

# Enhancing Interpretation of Patient-reported Outcomes: Responder Analysis, Cumulative Distributions, and Regulatory Insights

**Joseph C. Cappelleri, PhD, MPH**

*Pfizer Inc*

**Lisa A. Kammerman, PhD**

*FDA*

**Kathy Wyrwich, PhD**

*United BioSource Corporation*

## ***SECOND ANNUAL PATIENT-REPORTED OUTCOME (PRO) CONSORTIUM WORKSHOP***

**March 15, 2011 ■ Silver Spring, MD**

**Co-sponsored by**



# Panel Overview



- Opening remarks – Introductions
- Interpretation of Patient-Reported Outcomes for Label and Promotional Claims Using a Responder Definition
- Interpretation of Patient-Reported Outcomes — Cumulative Distributions and Other Techniques
- CDER's Use of Responder Analyses and Cumulative Distribution Functions
- Commentary
- Panel Questions & Answers

# Disclaimer



The views expressed here do not reflect the views of Pfizer Inc, United BioSource Corporation, or FDA

# **Part 1: Interpretation of Patient-Reported Outcomes for Label and Promotional Claims Using a Responder Definition**

**Kathy Wyrwich**

*United BioSource Corporation*

# Introduction



- Key is to focus on pre-specified patient-reported outcome (PRO)
- Important to report all pre-specified PROs (not just those that are “significant”)
- Also important to report all PROs, pre-specified or not

# Guidance for Industry Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims



*Additional copies are available from:*

*Office of Communications, Division of Drug Information  
Center for Drug Evaluation and Research  
Food and Drug Administration  
10903 New Hampshire Ave., Bldg 51, rm 2201  
Silver Spring, MD 20993-0002*

*Tel: 301-796-3400; Fax: 301-847-8714; E-mail: [druginfo@fda.hhs.gov](mailto:druginfo@fda.hhs.gov)  
<http://www.fda.gov/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/default.htm>*

*or*

*Office of Communication, Outreach, and Development, HFM-40  
Center for Biologics Evaluation and Research  
Food and Drug Administration*

*1401 Rockville Pike, Suite 200N, Rockville, MD 20852-1448  
Tel: 800-835-4709 or 301-827-1800; E-mail: [ocod@fda.hhs.gov](mailto:ocod@fda.hhs.gov)  
<http://www.fda.gov/BiologicsBloodVaccines/GuidanceComplianceRegulatoryInformation/default.htm>*

*or*

*Office of Communication, Education, and Radiation Programs  
Division of Small Manufacturers, International, and Consumer Assistance, HFZ-220  
Center for Devices and Radiological Health  
Food and Drug Administration*

*1350 Piccard Drive, Rockville, MD 20850-4307  
DSMICA E-mail: [ds mica@cdrh.fda.gov](mailto:ds mica@cdrh.fda.gov)  
DSMICA Fax: 301-443-8818  
(Tel) Manufacturers Assistance: 800-638-2041 or 301-443-6597  
(Tel) International Staff: 301-827-3993  
<http://www.fda.gov/MedicalDevices/DeviceRegulationandGuidance/GuidanceDocuments/default.htm>*

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

**December 2009  
Clinical/Medical**

# Interpretation of PRO Change over Time



- Not considered a measurement property
- Interpretation of PRO endpoints follows similar considerations as for all other endpoint types used to evaluate treatment benefit of a medical product
  - Interpretation of clinical data that supports the development of the instrument in Stage IV on the Wheel & Spoke diagram and
  - Interpretation of the clinical outcome of the confirmatory trial

# PRO Guidance – Interpretation of Data

## i. Hypothesize Conceptual Framework

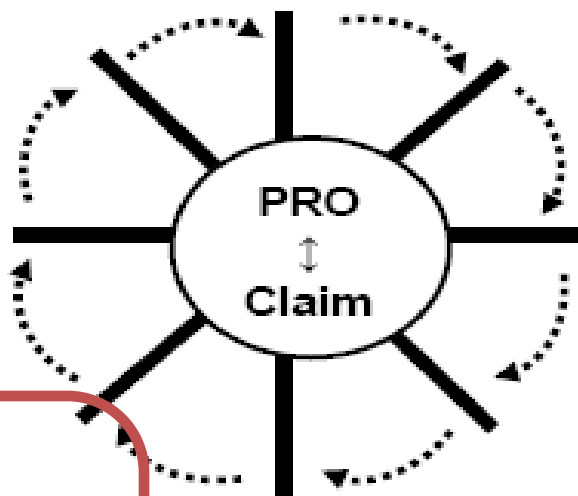
- Outline hypothesized concepts and potential claims
- Determine intended population
- Determine intended application/characteristics (type of scores, mode and frequency of administration)
- Perform literature/expert review
- Develop hypothesized conceptual framework
- Place PROs within preliminary endpoint model
- Document preliminary instrument development

## ii. Adjust Conceptual Framework and Draft Instrument

- Obtain patient input
- Generate new items
- Select recall period, response options and format
- Select mode/method of administration/data collection
- Conduct patient cognitive interviewing
- Pilot test draft instrument
- Document content validity

## iii. Confirm Conceptual Framework and Assess Other Measurement Properties

- Confirm conceptual framework with scoring rule
- Assess score reliability, construct validity, and ability to detect change
- Finalize instrument content, formats, scoring, procedures and training materials
- Document measurement development



## v. Modify Instrument

- Change wording of items, populations, response options, recall period, or mode/method of administration/data collection
- Translate and culturally adapt to other languages
- Evaluate modifications as appropriate
- Document all changes

## iv. Collect, Analyze, and Interpret Data

- Prepare protocol and statistical analysis plan (final endpoint model and responder definition)
- Collect and analyze data
- Evaluate treatment response using cumulative distribution and responder definition
- Document interpretation of treatment benefit in relation to claim



# Necessary Before Clinical Trial Data Interpretation



- Endpoint model supports indication/claim
- Adequate instrument development
  - Conceptual framework
  - Content validity
  - Construct validity, reliability, and ability to detect change
  - Cultural and linguistic adaptation
- Adequate clinical trial design
  - Blinding and randomization
  - Method for handling missing data
  - Statistical analysis plan

# Interpretation is More Than $p < .05$



- Need to achieve statistically significant differences between the active treatment and placebo arms for clinical trials, but it's just not enough (but is the way to properly power most trials!)
- Need a way to interpret if statistically significant differences are meaningful and important to clinical trial participants
- Can't rely on  $p < 0.05$  to demonstrate an interpretable difference
  - Many PRO scales are new to label readers and familiarity with what types of changes are important requires experience over time

# How Do You Determine the *Responder Definition* for a PRO Instrument?

# Key to Interpretation: Responder Definition



- Defined as the trial-specific important difference standard or threshold applied at the individual level of analysis
- This represents the individual patient PRO score change over a predetermined time period that should be interpreted as a treatment benefit

# Responder Definition



- The responder definition is determined empirically and may vary by target population or other clinical trial design characteristics
- FDA reviewers will evaluate a PRO instrument's responder definition in the context of each specific clinical trial

# Anchor-Based Methods



- Anchor-based methods explore the associations between the targeted concept of the PRO instrument and the concept measured by the anchors
- To be useful, the anchors chosen should be easier to interpret than the PRO measure itself

# Anchor-Based Methods



- Correlated response between targeted PRO instrument and measures of closely-related concepts
- Provide meaning or interpretation of change in a measure
- Anchor selection should have intuitive meaning

# Types of Anchors

- Clinical measure
  - a 50% reduction in incontinence episodes might be proposed as the anchor for defining a responder
- Clinician-reported outcome
  - Clinician global rating of change (CGIC) in mental health conditions
- Patient global ratings
  - Patient global rating of *change*
  - Patient global rating of *concept*



## Example of Responder Definition: Pain Intensity Numerical Rating Scale (PI-NRS)



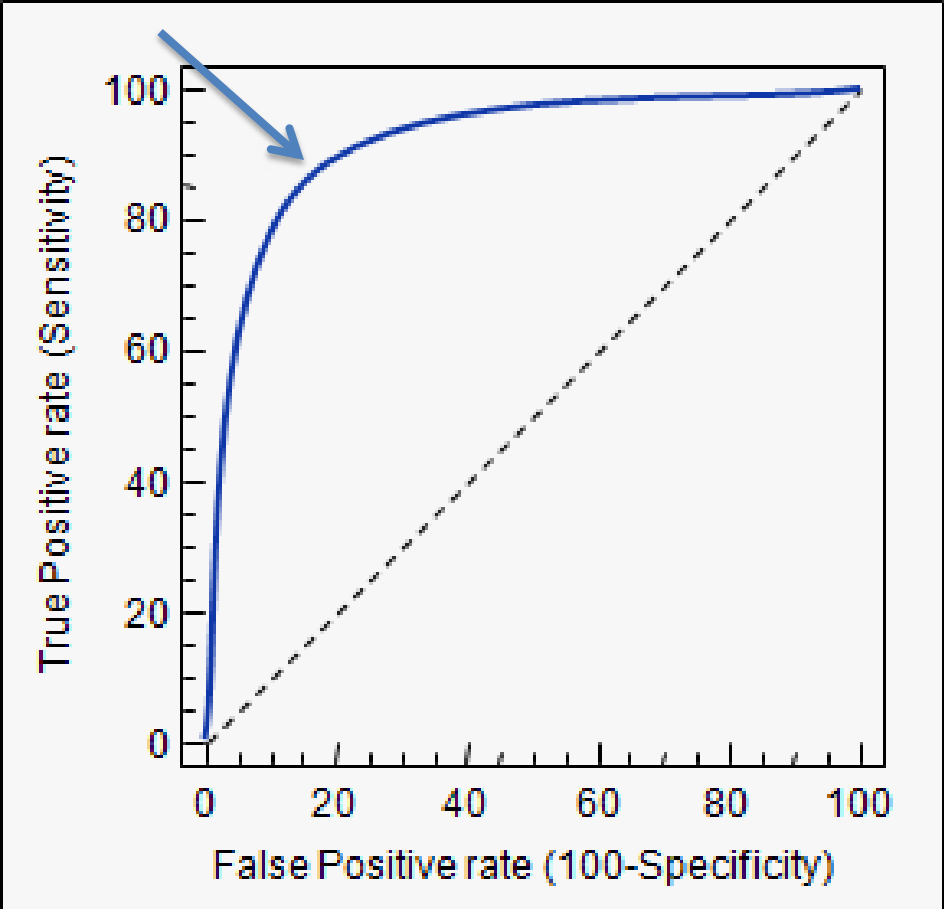
- Farrar JT et al. *Pain* 2001; 94:149-158
- **11-point pain scale: 0 = no pain to 10 = worst pain**
  - Baseline score = mean of 7 diary entries prior to drug
  - Endpoint score = mean of last 7 diary entries
  - Interest centers on change score
  - Primary endpoint in pregabalin program
- 10 chronic pain studies with 2724 subjects
  - Placebo-controlled trials of pregabalin
  - Several conditions (e.g., fibromyalgia and osteoarthritis)

# Example of Responder Definition: Pain Intensity Numerical Rating Scale (PI-NRS)



- Patient Global Impression of Change (anchor)
  - Clinical improvement of interest
  - Best change score for distinguishing ‘much improved’ or better on PGIC
- Since the start of the study, my overall status is:
  1. Very much improved
  2. Much improved
  3. Minimally improved
  4. No change
  5. Minimally worse
  6. Much worse
  7. Very much worse

PI-NRS score type	Model	Area under the curve	Change	Sensitivity (%)	Specificity (%)
Raw change	Very much improved	0.873	−2.76	79.2	80.1
Raw change	Much or very much improved	0.853	−1.74	77.0	78.6
Raw change	Minimally, much or very much improved	0.832	−1.0	77.9	75.3
Percent change	Very much improved	0.890	−46.51	81.5	81.5
Percent change	Much or very much improved	0.859	−27.9	78.4	78.4
Percent change	Minimally, much or very much improved	0.832	−14.5	76.8	76.8



# Types of Anchors: Patient Global Ratings

- Useful in defining a responder definition *a priori*
- Not intended as labeling claims
- Two Types:
  - Patient global rating of change
  - Patient global rating of concept



# Patient Global Rating of Change



- Comprehensive evaluation of complex concept
- Comparative
- Usually a long recall period
- Less desirable due to long recall period

# Patient Global Rating of Concept

- Comprehensive evaluation of complex concept
- Non-comparative (e.g., rating of current condition)
- Minimal or no requirement for patients to average their condition over long periods of time
- Example: “How would you rate your IBS symptoms overall over the past seven days?”
  - Question used at baseline and at endpoint (cross-sectional)

# Missing from the PRO Guidance



Minimum Important  
Difference (MID)!

# Minimum Important Difference (MID) / Minimum Clinically Important Difference (MCID)

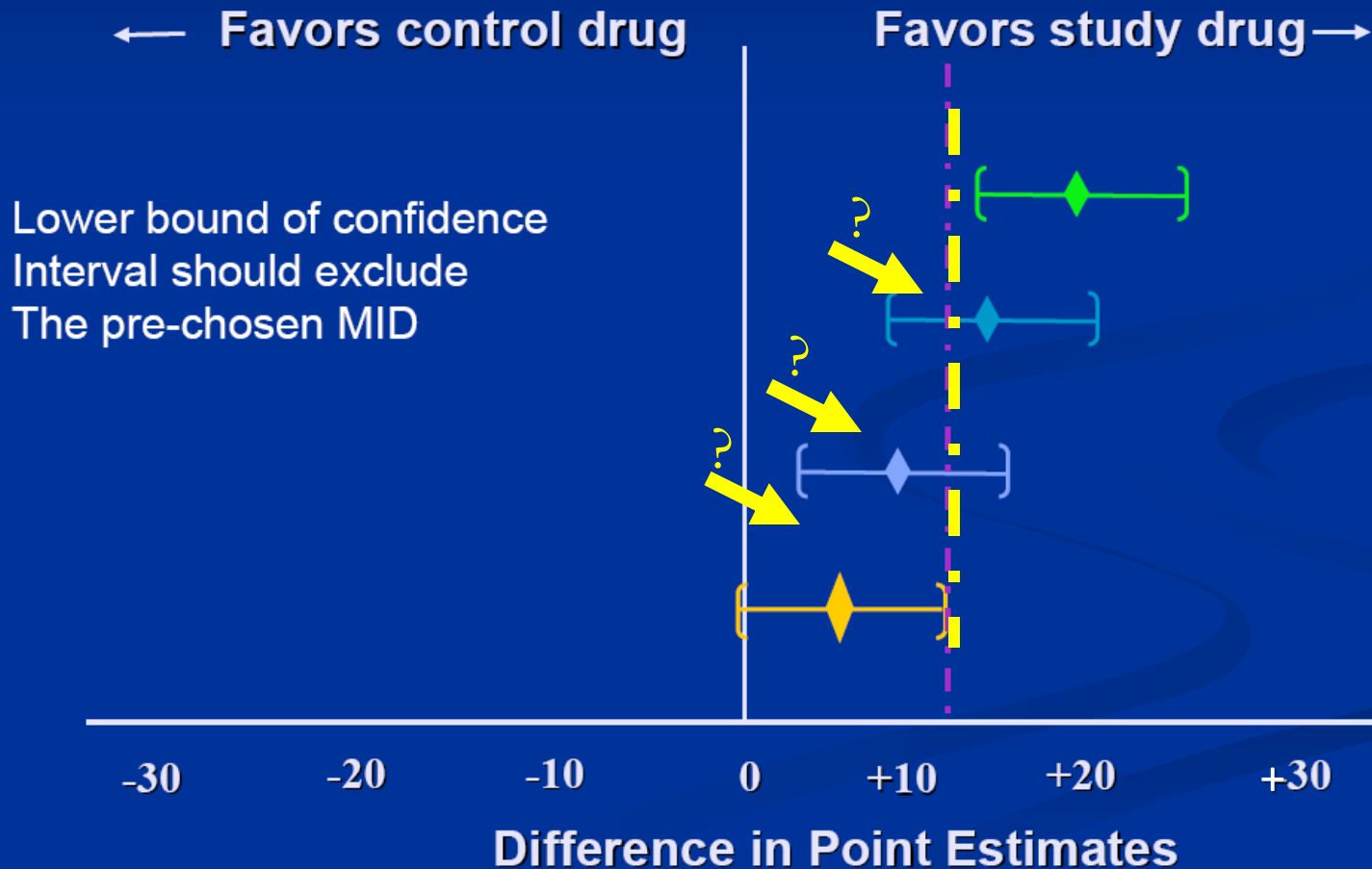


- Defined in *draft* PRO Guidance (2006) as:
  - The smallest difference between clinical trial treatment arm mean change from baseline (point estimates) that will be interpreted as important
- MID represents the between groups criterion that needs to be met or exceeded in order for study results to be considered clinically meaningful



# Ruling Out an Important Difference

To demonstrate that the treatment difference is statistically significantly larger than the MID, the lower bound of the confidence interval for the treatment difference should exceed the chosen MID



- Why is MID not included in Final PRO Guidance?
  - Term is interpreted inconsistently (intra-patient change vs. inter-group difference of mean change from baseline)
  - Point estimates of the difference in means between two groups may mask important changes for individual patients or types of patients in each group
  - Responder definitions offer a direct approach to intra-patient change and treatment differences across a range of clinical anchors that can be presented in a cumulative distribution function

**Part 2:**

**Interpretation of Patient-Reported  
Outcomes — Cumulative Distributions  
and Other Techniques**

**Joseph C. Cappelleri**  
*Pfizer Inc*

# Cumulative Distribution Function

# Cumulative Distribution Function

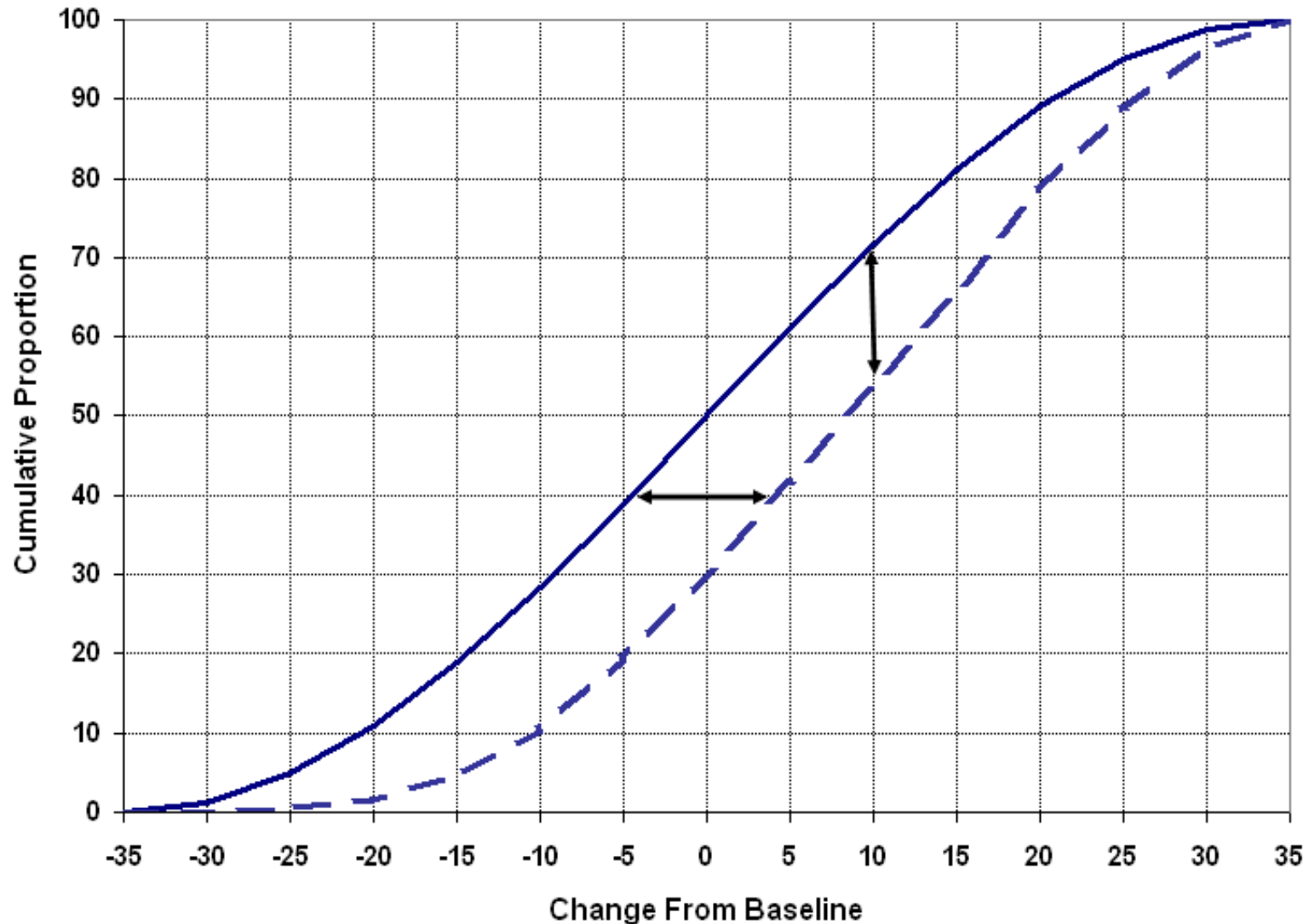
- An alternative or supplement to responder analysis
- Mentioned prominently in the FDA Guidance on PRO label and promotional claims

# Cumulative Distribution Function

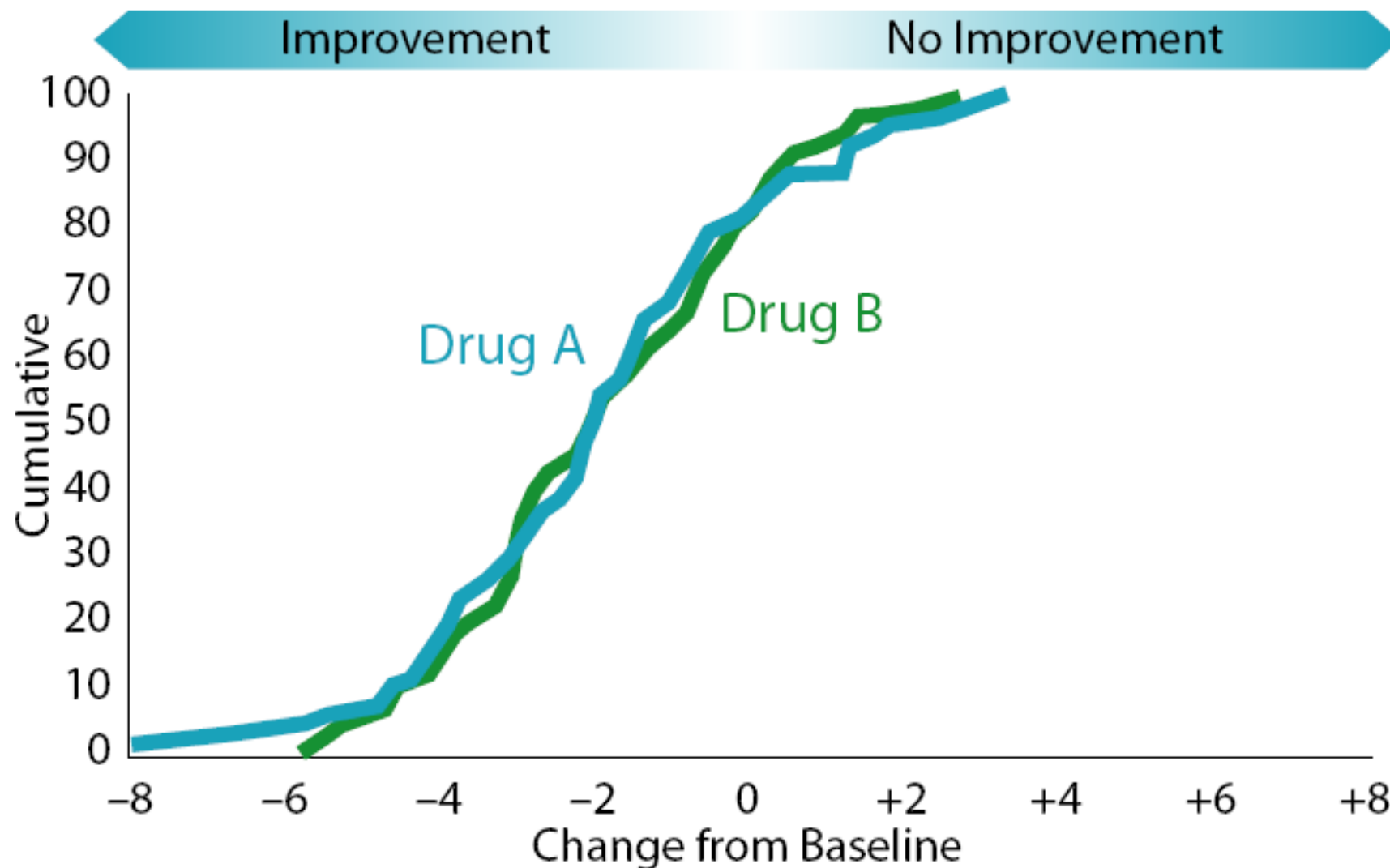


- Display a continuous plot of the percent change (or numeric change) from baseline on the horizontal axis and the cumulative percent of patients experiencing up to that change on the vertical axis
- Such a cumulative distribution of response curve – one for each treatment group – would allow a variety of response thresholds to be examined simultaneously and collectively, encompassing all available data
- Kolmogorov-Smirnov test can be used to test whether two empirical distributions are different

# Illustrative Cumulative Distribution Function: Experimental Treatment (solid line) better than Control Treatment (dash line) -- Negative changes indicate improvement

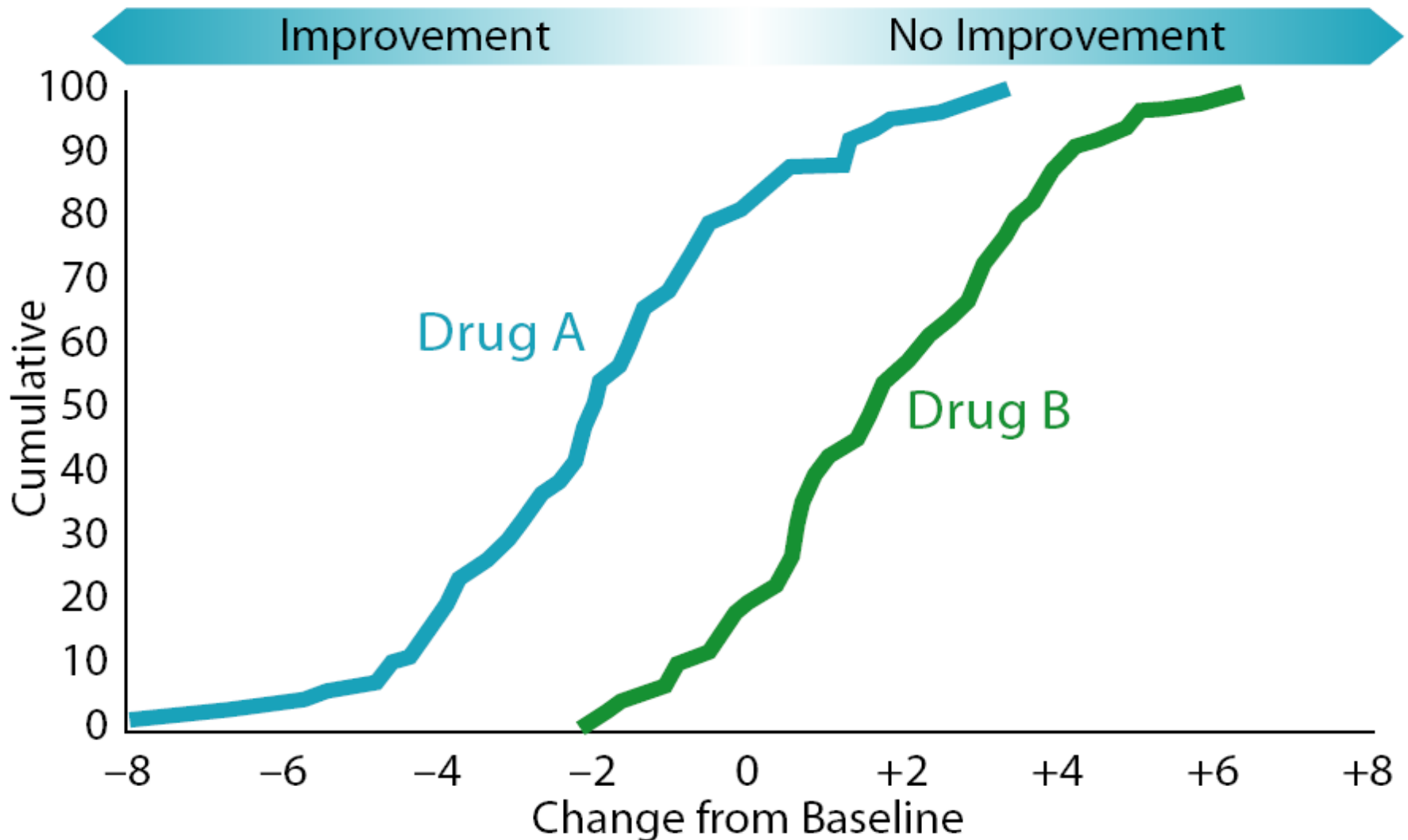


# CDF results that do not demonstrate the comparative efficacy of Drug A or Drug B





# Better Result for Demonstrating the Efficacy of Drug A over Drug B



# Aricept® Label from 10/13/2006

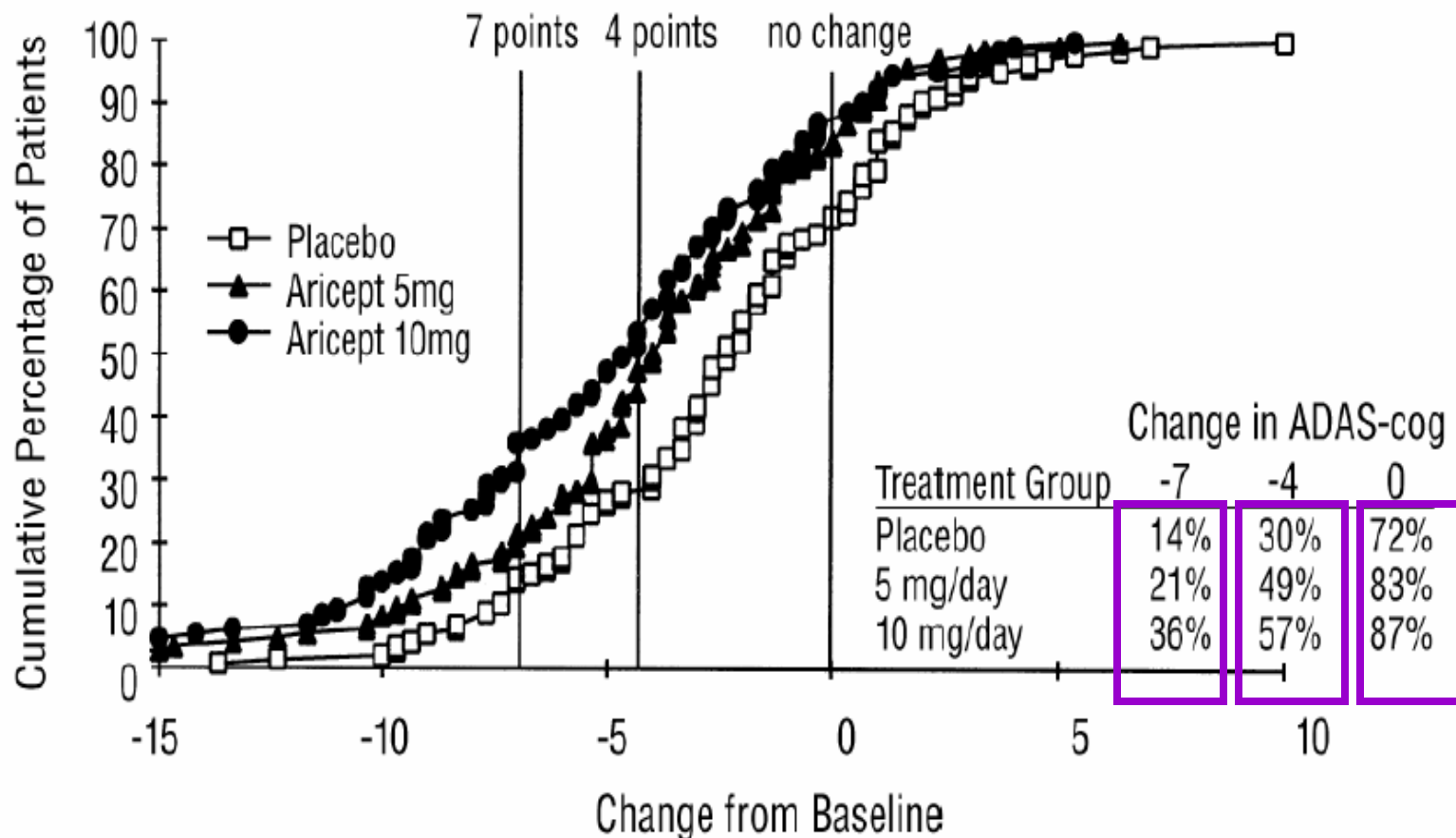
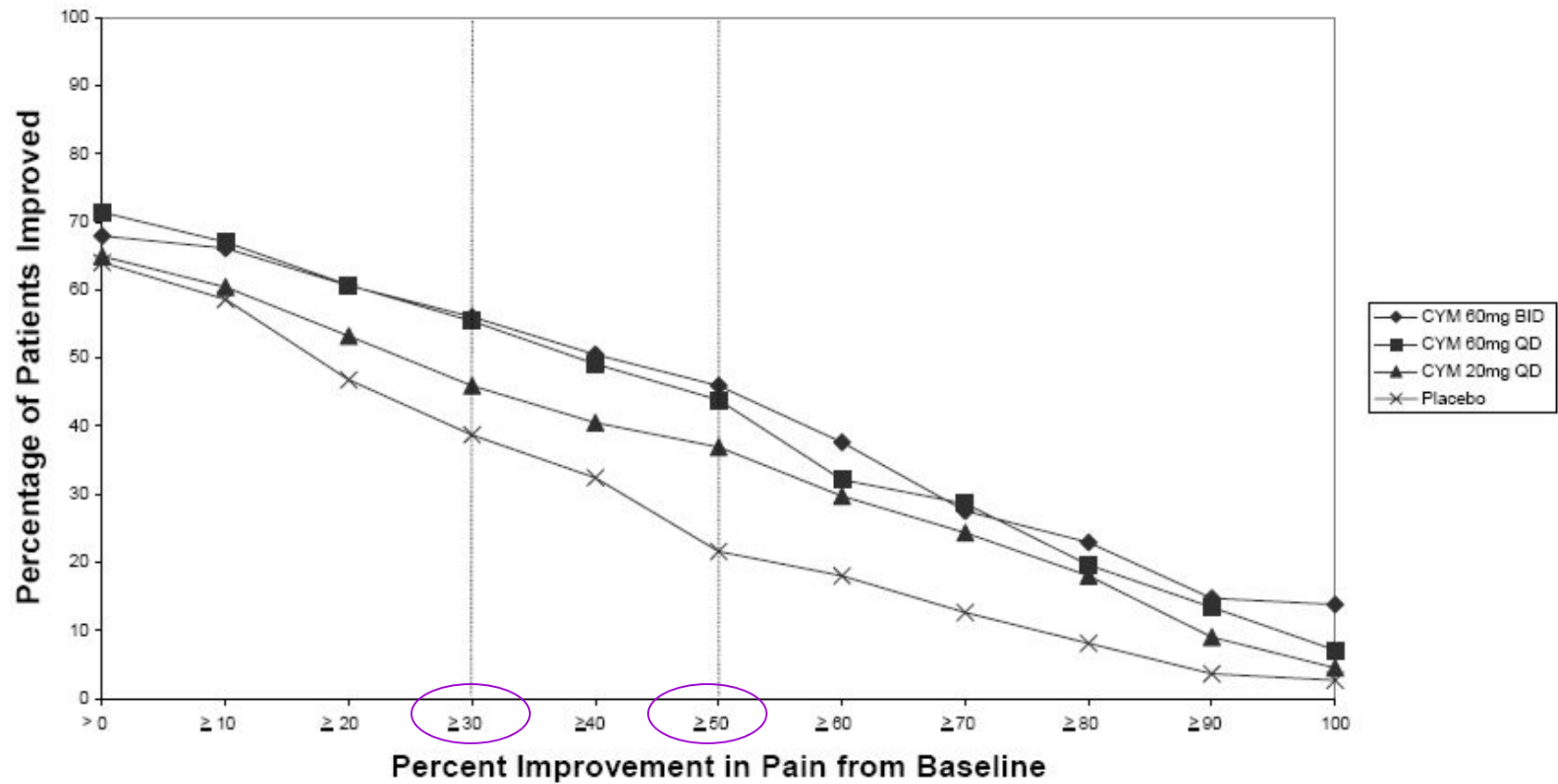


Figure 5. Cumulative Percentage of Patients with Specified Changes from Baseline ADAS-cog Scores. The Percentages of Randomized Patients Within Each Treatment Group Who Completed the Study Were: Placebo 93%, 5 mg/day 90% and 10 mg/day 82%.

# Cymbalta® Label from 11/19/2009



(x-axis reversed)



**Figure 1: Percentage of Patients Achieving Various Levels of Pain Relief as Measured by 24-Hour Average Pain Severity - Study 1**

# **Content-based Interpretation: Variation of Anchor-based Approach**

# Content-based Interpretation



- Uses a representative (anchor) item on multi-item PRO, along with its response categories, internal to the measure itself
- Item response theory
- Logistic models with binary or ordinal outcomes
- Observed proportions

# National Eye Institute – Visual Function Questionnaire



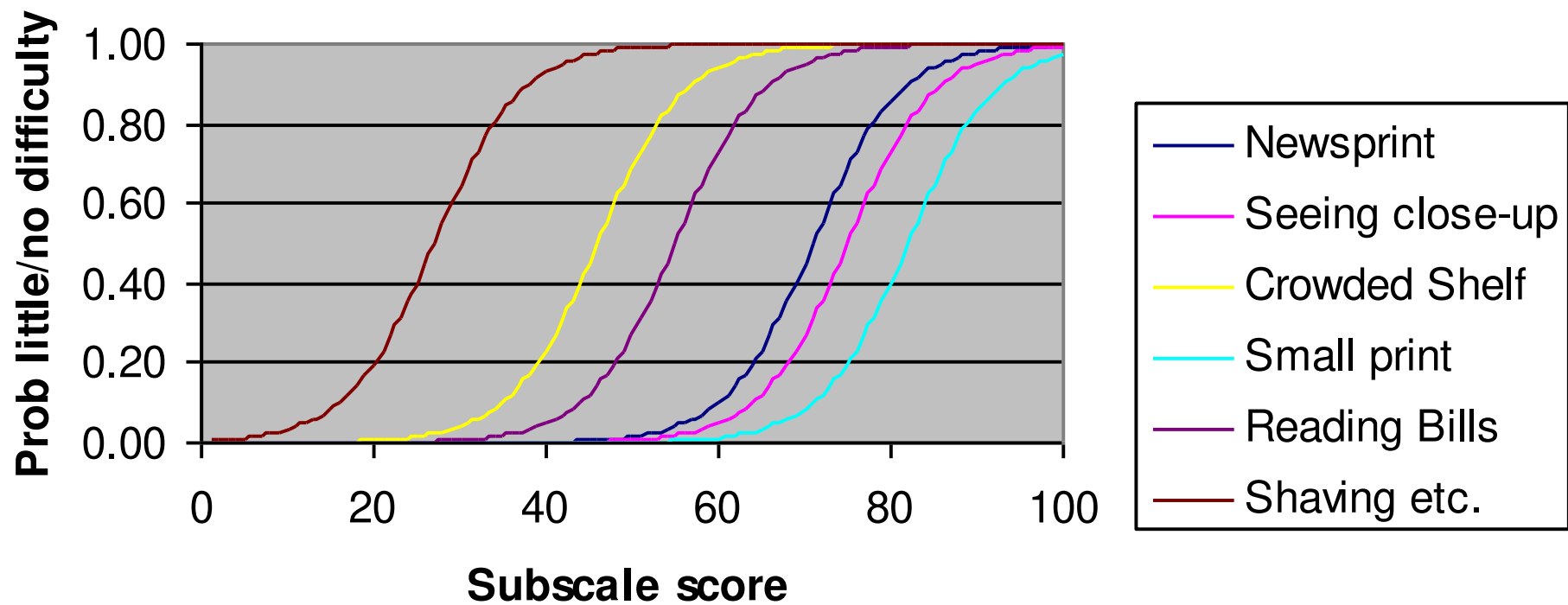
- NEI-VFQ assesses dimensions of vision-targeted functional health status that are most important to persons with chronic eye diseases
- Subscale scores range from 0 (worst) to 100 (best)
- Consider the near-vision subscale with its six items, each of which can be used as an anchor
  - Each item has five ordinal categories
  - No difficulty, little difficulty, moderate difficulty, extreme difficulty, not doing

# Example of Content-based Interpretation:

Thompson et al. Enhanced interpretation of instrument scales using the Rasch model. *Drug Information Journal* 2007; 41:541-550



## Near Vision Subscale from National Eye Institute-Visual Functioning Questionnaire



# Distribution-based Methods

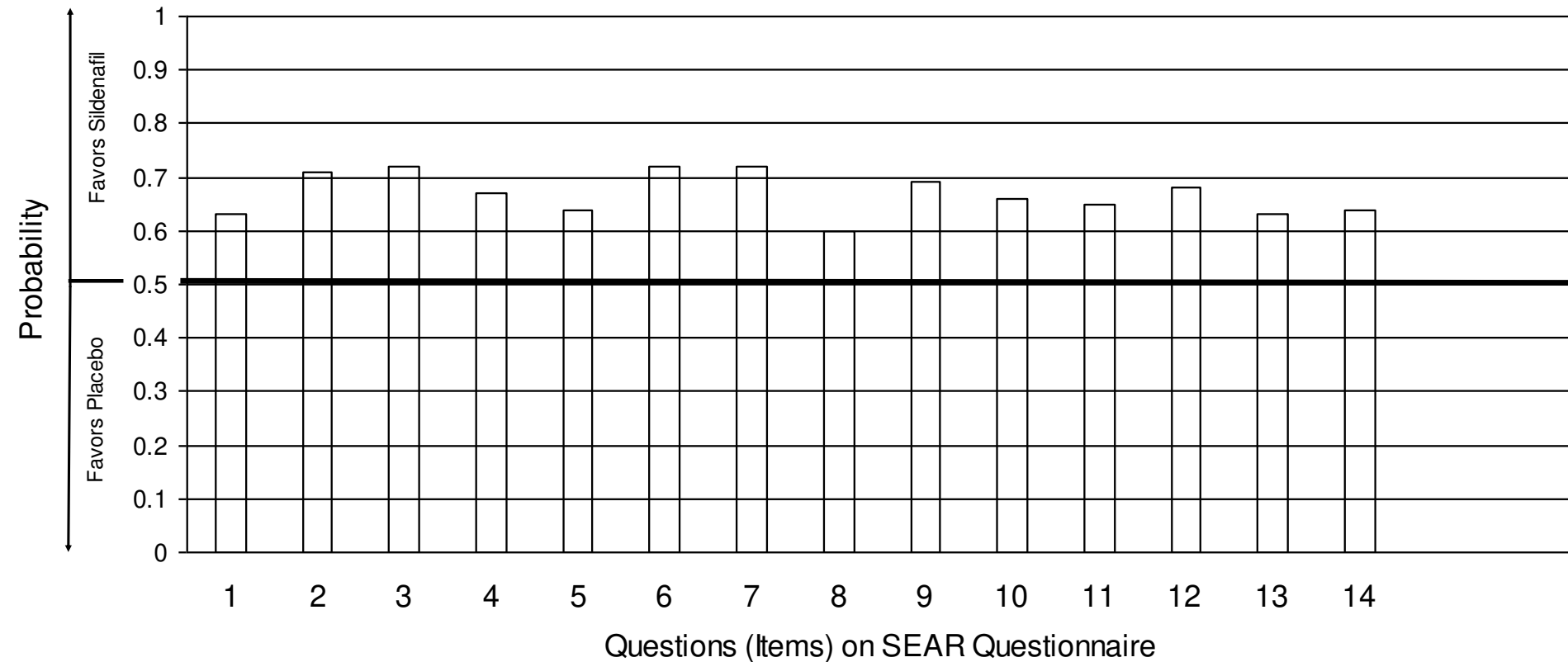


# Example: Probability of Relative Benefit



- Cappelleri et al. *BJU International* 2008; 101:861-866.
- Two 12-week, double-blind, placebo-controlled, flexible-dose sildenafil trials on Self-Esteem And Relationship (SEAR) questionnaire in men with erectile dysfunction
- Difference (sildenafil versus placebo) in SEAR from baseline to week 12 was evaluated with a Wilcoxon rank-sum test using rdit analysis

# Example: Probability of Relative Benefit



- All p values < 0.001
- Across all items, average probability was 0.67 (standard deviation of 0.04)

# Part 3: CDER's use of responder analyses and cumulative distribution functions

**Lisa A. Kammerman**

**Office of Biostatistics**

**Center for Drug Evaluation and Research**

**Food and Drug Administration**

## ***SECOND ANNUAL PATIENT-REPORTED OUTCOME (PRO) CONSORTIUM WORKSHOP***

**March 15, 2011 ■ Silver Spring, MD**

**Co-sponsored by**



# Disclaimer

The views expressed in this presentation do not necessarily represent those of the U.S. Food and Drug Administration

# Role of PRO as an endpoint



1. Primary evidence of efficacy?
2. Secondary endpoint?
  - Pre-specify
  - Report findings, regardless of statistical significance and direction

# Primary evidence of efficacy



- Instruments
  - Pain
  - Depression
  - Alzheimer's
- Diary – counting events
  - Hot flushes
  - Urinary incontinence
  - Irritable bowel syndrome with constipation

# Example of “Responder” used as a Primary Outcome

# Amitiza



- Indication
  - The treatment of irritable bowel syndrome with constipation (IBS-C) in women  $\geq 18$  years old
  - The treatment of chronic idiopathic constipation in adults.
- Initial approval – 2006
- Source: Drugs@FDA ([www.fda.gov](http://www.fda.gov))



# Amitiza (IBS-C)



- Primary endpoint, assessed weekly
  - Global symptom relief question, 7-point scale
  - “How would you rate your relief of IBS symptoms (abdominal discomfort/pain, bowel habits, and other IBS symptoms) over the past week compared to how you felt before you entered the study?”

# Amitiza (IBS-C)

- Primary Efficacy Analysis
  - Compared “Overall Responders”
- Definition of an “Overall Responder”
  - A “monthly responder” in at least 2 of the 3 months on study.
  - A “monthly responder” -- “significantly relieved” for at least 2 weeks of the month or at least “moderately relieved” in all 4 weeks of that month.

# Amitiza (IBS-C)

- Definition of a “Non-responder” (monthly)
  - “moderately worse” or “significantly worse” relief
  - An increase in rescue medication use
  - Discontinued due to lack of efficacy, were deemed non-responders.

# Amitiza (IBS-C)

## “Overall responders”

- Treatment differences between the placebo and Amitiza groups were statistically significant. Treatment effect ~ 7%
- Study 1
  - 13.8% of patients in the Amitiza 8 mcg group
  - 7.8% of patients in the placebo group
- Study 2
  - 12.1% of patients in the Amitiza 8 mcg group
  - 5.7% of patients in the placebo group
- What do the results mean for a patient coming into a provider's office?

# Amitiza (IBS-C)

## Comments

- Patient global
- Definition of responder seems ad hoc
- Irritable Bowel Syndrome Draft Guidance (March 2010) addresses clinical study design issues, including the definition of endpoints

# Irritable Bowel Syndrome

## Draft Guidance (March 2010)



<http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM205269.pdf>

- A definition of a responder for use in an analysis of proportions for evaluation of the co-primary endpoints should be prospectively described in the protocol and statistical analysis plan. ***Statistical power calculations should be based on a predefined difference in proportions. The predefined difference that would be considered clinically meaningful should be discussed during protocol development with the review division.***

# Irritable Bowel Syndrome

## Draft Guidance (March 2010)



<http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM205269.pdf>:

- Proposes responder definitions for IBS-C and IBS-D
  - IBS-C: both pain intensity and stool frequency
  - IBS-D: both pain intensity and stool consistency
  - Weekly responder
  - Overall responder – prespecified improvement in symptoms for at least 50 percent of the time.
  - Response should be observed at several points throughout the trial to establish sustained improvement
- Pain intensity responder
  - Decrease in worst abdominal pain in past 24 hours of at least 30 percent, compared with baseline

# Points to consider on the evaluations of medicinal products for the treatment of irritable bowel syndrome (March 2003)



[http://www.ema.europa.eu/docs/en\\_GB/document\\_library/Scientific\\_guideline/2009/09/WC500003187.pdf](http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2009/09/WC500003187.pdf):

The protocol should define a priori a responder in terms of a clinically meaningful change in the primary endpoints. The study should ***compare the proportion of patients who achieve the stipulated amount of improvement necessary to be qualified as a responder rather than a mean change in a score.*** An a priori specification of the time interval over which a responder or a response occurs should be included. This should be consistent over time but usually be towards the end of the trial.

It is acknowledged that the assessment of efficacy may depend on the specific characteristics of the drug and its intended use (e.g. on demand or continuous). It is recommended that for short-term studies of about 4 weeks duration, a ***positive response would require a pre-specified improvement in symptoms for at least 50% of the time.*** The study should include measures of change for each of the symptoms that was part of the entry criteria.



# “Responder” as an outcome

- Not necessarily a requirement
- Continuous data usually preferred
- Responder analyses can help interpret results
- “Points to consider on multiplicity issues in clinical trials” (September 2002) – useful guidelines

[http://www.ema.europa.eu/docs/en\\_GB/document\\_library/Scientific\\_guideline/2009/09/WC500003640.pdf](http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2009/09/WC500003640.pdf)

# Who is a “Responder”?



- There are two types:
  - Those who respond *positively*
  - Those who respond *negatively*
- A third group of subjects –  
Neither a positive responder nor a negative responder
- Most discussions seem to focus on the positive responders
- However, we need to know about those who are negatively affected by treatment. They too are responders!
- ICH E9: Uses “success” as an example of categorization

# Treatment effect vs. Patient-level effect



- Can clinical trials do both?
- Treatment effect (difference between groups)
  - Use for drug approval
  - Conceptually, the results tell us the proportion of patients expected to benefit
  - In risk-benefit setting, may still need to know how much of a difference is important
- Patient-level effect
  - What is the chance my patient will respond?
  - What about the placebo effect?

# Significance testing vs. effect size estimation



- M. Borenstein, “Hypothesis testing and effect size estimation in clinical trials”, Annals of Allergy, Asthma & Immunology, 1997.
- Compelling effect:
  - Size of the effect
  - Size of the sample
  - Stringency of evidence supporting an effect – Risk- Benefit
- Significance testing: Does the treatment have *any* effect?
- Power analysis -- the effect size is theoretical
- Effect size observed in the study is our “best guess” about the size of the effect in the population
- Effect size observed in any given study will be somewhat *lower or higher than the true effect size*. 50-50 chance.

# Example of Cumulative Distribution Function in Labeling

# Egrifta



- Approved November 2010
- Indicated for the reduction of excess abdominal fat in HIV-infected patients with lipodystrophy
- Source: Drugs@FDA ([www.fda.gov](http://www.fda.gov))

# Egrifta (tesamorelin for injection)



## Endpoints in labeling

- Visceral adipose tissue
- IGF-I
- IGFBP-3
- Weight
- Waist circumference
- *Degree of distress associated with belly appearance*

# Egrifta

## Degree of Distress



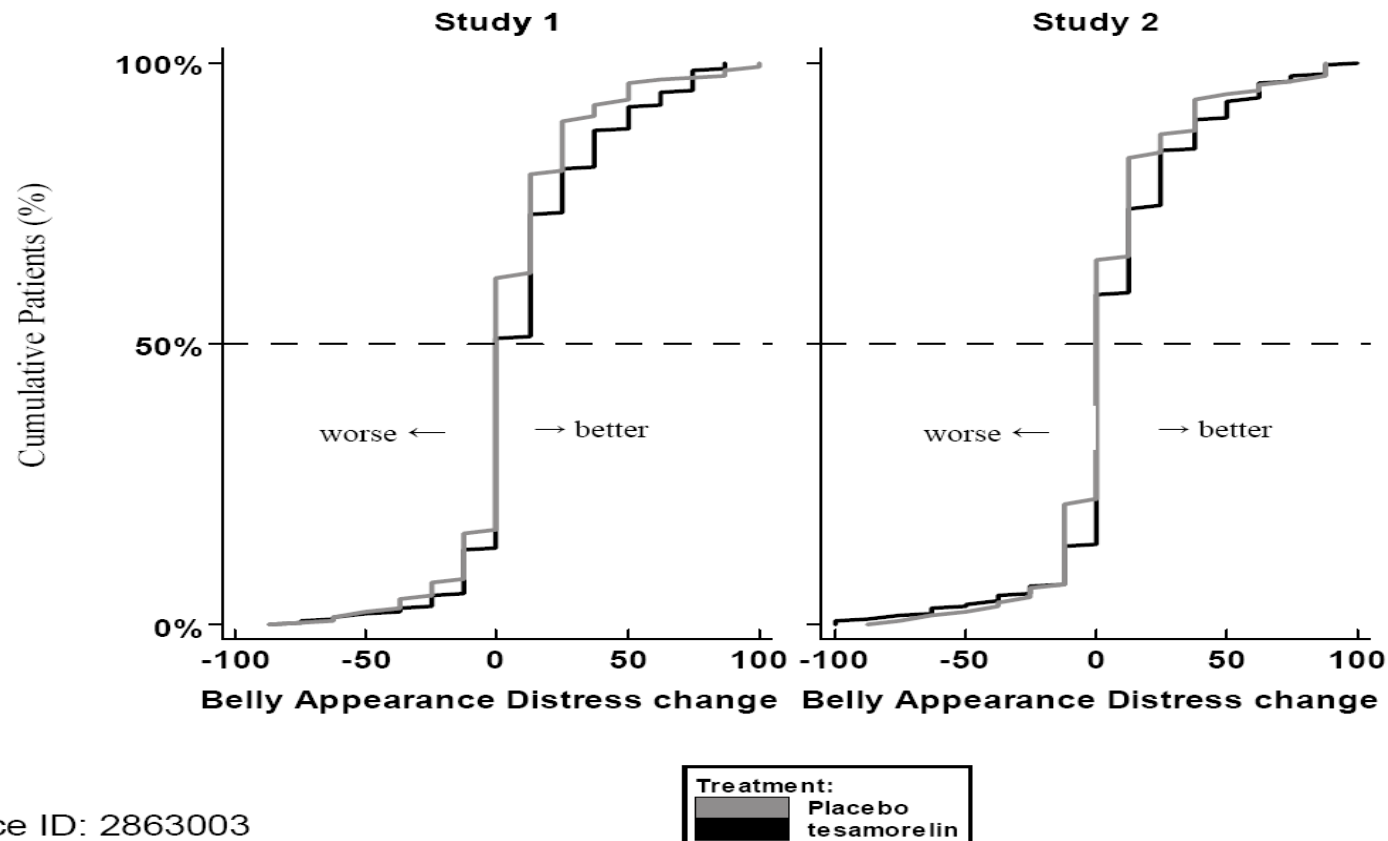
- Degree of distress associated with their belly appearance
- 9-point scale
- Transformed to a score from 0 to 100
  - 0 (extremely upsetting and distressing)
  - 50 (neutral – no feeling either way)
  - 100 (extremely encouraging)
- Positive change from baseline = improvement



# Egrifta – CDF

(Belly Appearance Distress, label)  
Change from baseline to 26 weeks

**Figure 1. Cumulative Distribution of Response for Belly Appearance Distress**



ference ID: 2863003

Source: [http://www.accessdata.fda.gov/drugsatfda\\_docs/label/2010/022505s000lbl.pdf](http://www.accessdata.fda.gov/drugsatfda_docs/label/2010/022505s000lbl.pdf)

# Guidance: Clinical Studies Section of Labeling for Human Prescription Drug and Biological Products – Content and Format



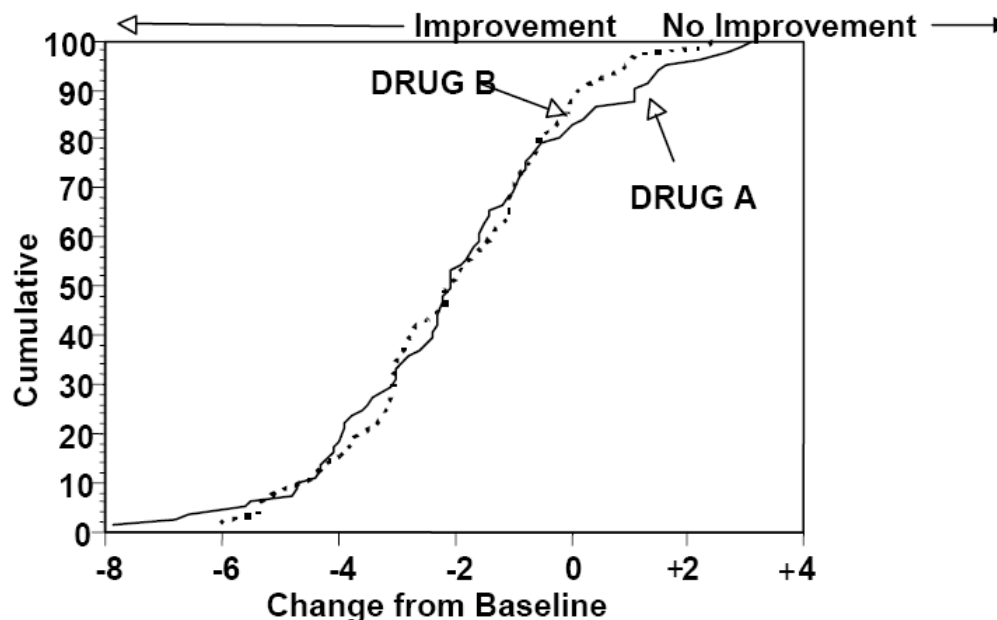
- <http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/ucm075059.pdf>
- Goals
  - What studies to include
  - How to describe the studies
  - How to present the data, including presentation in tables and graphs
- “... providing individual subject data for all treatment groups can be a useful alternative for describing the clinical effect of a drug. This can be done by including a graphical presentation of the distribution or cumulative distribution of responses among individual subjects (see Appendix for examples ... ). Individual data can also be presented as categorical outcomes” (page 8)

# Guidance: Clinical Studies Section of Labeling for Human Prescription Drug and Biological Products – Content and Format



- A cumulative distribution plot may need a footnote and additional text in the body of the label describing how to read the graph. For example, the following text could accompany the graph shown above: “Approximately 50% of the patients in each group had a decrease of at least 2 mg/dL at endpoint.”

Cumulative Distribution Plot of Change From Baseline for Study 1



This graph shows the percentage of subjects (y-axis) attaining a change from baseline less than or equal to the value on the x-axis. A curve that shifts to the left indicates a better response.

Source:

<http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/ucm075059.pdf>

# Conclusions

- PRO Endpoint Considerations
  - Consider the study objective and the role of the PRO endpoint
  - Plan the design and analysis
  - Testing vs Estimation vs Powering
- “Responders”
  - Positive vs Negative vs Neither
  - Appears in many guidance documents and draft guidance documents
- Responder analyses
  - Helpful in interpretation of study results
  - Not necessarily needed as a primary endpoint
  - Continuous variables are often preferred

# Commentary

## Panel Questions & Answers

**Moderator:**

**David Reasner**

*Sunovion Pharmaceuticals*

# Commentary (1)



- The conceptual framework and endpoint model will not only guide development of the PRO but ensure the clinical relevance of the responder definition.
- Beyond content validity, rely on convergent, discriminant, and predictive validity (Evidence based on relations to other variables) to anchor the clinical interpretation of the responder definition.

## Commentary (2)

- The cumulative distribution function allows simultaneous review of all responder definitions.
- The CDF anchors the study in the literature and avoids post-hoc bias while enabling intuitive, graphical data analysis.
- Examination of the CDF is amenable to confirmation by evidence and a toolkit of statistical methods.

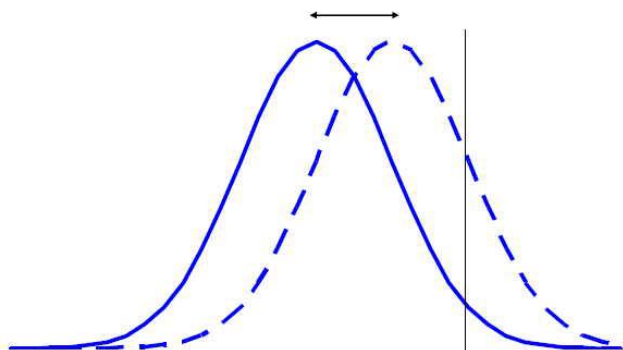
# Commentary (3)



- The endpoint model will declare the PRO for regulatory consideration as primary evidence or secondary objective.
- Control of the Type I error rate leads to an intuitive process by which hypothesis testing of the PRO ( $H_0$  rejected) is followed by an exploration of response that provides clinical characterization of the treatment effect.
- Nonetheless, the regulator is left to ponder the clinical relevance of the difference in proportions and how the label can inform the physician in their treatment of an individual patient.

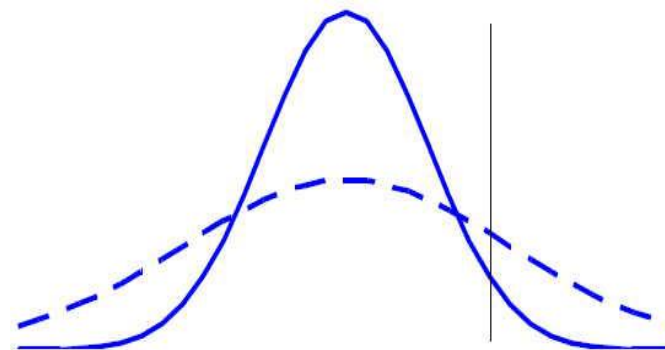


# Responder analyses and the assessment of a clinically relevant treatment effect (Snapinn & Jiang, 2007)



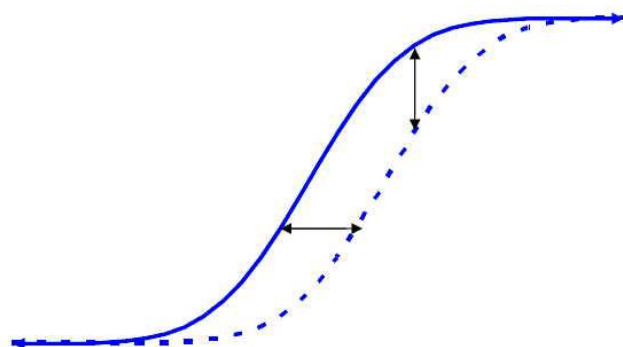
**Figure 1**

Distribution of Outcomes in the Experimental Group (Dashed Line) Has Greater Mean Value Than Control Group (Solid Line) And Greater Proportion of Responders.



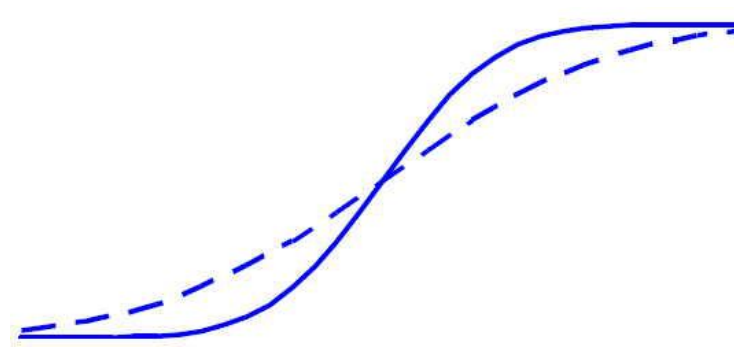
**Figure 3**

Distribution of Outcomes in the Experimental Group (Dashed Line) Has Equal Mean Value to That of the Control Group (Solid Line), But a Greater Proportion of Responders.



**Figure 2**

Cumulative Distribution Functions for Two Treatment Groups When the Outcome Variable Distributions Differ in Mean But Not Variance; Horizontal Displacement Represents the Mean Difference and Vertical Displacement Represents the Difference in Response Rates.



**Figure 4**

Cumulative Distribution Functions for Two Treatment Groups When the Outcome Variable Distributions Have the Same Mean But Different Variance.

# Moderator Notes



- The homogeneous normal population is a useful mischaracterization.
- A difference in group means implies a difference in responder rates (see Figure 1 and Figure 2).
- Differences between treatment (e.g., heterogeneity of variance) lead to an interaction between the responder threshold and the relative response (see Figure 3 and Figure 4).
- Qualitatively different patient subpopulations may create “breaks” in the PRO score continuum that surface a potential responder definition in the exploratory phase.
- Breaking adjacent scores on a true continuum raises the “dichotomization” problem.
- Absolute (7 point remission criterion) versus relative (50% responder criterion) operate differently in the study sample and the population.
- As with PRO development, responder analysis is in transition from clinical convention to evidence-based decision; the basis for approval is broader than the summation in the label.

# Responder analyses and the assessment of a clinically relevant treatment effect



Responder analyses and the assessment of a clinically relevant treatment effect

Steven Snapinn and Qi Jiang

Trials 2007, 8:31

## Abstract

- Ideally, a clinical trial should be able to demonstrate not only a statistically significant improvement in the primary efficacy endpoint, but also that the magnitude of the effect is clinically relevant. One proposed approach to address this question is a responder analysis, in which a continuous primary efficacy measure is dichotomized into "responders" and "non-responders." In this paper we discuss various weaknesses with this approach, including a potentially large cost in statistical efficiency, as well as its failure to achieve its main goal. We propose an approach in which the assessments of statistical significance and clinical relevance are separated.