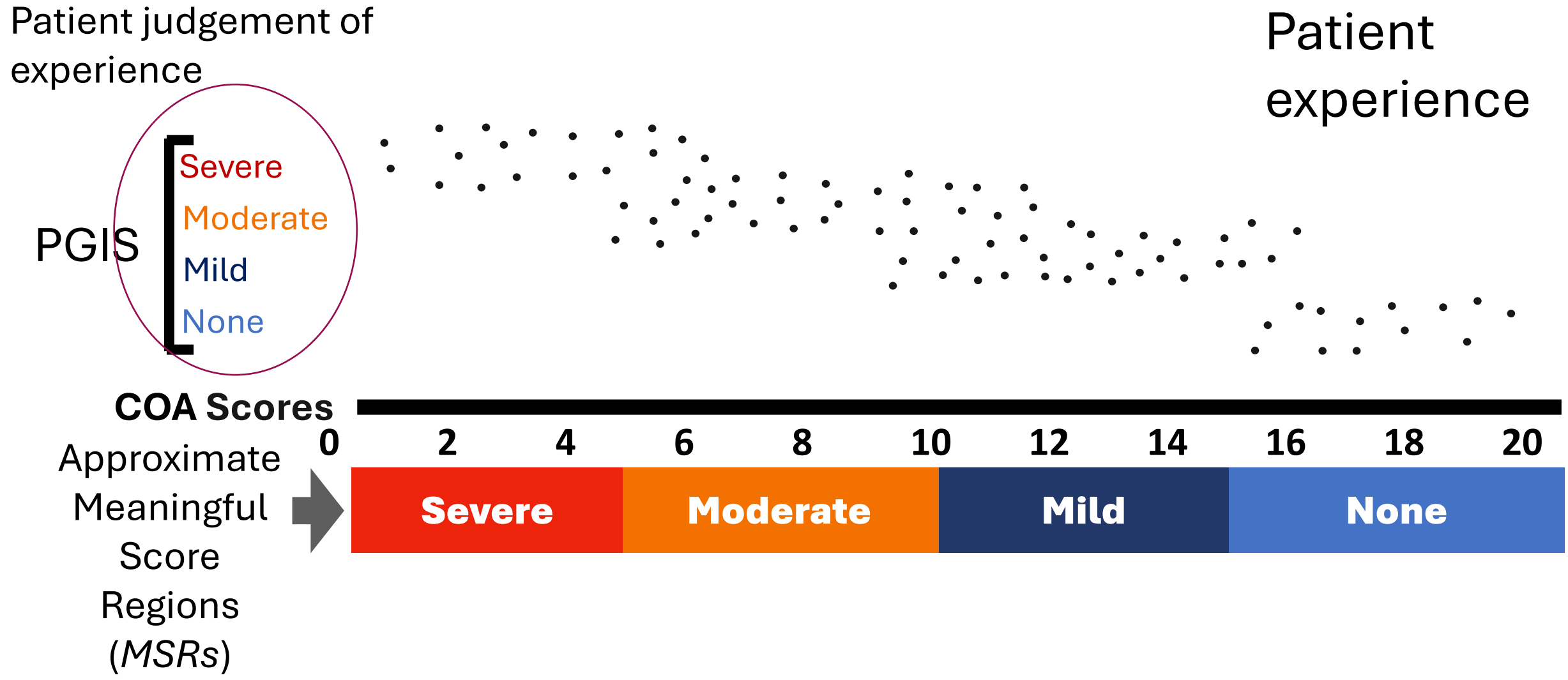
- **Anchor**: variable not part of COA being interpreted for which meaningful differences are directly interpretable or already known

Patient Global Impression of Severity (PGIS)
How would you rate the severity of your X Symptoms over the past 7 days?
None, Mild, Moderate, Severe, Very Severe

- Qualities of good anchors:
  - **Plainly understood** by respondents
  - Assesses **same concept** measured by COA-based endpoint
  - Assess **comparable time periods** as COA-based endpoint
  - Meaningful change/increments are **well-justified**
  - Use of **multiple anchors** is encouraged

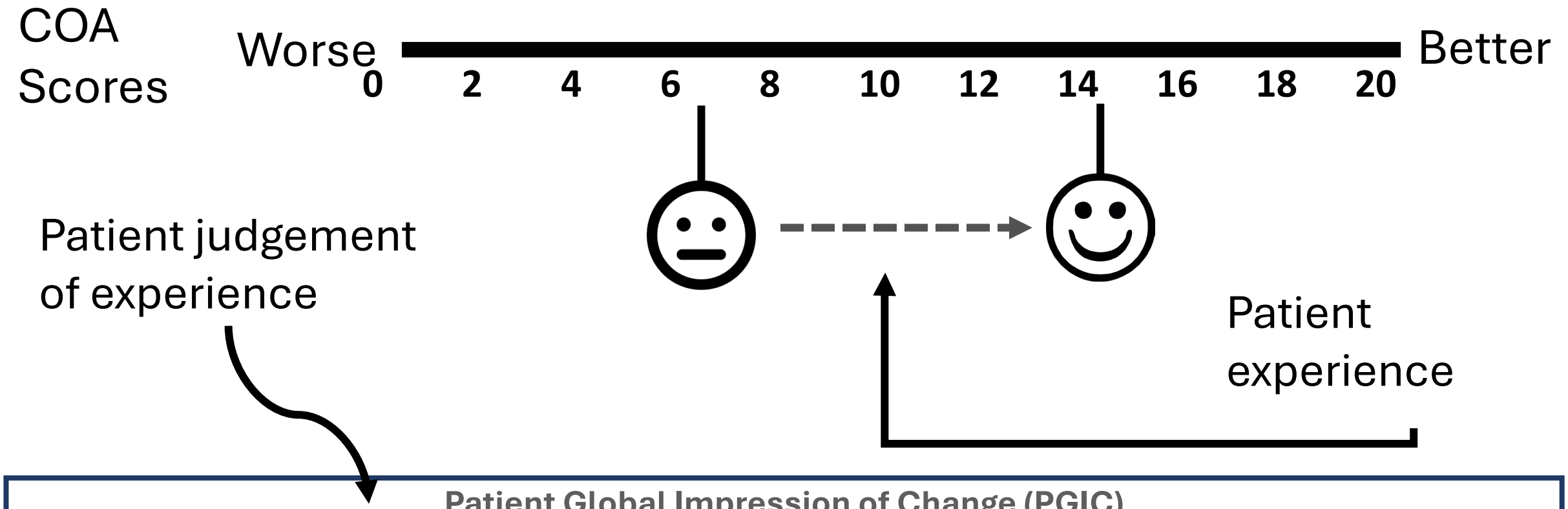
Meaningful Score Regions Approaches

What meaningful patient experiences correspond to different COA scores?



Meaningful Score Differences Approaches

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



Patient Global Impression of Change (PGIC)

How would you rate the change in severity of your X Symptoms since the start of study medication?

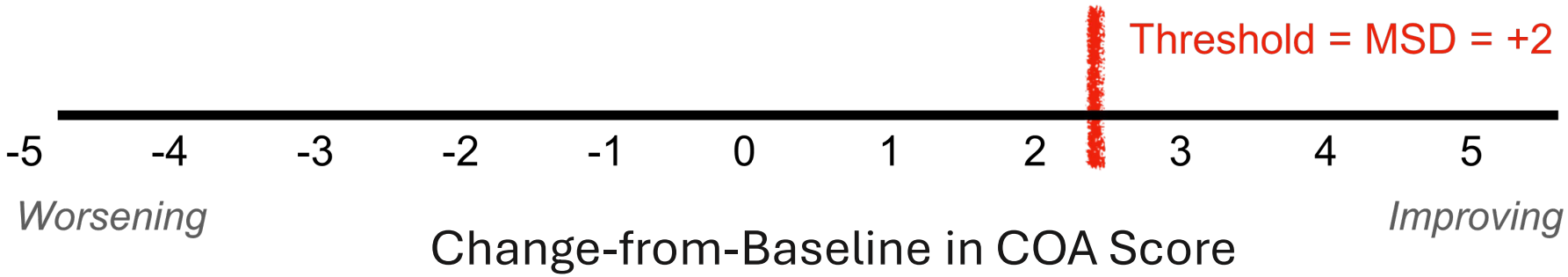
Much Worse, A Little Worse, No change, A Little Better, ***Much Better***

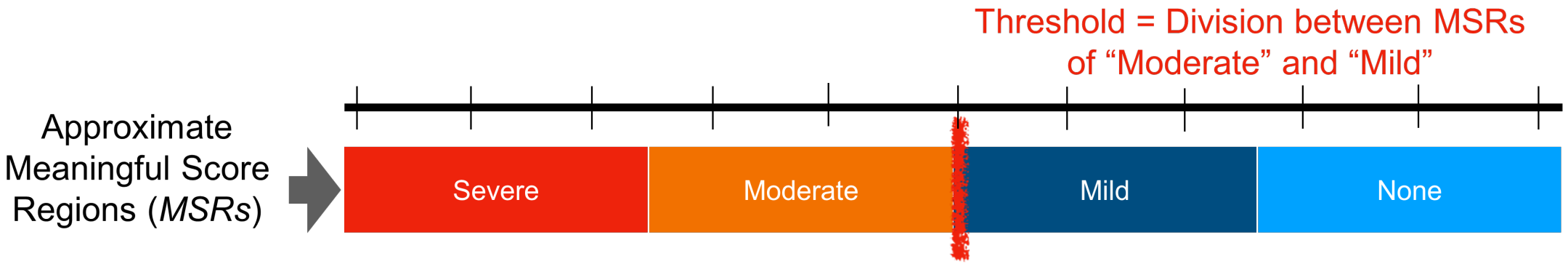
Applying Score Interpretation Metrics MSRs and MSDs

Applying Score Interpretation Metrics

Both MSRs and MSDs can be used in two ways

1. **Define an endpoint** (e.g., Responder endpoint, time-to-event endpoint)





Applying Score Interpretation Metrics

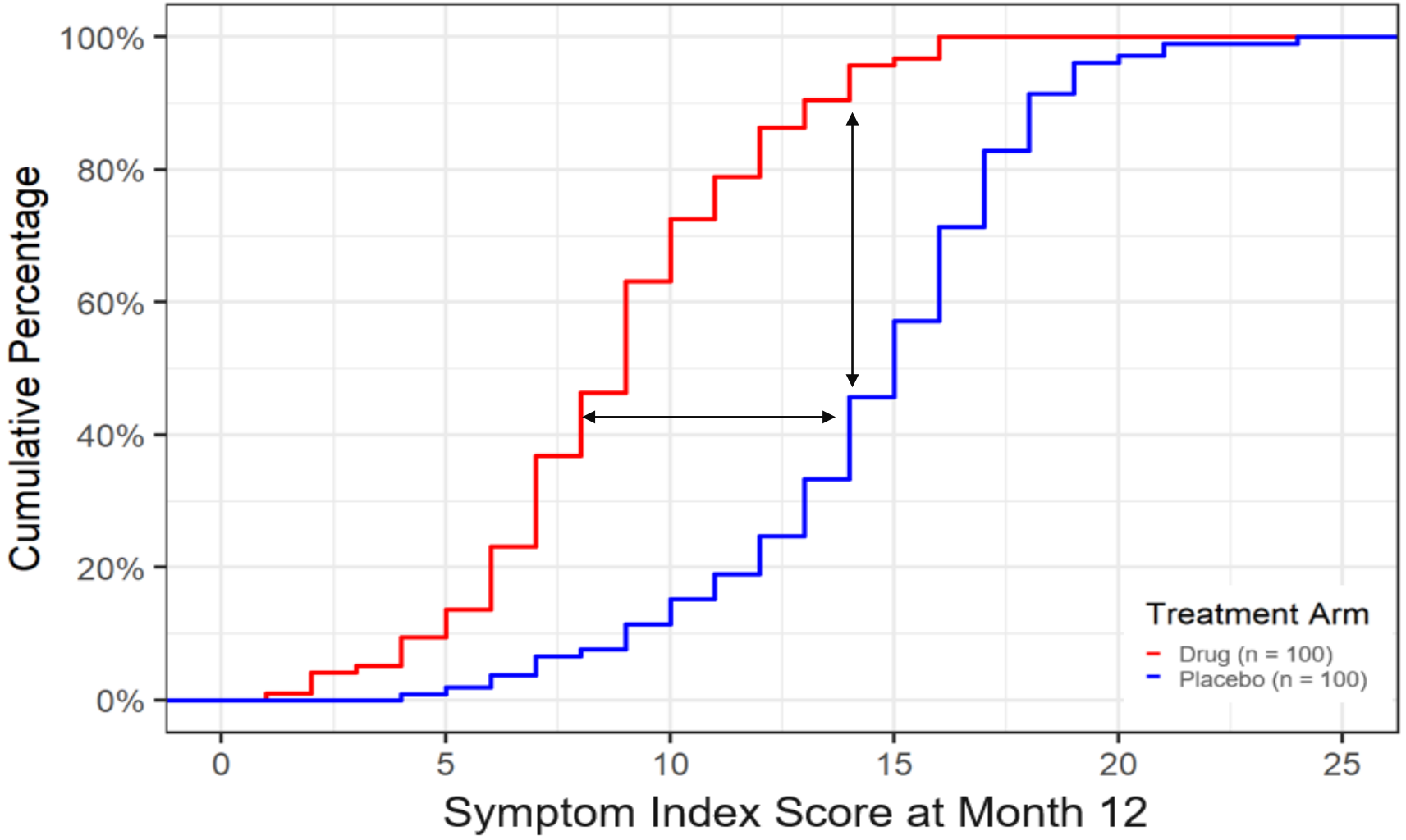
Both MSRs and MSDs can be used in two ways

1. **Define an endpoint** (e.g., responder endpoint, time-to-event endpoint)

2. **Interpret the meaningfulness of an estimated treatment effect on a COA-based endpoint** (e.g., COA score at fixed timepoint, change-from-baseline in COA score)

Interpreting the Treatment Effect using the Horizontal Approach and Vertical Approach

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



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

Average Horizontal Gap

Expected difference in COA-based endpoint score across treatment arms

This corresponds to the following question:

- **How much better is a patient expected to feel if given treatment rather than control?**
- e.g., The average patient's symptoms are likely to be meaningfully better on drug than on placebo.

Vertical Gap

Expected difference in probability of exceeding a specific COA-based endpoint score (i.e., MSRs or MSDs)

This corresponds to the following question:

- **How much more likely is a patient to experience a meaningful benefit in how they feel if given treatment rather than control?**
- e.g., Patients are about 18% to 20% more likely to experience a meaningful improvement in their symptoms if given drug rather than placebo.

Hypothetical Example

Using **Meaningful Score Differences** to aid in the interpretation of the meaningfulness of an estimated treatment effect on a change-from-baseline COA-based endpoint

- Deriving MSDs using the anchor-based analysis
- Applying MSDs to interpret the treatment effect

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 to FDA

Hypothetical Example: Study A

Meaningful Score Difference Derivation Study

- Pre-registrational, randomized trial, parallel groups design
- N = 250
- COA: Patient-reported *ABC Symptom Index*
  - Scores 0 (better) to 60 (worst)
- Anchor: Patient Global Impression of Severity (PGIS)
  - Response options: *None, Mild, Moderate, Severe, Very Severe*
- ABC Symptom Index and PGIS both have 7-day recall periods
- Administered at baseline and every 4 weeks thereafter to Week 24

Based on qualitative interviews with patients, a **1-category improvement** on the PGIS **was prespecified** as representing a meaningful improvement in severity.

A 1-category improvement on the PGIS anchor scale could occur in the following ways:

- Change from “very severe” to “severe”
- Change from “severe” to “moderate”
- Change from “moderate” to “mild”
- Change from “mild” to “none”

MSD approach assumes that the MSD value is the same regardless of how the prespecified-category change occurred

Category Change (n(%)) in PGIS from Baseline to Week 24, by Baseline PGIS

PGIS at Baseline	N	Worsened 2 Categories	Worsened 1 Category	No Change	Improved 1 Category	Improved 2 Categories
None	0	0	0	0	NA	NA
Mild	84	0	6 (7%)	48 (57%)	30 (36%)	NA
Moderate	107	2 (2%)	8 (7%)	23 (22%)	63 (59%)	11 (10%)
Severe	50	NA	0	9 (18%)	32 (64%)	9 (18%)
Very Severe	9	NA	NA	0	5 (55%)	4 (45%)

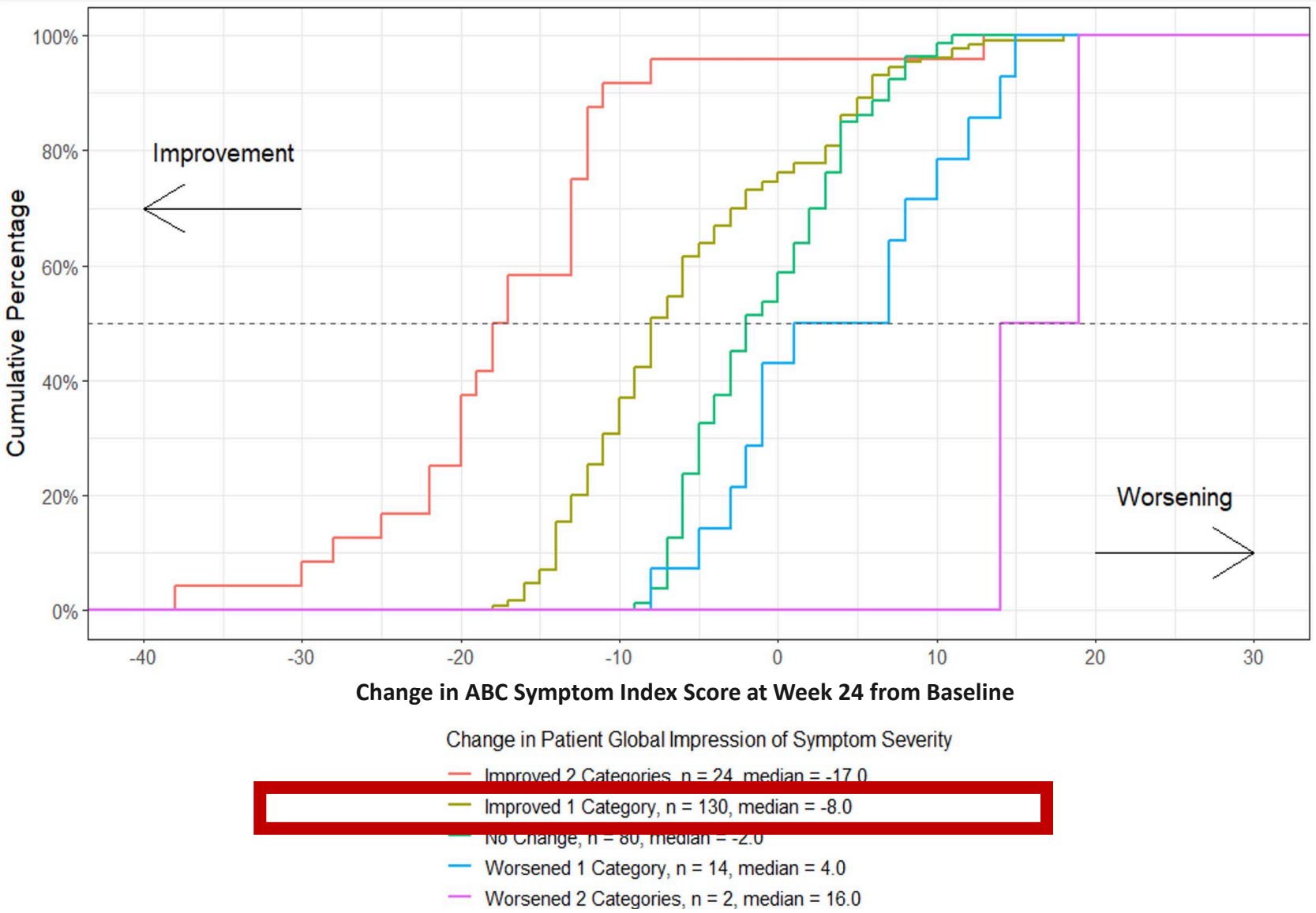
Category Change (n(%)) in PGIS from Baseline to Week 24, by Baseline PGIS

PGIS at Baseline	N	Worsened 2 Categories	Worsened 1 Category	No Change	Improved 1 Category	Improved 2 Categories
None	0	0	0	0	NA	NA
Mild	84	0	6 (7%)	48 (57%)	30 (36%)	NA
Moderate	107	2 (2%)	8 (7%)	23 (22%)	63 (59%)	11 (10%)
Severe	50	NA	0	9 (18%)	32 (64%)	9 (18%)
Very Severe	9	NA	NA	0	5 (55%)	4 (45%)

Change in ABC Symptom Index Score at Week 24 from Baseline, by Baseline PGIS

		Change from Baseline to Week 24 in ABC Symptom Index Score				
PGIS Baseline	**N**	**10th**	**25th**	**50th**	**75th**	**90th**
None	0	-	-	-	-	-
Mild	84	-13.0	-10.0	-5.0	4.0	12.0
Moderate	107	-23.0	-15.5	-6.0	3.5	10.4
Severe	50	-16.1	-12.0	-7.5	3.75	13.0
Very Severe	9	-12.2	-11.0	-6.5	4.0	7.6

Empirical Cumulative Distribution Functions for Change from Baseline to Week 24 in ABC Symptom Index Score by Change in PGIS



r = 0.59

Change in ABC Symptom Index Score at Week 24 from Baseline for Subjects who Achieved a 1-Category Improvement on PGIS by Baseline PGIS

		Change from Baseline to Week 24 in ABC Symptom Index Score Percentiles				
PGIS Baseline	N (%)	10th	25th	50th	75th	90th
Mild	30 (23%)	-14.0	-11.5	-8.0	3.3	11.1
Moderate	63 (48%)	-21.0	-18.0	-7.5	-0.5	7.0
Severe	33 (25%)	-16.0	-12.0	-8.5	-3.0	6.0
Very Severe	4 (3%)	-12.7	-12.7	-11.5	-3.8	9.3

Conclusion: MSD range is -8.5 to -7.5

Specifying the Meaningful Score Difference Range

- Sponsor proposed 1-category improvement* in ABC symptoms (per PGIS) is meaningful to patients based on qualitative data
- Examined distribution of PGIS change scores by baseline symptom severity
- Examined ABC Symptom Index change scores by baseline symptom severity
- Evaluated appropriateness of anchor measure
  - (Qualitative data)
  - Correlation with ABC Symptom Index change scores
  - PDF curves
  - eCDF curves
- Median ABC Symptom Index change score for subjects experiencing 1-category improvement on PGIS
  - Examined misclassification rate
- Explored ABC Symptom Index change scores by baseline severity
  - Subjects who experienced a 1-category improvement on PGIS

- **Conclusion: MSD range is -8.5 to -7.5**

1-category improvement in PGIS is meaningful in this hypothetical example. In other cases, more than 1-category improvement may be needed.

Hypothetical Example: Study X

FAKE DATA

- Randomized trial, parallel groups design
- Evaluated efficacy and safety of Drug A in 250 patients with metastatic/advanced cancer
  - N = 125 in Placebo
  - N = 125 in Drug
- COA Endpoint: Change in *ABC Symptom Index* weekly score from baseline to Week 24
  - COA weekly scores 0 (better) to 60 (worst)
  - COA collected every week for first 8 weeks of the study, and every 4 weeks thereafter to 24 weeks post-randomization
- *Main COA Analysis:* Comparison of study group mean change-from-baseline scores using mixed model for repeated measures (MMRM) with baseline *ABC Symptom Index* weekly score as covariate

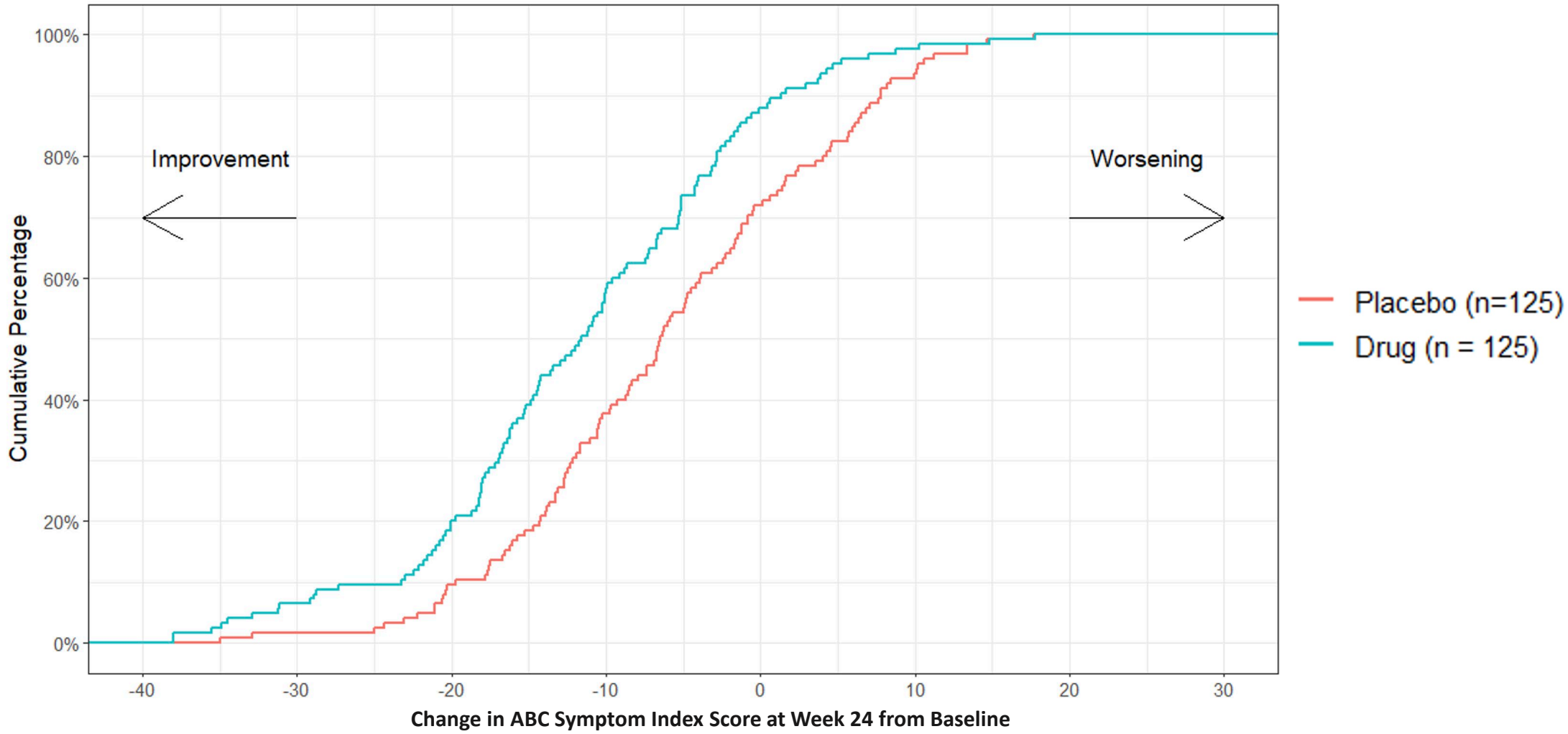
Hypothetical Example: Study X

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 MMRM 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

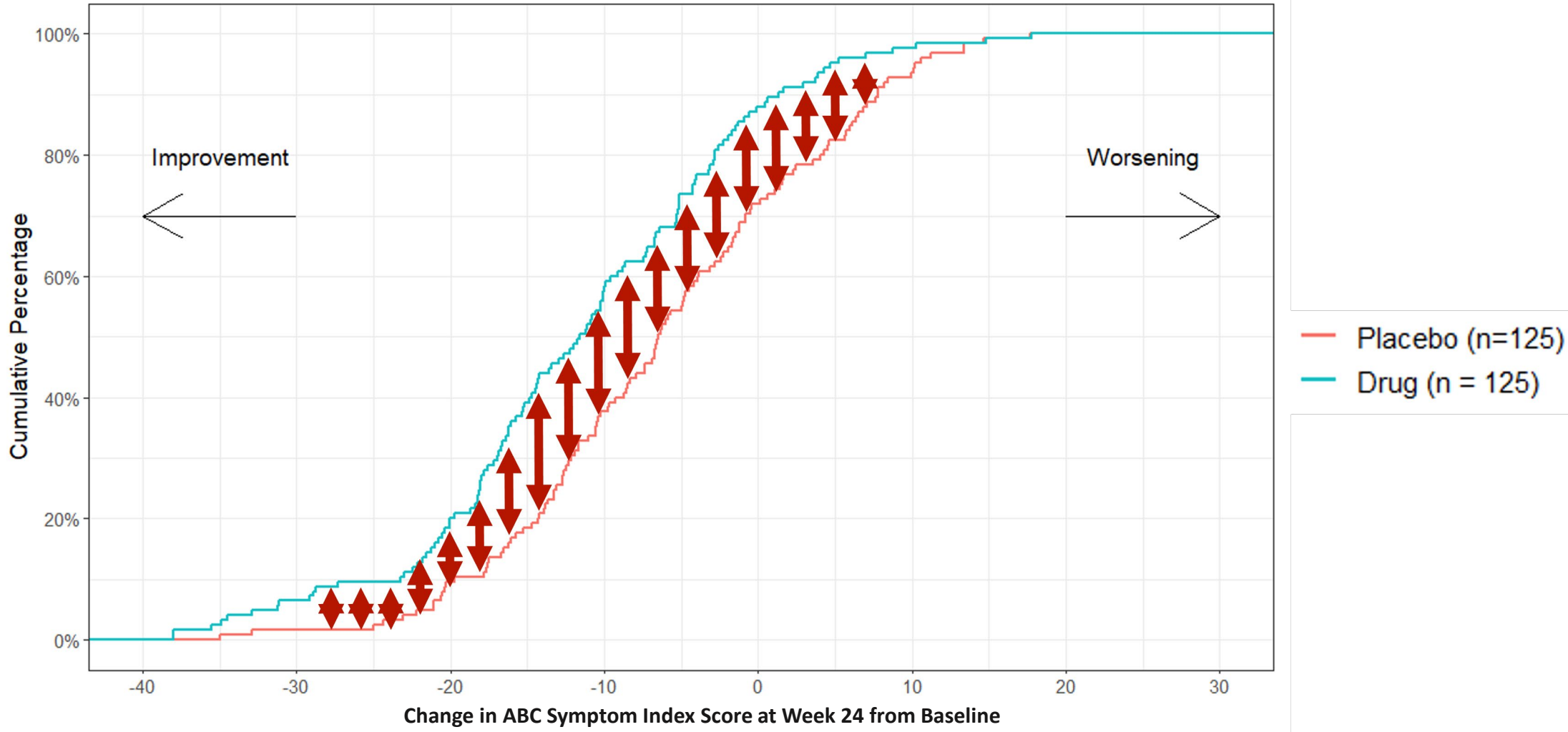
Empirical Cumulative Distribution Function: Change from Baseline to Week 24 in ABC Symptom Index Score by Treatment Group



“Vertical Approach”: Expected difference in the probability of exceeding MSDs thresholds (-8.5 to -7.5)

How much more likely is 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



- MSD Range: -8.5 to -7.5

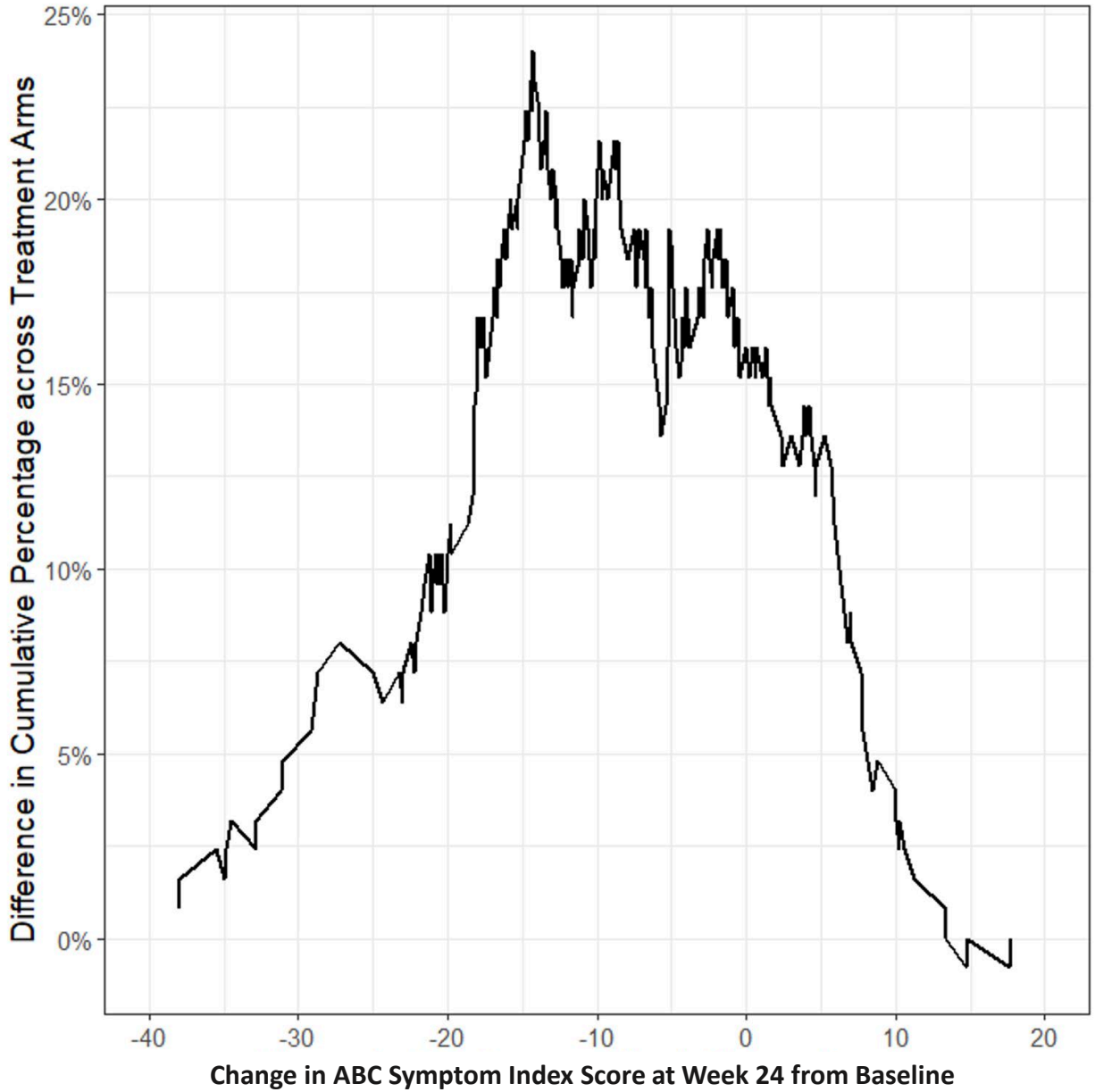


Figure based on Revicki DA, Erickson PA, Sloan JA, et al. Interpreting and Reporting Results Based on Patient-Reported Outcomes. *Value Health*. 2007;10(s2):S116-S124.

Difference between treatment arms ranges from 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?

MSD Range: -8.5 to -7.5

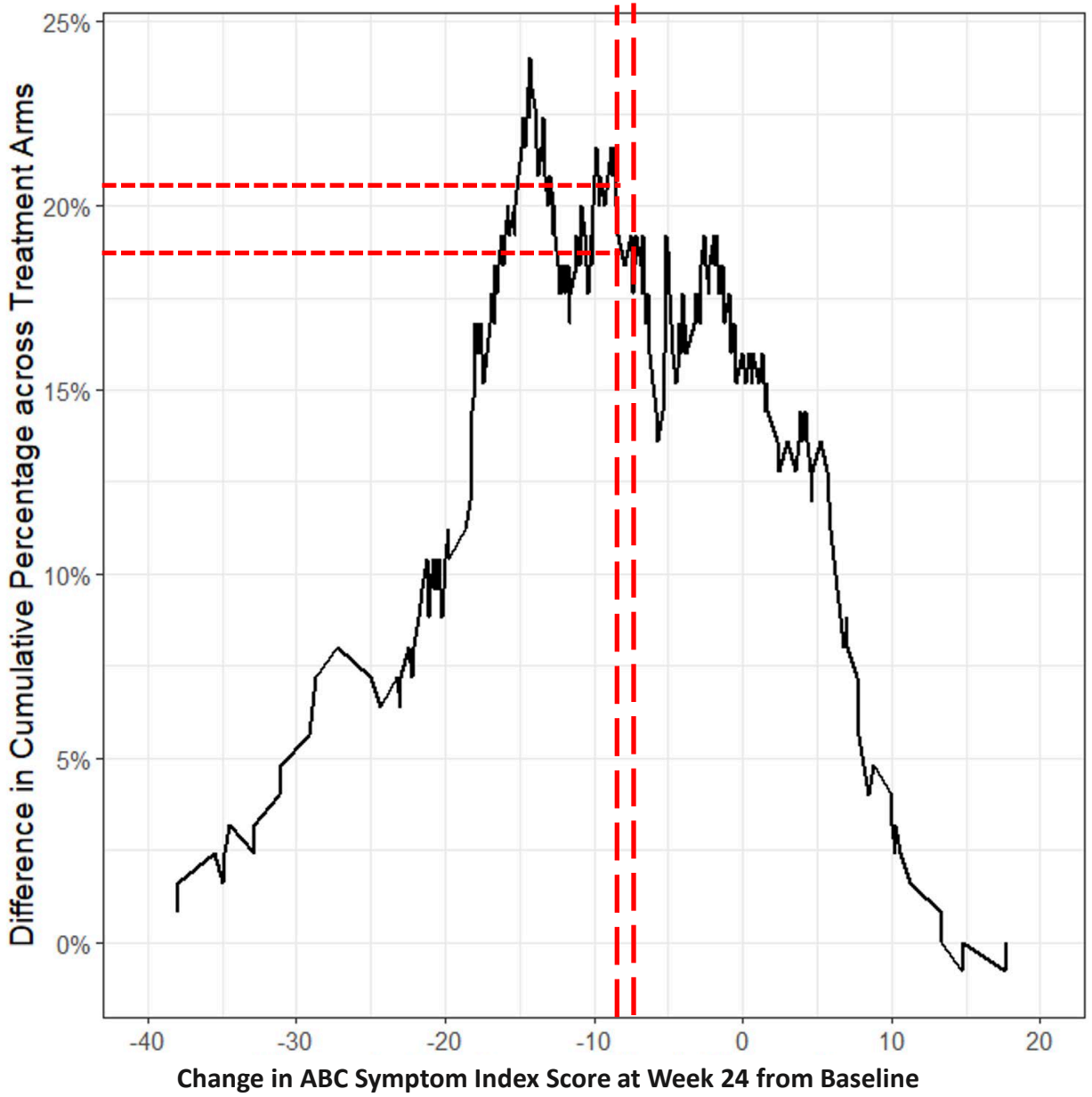
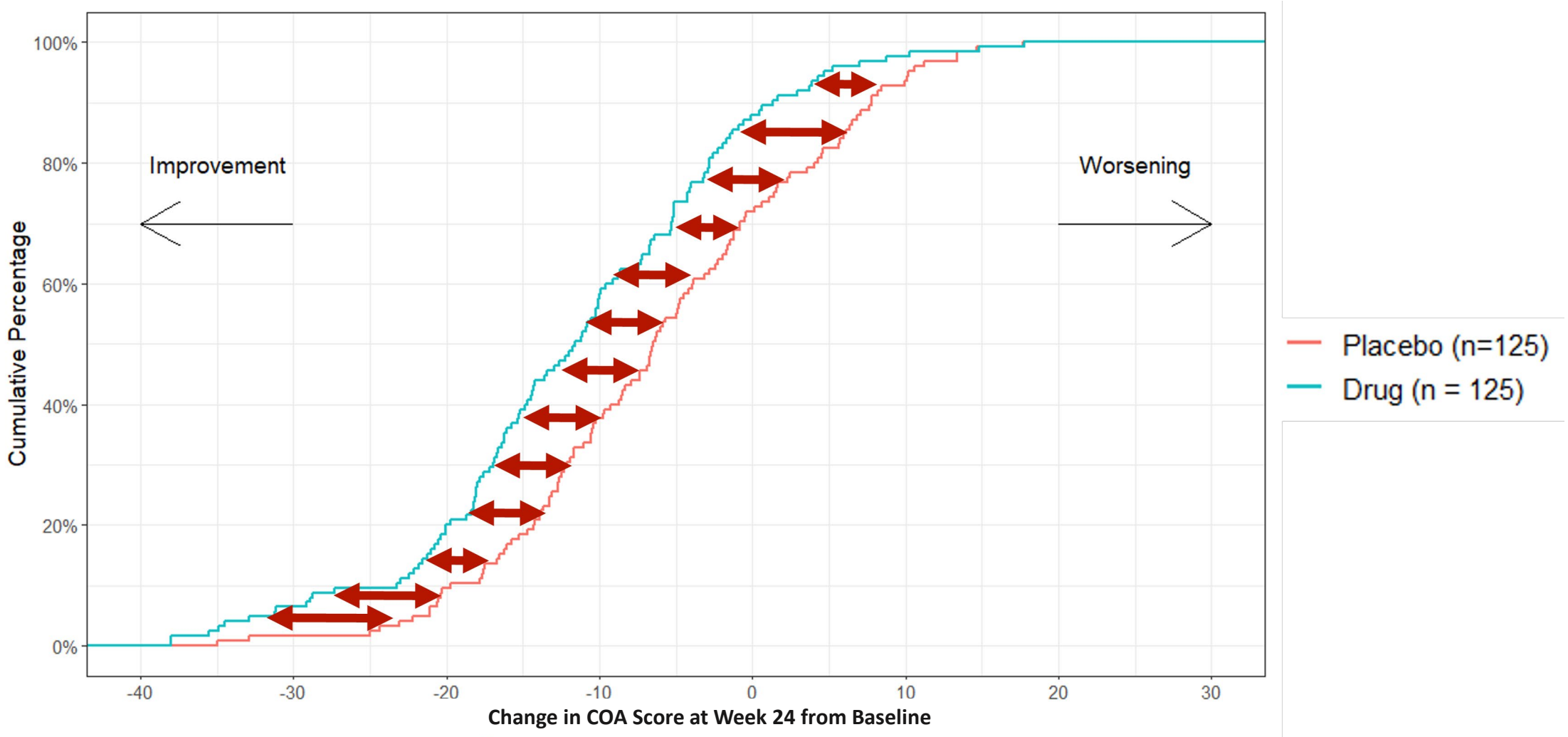


Figure based on Revicki DA, Erickson PA, Sloan JA, et al. Interpreting and Reporting Results Based on Patient-Reported Outcomes. *Value Health*. 2007;10(s2):S116-S124.

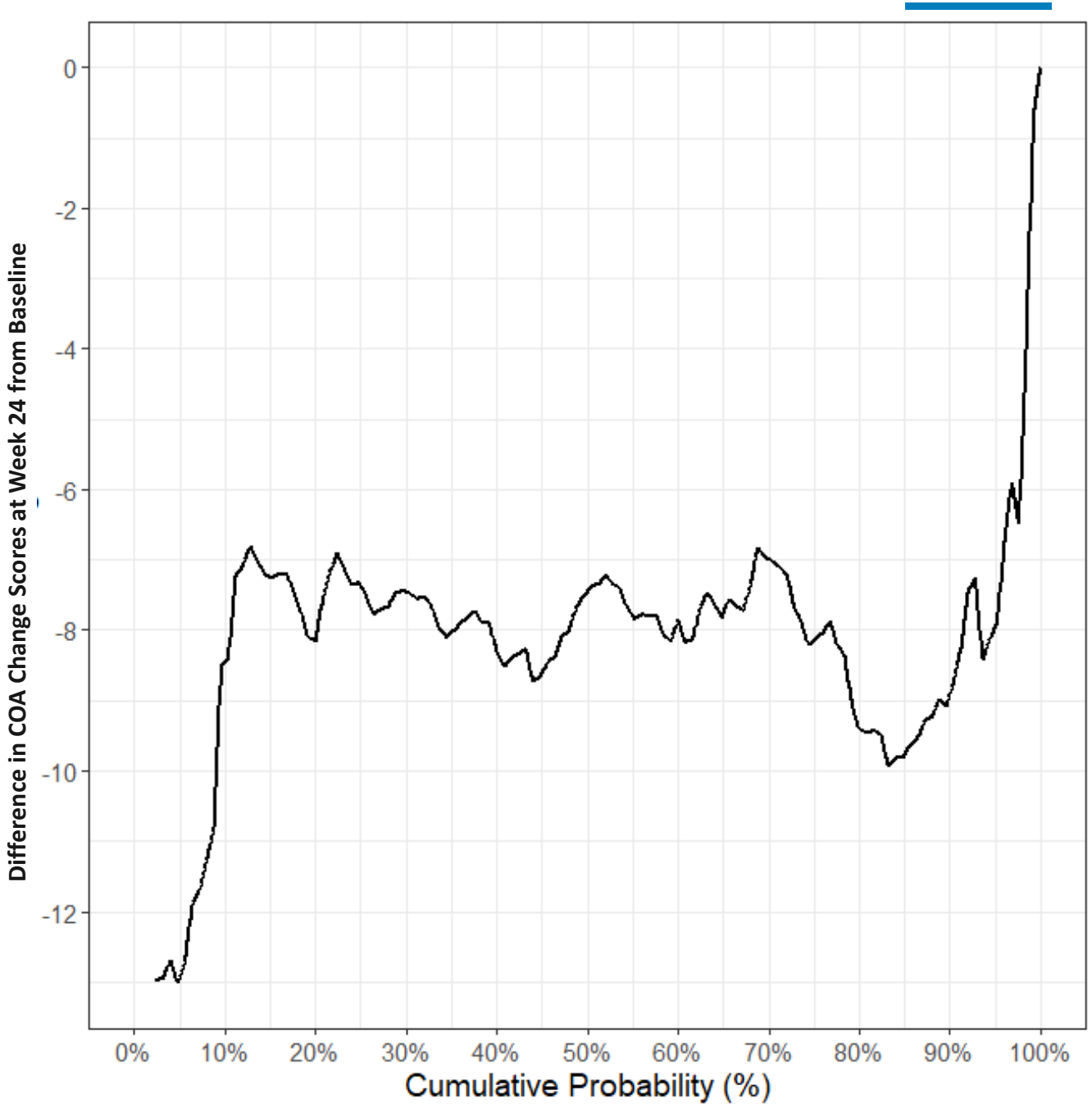
“Horizontal Approach”: Expected difference in change from baseline to week 24 ABC Symptom Index scores

How much better is the average patient’s ABC symptoms likely to be if they receive drug rather than placebo?

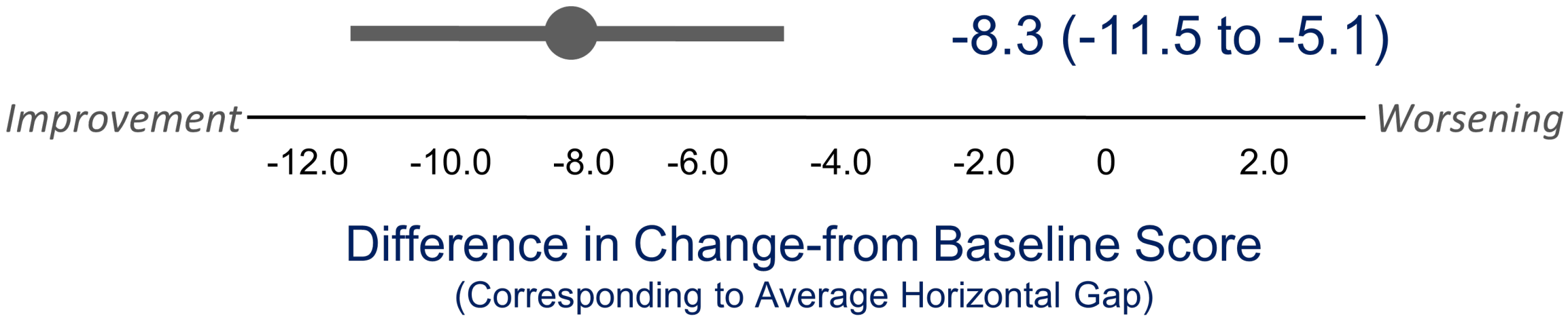
- Overall estimate of treatment effect (difference between group means) corresponds to the average horizontal gap between eCDFs
- Check to see if it is relatively consistent



- Directly plotting the horizontal gap shows it is relatively consistent
- Supports use of difference in group means to estimate size of treatment effect

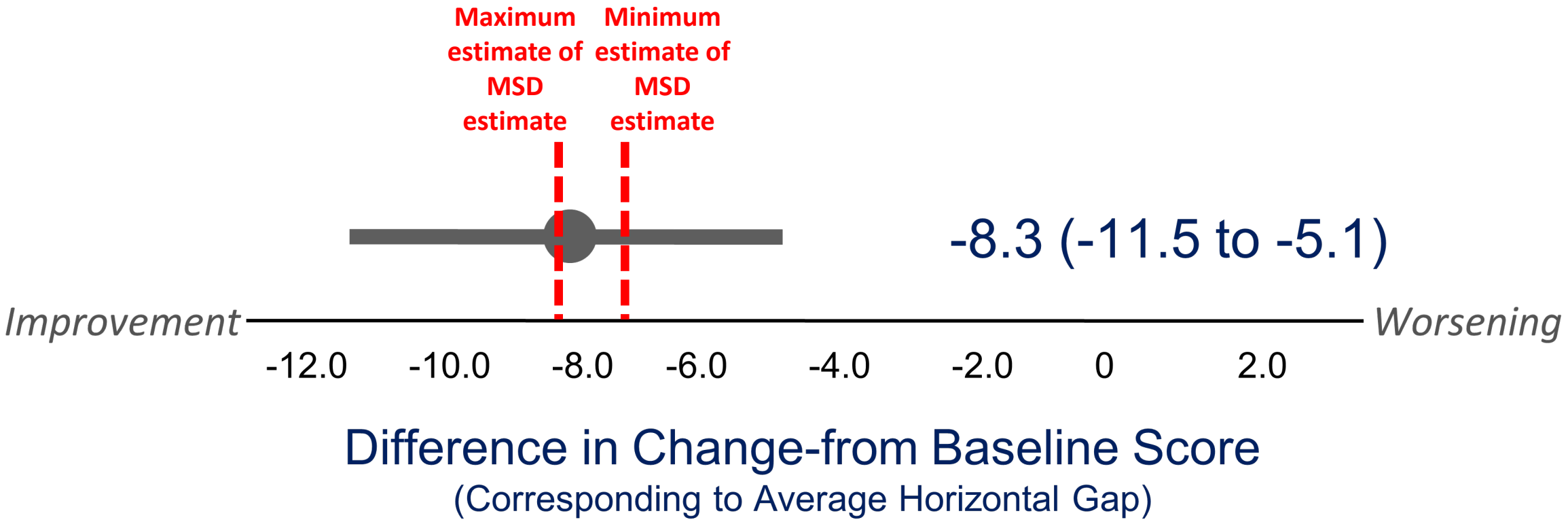


Estimated Difference in Adjusted Means (With 95% Confidence Interval) Between Treatment and Placebo on Change-from Baseline



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



- **<u>Not</u> another form of statistical hypothesis testing**

Summary of Example

Interpreting the Treatment Effect using MSDs

- Endpoint: Change in *ABC Symptom Index* score from baseline to week 24
- Expected differences in probability of exceeding MSD thresholds (“**Vertical Approach**”)
- Expected difference in change from baseline to week 24 ABC Symptom Index score (“**Horizontal Approach**”)

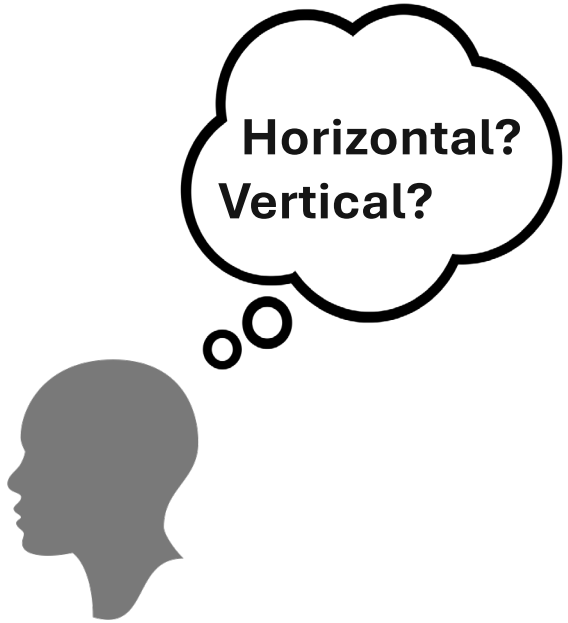
Important Reminders

- MSDs and MSRs are **approximate points of reference** that help put the treatment effect in **context**
- Aim to have **multiple estimates** of MSDs or MSRs using **different methods** (hypothetical example had only one)
- Qualitative evidence (e.g., exit interviews)
- The use of MSDs or MSRs is just **one part** of assessing the meaningfulness of treatment effects

Final Thoughts

Interpreting the Meaningfulness of Treatment effects

- Understanding what COA scores mean
  - MSRs: connecting **scores**
  - MSDs: connecting **score differences**
- Aim to have **multiple estimates** of MSRs or MSDs using **different methods**
  - Using MSRs or MSDs 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
- The use of MSRs or MSDs is just **one part** of assessing the meaningfulness of treatment effects
  - Multiple endpoints (e.g., primary and secondary endpoints)
  - Prespecified sensitivity analyses to supplement the main trial analysis of the COA-based endpoints
  - Qualitative evidence (e.g., cognitive interviews)



- **Select approach, and provide a rationale, that best suits the COA, endpoint, and context of use**

Additional Resources

- FDA's patient-focused drug development (PFDD) Guidance 4 Incorporating Clinical Outcome Assessments into Endpoints for Regulatory Decision Making
- FDA PFDD guidance 4 public webinar: https://www.fda.gov/drugs/news-events-human-drugs/public-webinar-patient-focused-drug-development-incorporating-clinical-outcome-assessments-endpoints
- FDA PFDD Guidance Series: https://www.fda.gov/drugs/development-approval-process-drugs/fda-patient-focused-drug-development-guidance-series-enhancing-incorporation-patients-voice-medical
- FDA Guidance - Core PRO in Cancer Clinical Trials: https://www.fda.gov/regulatory-information/search-fda-guidance-documents/core-patient-reported-outcomes-cancer-clinical-trials

Questions?

Thank you!

Special thanks to Monica Morell, Laura Lee Johnson, Lili Garrard