RTI HEALTH SOLUTIONS®

Defining a Responder: Implementing the Patient-Reported Outcome (PRO) Guidance Recommendations

Lori McLeod,[1] Ron D. Hays,[2] Susan Martin,[1] and Sheri Fehnel[1]

[1]RTI Health Solutions

[2]UCLA Department of Medicine/Division of General Internal Medicine & Health Services Research and RAND

RTI Health Solutions

Research Triangle Park, NC, US
+1.800.262.3011

Ann Arbor, MI, US
+1.734.213.5372

Barcelona, Spain
+34.93.241.77.66

Lund, Sweden
+46.46.14.7040

Manchester, UK
+44(0)161.232.3400

Sheffield, UK
+44(0)114.213.3390

Waltham, MA. US
+1.781.434.1700

www.rtihs.org

E-mail: rtihealthsolutions@rti.org

ISPOR 15th Annual International Meeting

Atlanta, Georgia

May 19, 2010

1:45 PM – 2:45 PM

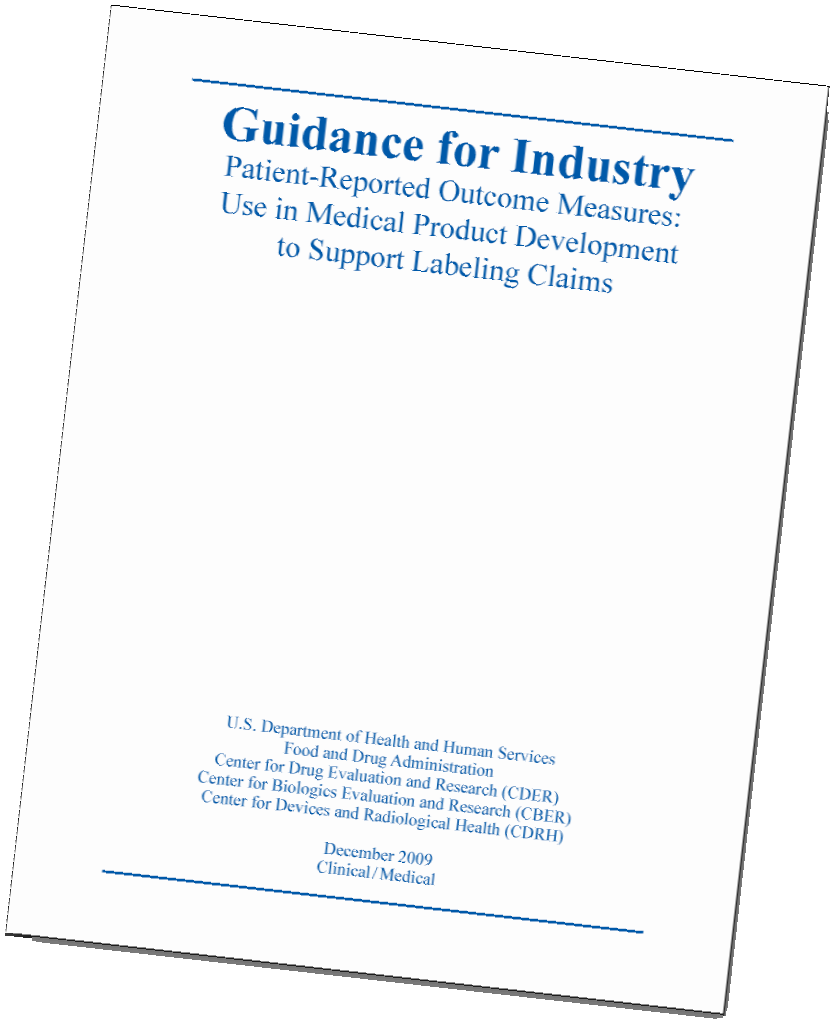
LEADING RESEARCH…
MEASURES THAT COUNT

Agenda

- Introductions
- Responder Overview from the PRO Guidance
- Minimal Important Difference
- Responder
- Cumulative Distribution Function
- Qualitative Methods for Interpretation
- Examples and Discussion

FDA PRO Guidance

- Draft FDA PRO Guidance: published February 2006
- Final FDA PRO Guidance: published December 2009
- Guidance developed by the SEALD group within the Office of New Drugs (OND) at FDA
- SEALD serves as an advisory group to all reviewing divisions



Final FDA PRO Guidance

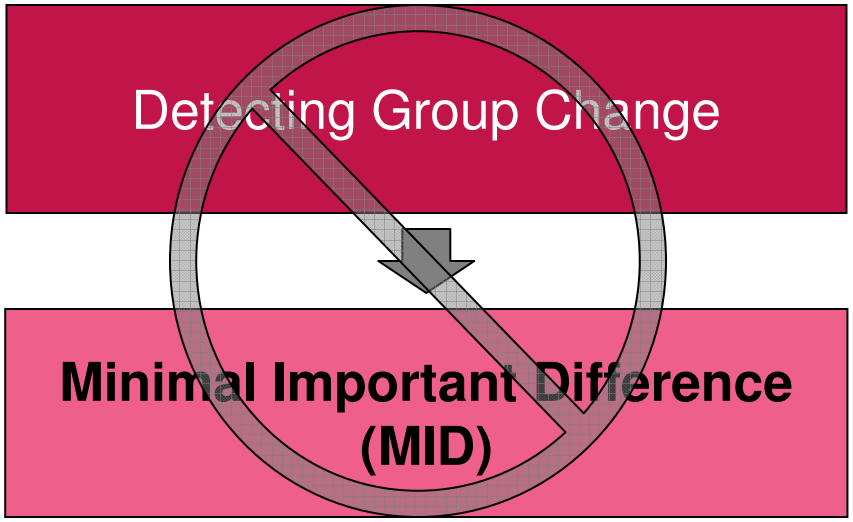
- The Final FDA PRO guidance describes:
  - How the FDA evaluates PRO instruments used as endpoints in clinical trials
  - The FDA's current thinking on how sponsors can develop and use study results measured by PRO instruments to support claims in approved product labeling
  - How the FDA evaluates instruments for their usefulness in measuring and characterizing the benefit of medical product treatment

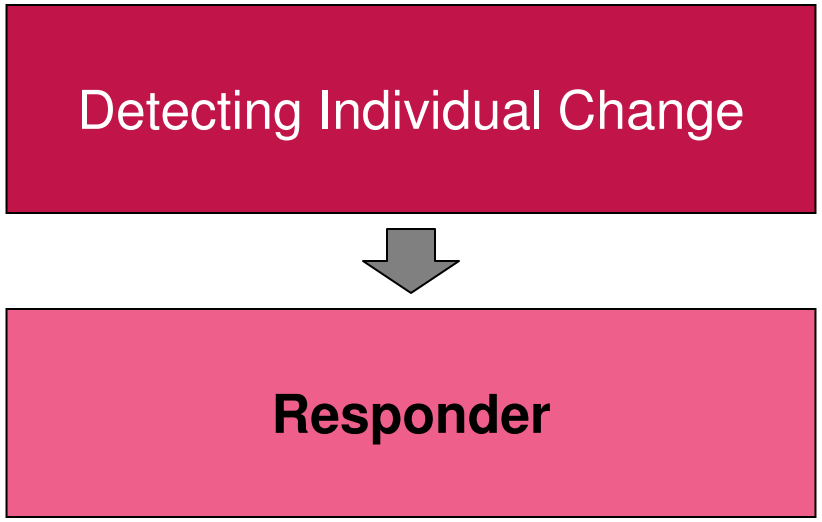
http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM193282.pdf

Draft Guidance

Detecting Group Change	Detecting Individual Change
Minimal Important Difference (MID)	**Responder**
"Amount of difference or change observed that would be interpreted as a treatment benefit." Draft FDA PRO Guidance, 2006, Glossary, p. 31	"Change in score that would be clear evidence that an individual patient experienced a treatment benefit." Draft FDA PRO Guidance, 2006, p. 17

Final Guidance





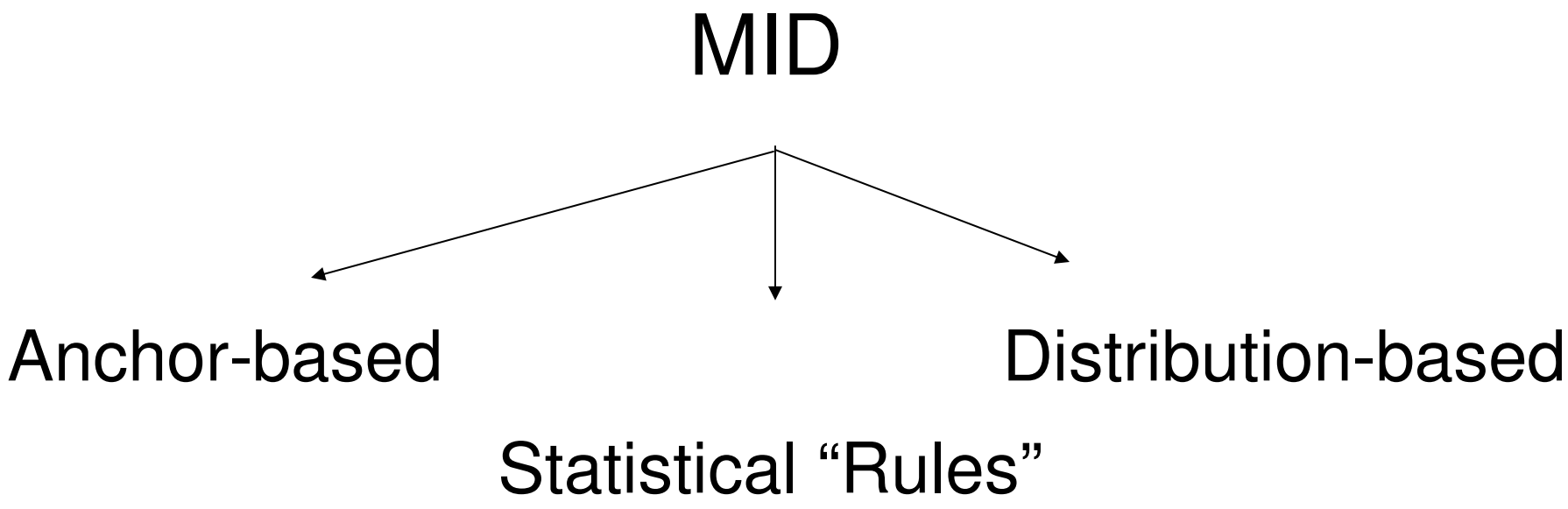
"A score change in a measure, experienced by an individual over a predetermined time period that has been demonstrated in the target population to have a significant treatment benefit."

Final FDA PRO Guidance, 2009, Glossary, p. 33

MID Versus Responder

- MID
  - Typically defined as the smallest difference that is considered clinically important or implies treatment benefit
  - Used as a benchmark for evaluating mean differences between treatments
- Responder
  - Typically larger than the MID
  - Used to categorize patients as having responded to treatment
- When effect sizes are small, the distribution of responses for treatment and placebo groups might be more informative than MID

Defining an MID: Common Approaches



Defining an MID: Anchor-based Approaches

- Patient- or Physician-based judgment using “anchor”
  - Categorical rating of change between baseline and end-of-study
  - MID = Mean change score on the PRO for those choosing “a little better” on the anchor question

Since the start of the study, how would you describe the change (if any) in <<symptom X, severity of condition>>?

- ❑ Much better
- ❑ Moderately better
- ❑ A little better ← *MID anchor*
- ❑ No change
- ❑ A little worse
- ❑ Moderately worse
- ❑ Much worse

Defining an MID: Statistical "Rules"

- Commonly used "rule of thumb"
  - 0.5 change per 7-category ordinal item

	Scale W	Scale X	Scale Y	Scale Z
Number of Items	5	10	15	20
MID Estimate	2.5	5	7.5	10

Juniper et al., 1994

Defining an MID: Distribution-Based

- Distribution-based methods categorize the size of a change on a PRO measure by considering the distribution of the scores themselves.
  - Change of 0.5*SD_{bl}
  - Effect size (ES)
    - Change of 0.2 ES = small response
  - Standard error of measurement (SEM) = $SD_{bl} \cdot \sqrt{1 - r_{xx}}$

Note: SD_{bl} = standard deviation at baseline
r_{xx} = reliability

Interpretation of PRO Results

- Methods for responder definition in final guidance
  - Primary method
    - Anchor-based
  - Supportive method
    - Distribution-based
- Statistical significance of individual change
- Final guidance also describes an alternative approach to use of a single responder definition
    - Cumulative distribution function (CDF)
- FDA has also requested direct patient input (qualitative research)

Defining a Responder: Guidance-Recommended Method

- Anchor-based methods use a relevant measurement that is easier to interpret than scores on the PRO.
- For example:
  - Mean change scores on the PRO for patients who have a 50% reduction in Hamilton Depression (HAM-D) scores
  - Mean change scores on the PRO for patients reporting an improvement in a patient-reported global impression of change question

Since the start of the study, how would you describe the change (if any) in <<symptom X, severity of condition>>?

- ❑ Much better
- ❑ Moderately better ← *Responder anchor*
- ❑ A little better ← *MID anchor*
- ❑ No change
- ❑ A little worse
- ❑ Moderately worse
- ❑ Much worse

Defining a Responder: Additional Methods

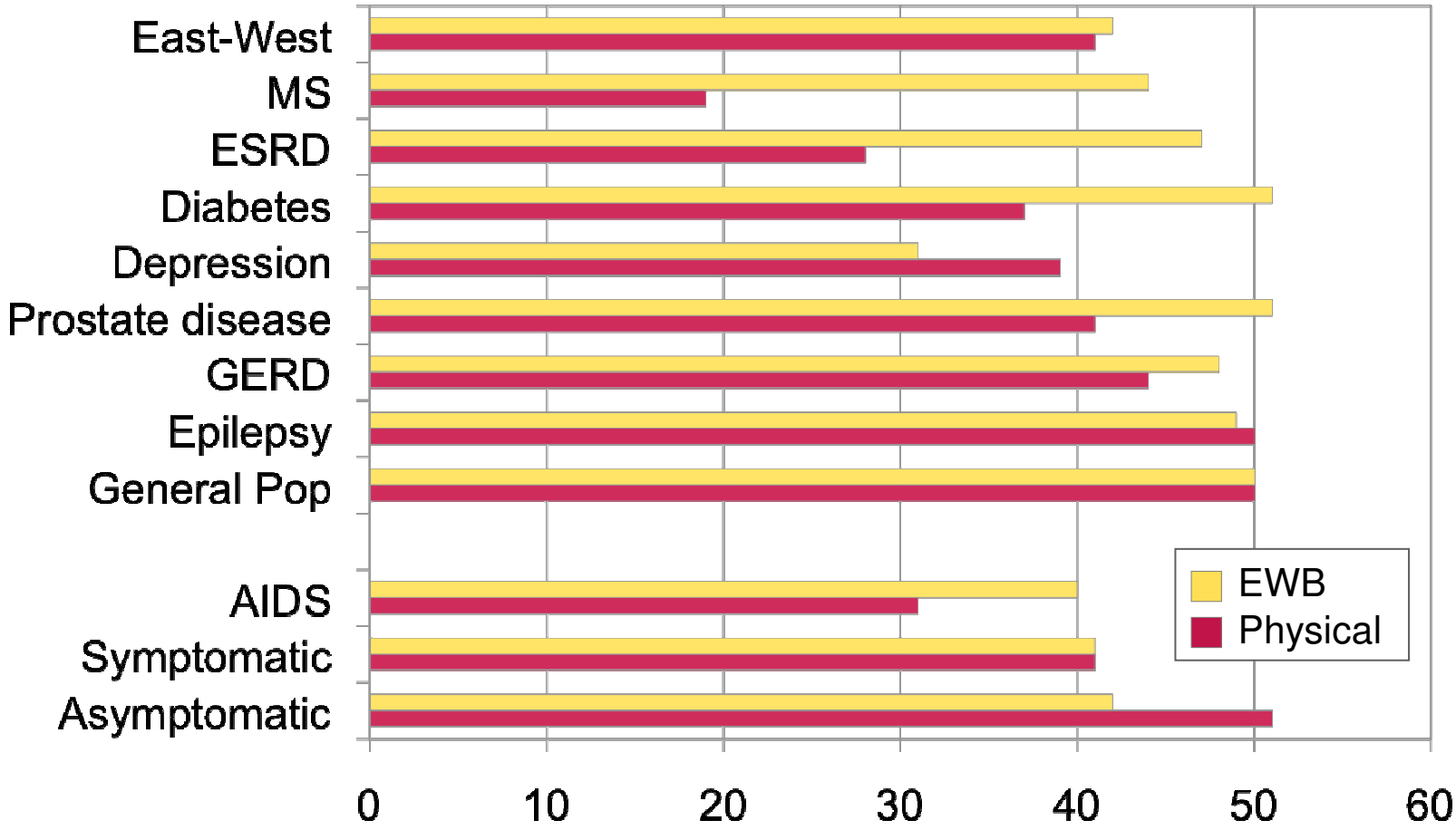
- Distribution-based methods categorize the size of a change on a PRO measure by considering the distribution of the scores themselves.
  - Change of 0.5*SD_{bl}
  - Effect size (ES)
    - Change of 0.2 ES = small response
    - Change of 0.5 ES = medium response
    - Change of 0.8 ES = large response
  - Standard error of measurement (SEM) = $SD_{bl} \times \sqrt{1 - r_{xx}}$

Note: SD_{bl} = standard deviation at baseline
r_{xx} = reliability

SEM and Significance of Individual Change

- Standard Error of Measurement (SEM) noted as one type of “distribution-based” method
- SEM actually used to estimate 95% confidence interval around an individual’s score
  - Observed score +/- (1.96 * SEM)
- Significance of individual change should be used to define responders

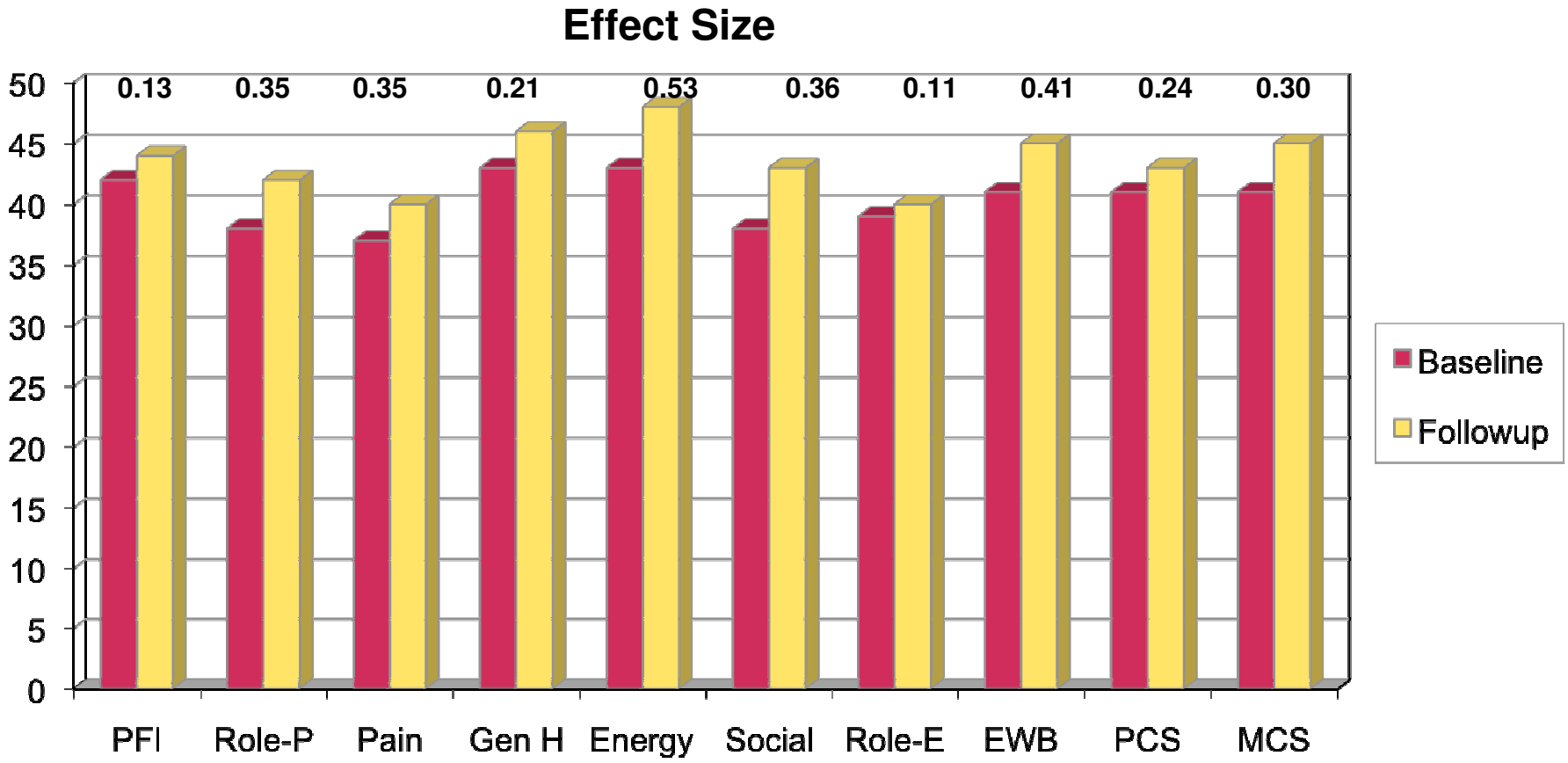
Physical Functioning and Emotional Well-Being at Baseline for 54 Patients at UCLA-Center for East West Medicine



ESRD = end-stage renal disease; GERD = gastroesophageal reflux disease; MS = multiple sclerosis.

Hays et al., 2000

Change in SF-36 Scores Over Time



Energy = Energy/Fatigue; EWB = Emotional Well-being; Gen H=General Health; MCS =Mental Component Summary; Pain = Bodily Pain; PCS = Physical Component Summary; PFI = Physical Functioning; Role-E = Role-Emotional; Role-P = Role-Physical; Social = Social Functioning

t-test for Within-Group Change

$$\frac{X_d}{SD_d / \sqrt{n}}$$

X_d = is mean difference, SD_d = standard deviation of difference

Significance of Group Change (T-scores)

	Change	t-test	prob.
PF-10	1.7	2.38	.0208
RP-4	4.1	3.81	.0004
BP-2	3.6	2.59	.0125
GH-5	2.4	2.86	.0061
EN-4	5.1	4.33	.0001
SF-2	4.7	3.51	.0009
RE-3	1.5	0.96	.3400 ←
EWB-5	4.3	3.20	.0023
PCS	2.8	3.23	.0021
MCS	3.9	2.82	.0067

Defining a Responder: Reliable Change Index (RCI)

$$\frac{X_2 - X_1}{(\sqrt{2})(SEM)}$$

$$SEM = SD_{bl} \times \sqrt{1 - r_{xx}}$$

Note: SD_{bl} = standard deviation at baseline
r_{xx} = reliability

Amount of Change in Observed Score Needed for Significant Individual Change

	RCI	Effect size	Cronbach's alpha
PF-10	8.4	0.67	0.94
RP-4	8.4	0.72	0.93
BP-2	10.4	1.01	0.87
GH-5	13.0	1.13	0.83
EN-4	12.8	1.33	0.77
SF-2	13.8	1.07	0.85
RE-3	9.7	0.71	0.94
EWB-5	13.4	1.26	0.79
PCS	7.1	0.62	0.94
MCS	9.7	0.73	0.93

Significant Change for 54 Cases

	% Improving	% Declining	Difference
PF-10	13%	2%	+ 11%
RP-4	31%	2%	+ 29%
BP-2	22%	7%	+ 15%
GH-5	7%	0%	+ 7%
EN-4	9%	2%	+ 7%
SF-2	17%	4%	+ 13%
RE-3	15%	15%	0%
EWB-5	19%	4%	+ 15%
PCS	24%	7%	+ 17%
MCS	22%	11%	+ 11%

Response Interpretation Alternative

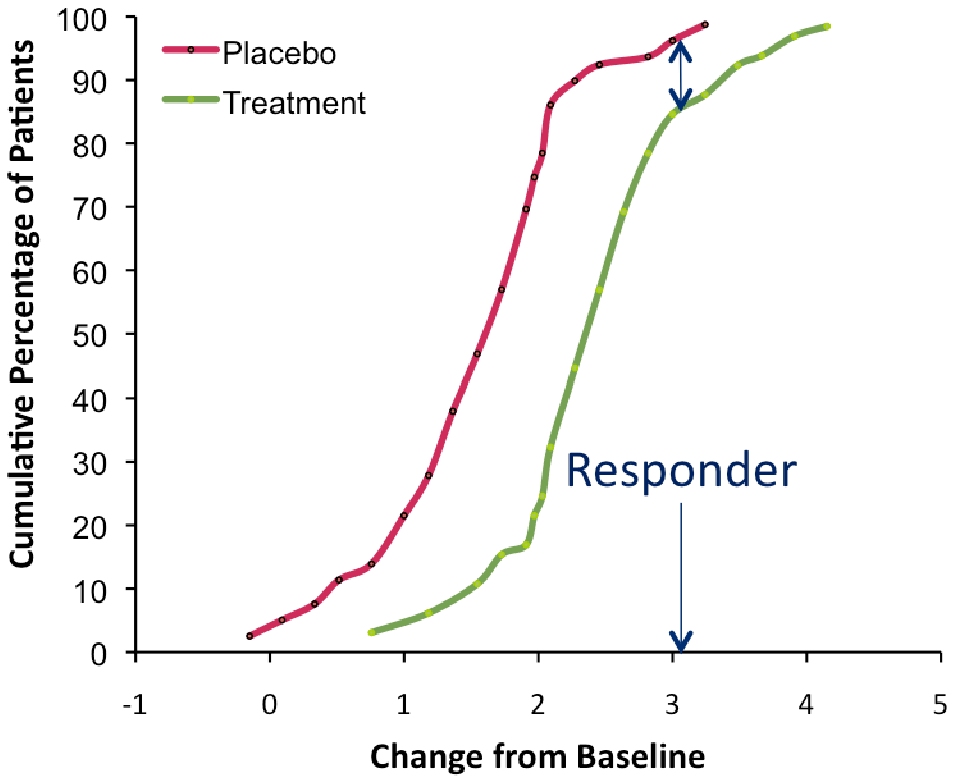
- "Alternatively, it is possible to present the entire distribution of responses for treatment and control group, avoiding the need to pick a (specific) responder criterion. …A variety of responder definitions can be identified along the cumulative distribution of response curve."

Final FDA PRO Guidance, 2009, p. 25

(presenters' addition)

Response Interpretation Alternative

- A CDF graphs the cumulative frequency of change in response across the PRO response scale separately by treatment group
  - "This display type may be preferable to attempting to provide categorical definitions of responders" or selecting one definition (presenters' addition)

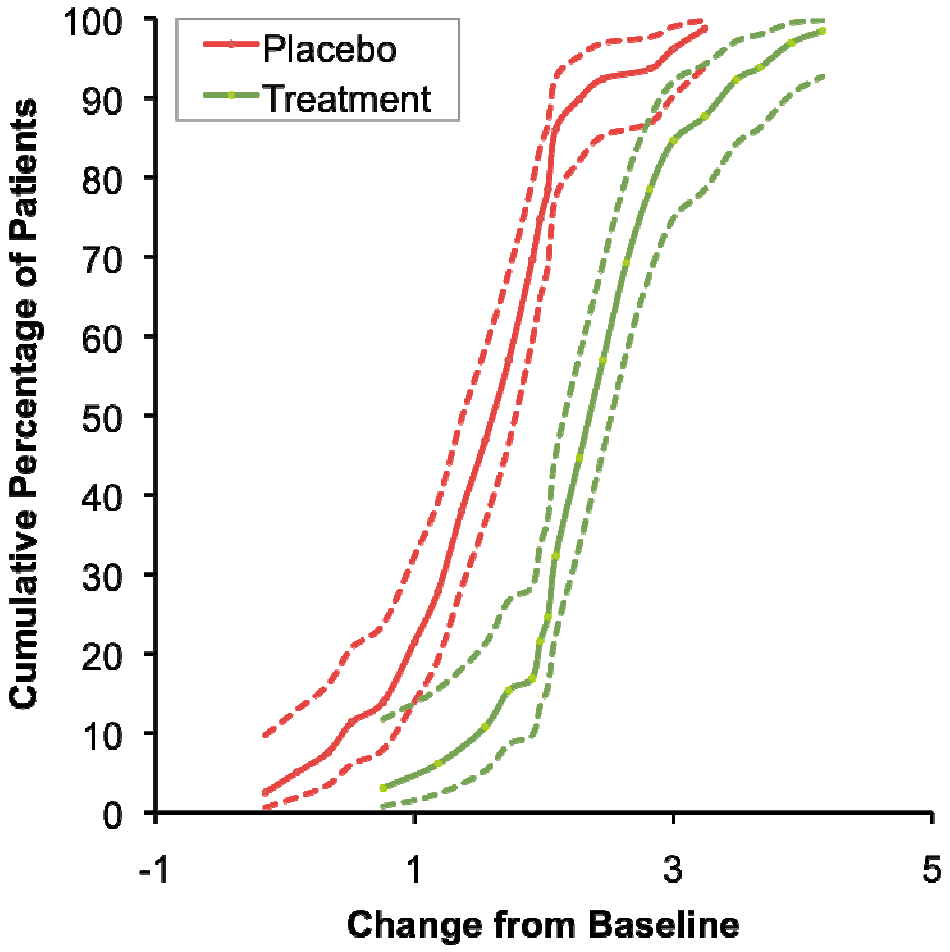


Final FDA PRO Guidance, 2009, p. 25

*Positive change indicates improvement

Potential Analyses Using CDF

- Are there more responders in the treatment group?
  - Chi-square tests at specific points
- Are the curves overlapping?
  - Kolmogorov-Smirnov (KSa) test
    - Riffenburgh, 1999
  - Tests of area under the curve
    - Farrar et al., 2006
  - Confidence bands
    - Diaz-Ramos et al.,1996
- External to Final FDA PRO Guidance



*Positive change indicates improvement

Defining a Responder: Qualitative Approaches

- Thus far we have explored quantitative approaches at the group and individual level.
- Eliciting patient input on changes that would constitute an MID or a response can complement these approaches and has been requested by the FDA.

Case Example - Asthma

- The concept of symptom-free days (SFD) is commonly used as a meaningful measure of treatment efficacy.
- The amount of additional SFD that either would be considered an MID or would define a response from the patient's perspective has not been established.

Qualitative Approach

- Identifying the patient cohort
  - Eligibility criteria should be similar to those required for entry into the clinical studies
  - Since probing on issues around response to medication, also required that patients had initiated a new treatment for their asthma in the last 6 months

Martin et al., 2010

Discussion Guide Development

- Important to establish the patient's current disease severity
  - How many days in a typical month do you currently experience no asthma symptoms?
- Need to present the concepts of responder and MID in terms patients are able to relate to
  - What number of additional days in 1 month with no asthma symptoms would be an important improvement for you? (Responder)
  - Thinking about that question again, what would be the fewest number of additional days with no asthma symptoms that you would still see as improvement (MID)

Demographic Characteristics (n = 11)

Characteristic	n
Gender	
Female	7
Male	4
Age, average years (range)	44 (29-59)
Race/ethnicity	
White (1 white participant was of Hispanic origin)	7
African American	4
Education	
High school diploma or GED	4
Some college	3
College degree	4

Patient Estimates of SFD Response

- Frequently, patients' first response was that they would want all of their days to be SFD days.
- However, this desire for a complete cure, was usually followed up by an amount of additional SFD that patient's would consider desirable or an important improvement in their asthma.
- These additional SFD days ranged from 2 to 15 days, with the average being an increase of 25% SFD days in 1 month.

Patient Estimates of SFD MID

- Additional SFD that would still be considered an improvement to patients ranged from 1 to 6 days, with the average being an increase of 11% in SFD days in 1 month.

Qualitative Findings

- It is possible to elicit patient perceptions of MID and responder values to help with the interpretation of PRO results from clinical trials.
- Recruiting patients that have recently started a new medication may improve the ability of the patients to quantify a desired response.
- These approaches should be considered complementary to quantitative methods, which together can provide an accumulation of evidence for meaningful changes in PRO measures.

Example: Acne-QoL

- Acne-QoL contains 19 items designed to measure the impact of facial acne across four dimensions of patient quality of life
- Primary Method: Patient-based judgment
  - Global patient rating of change in severity at study end
- Secondary Methods:
  - Physician-based judgment
    - Change in a categorical physician rating of acne severity between mid-study and end-of-study (2 categories of change required for responder)
  - Distribution-based
    - 0.5 SD at baseline
  - Reliable Change Index
- Alternative Approach:
    - CDF

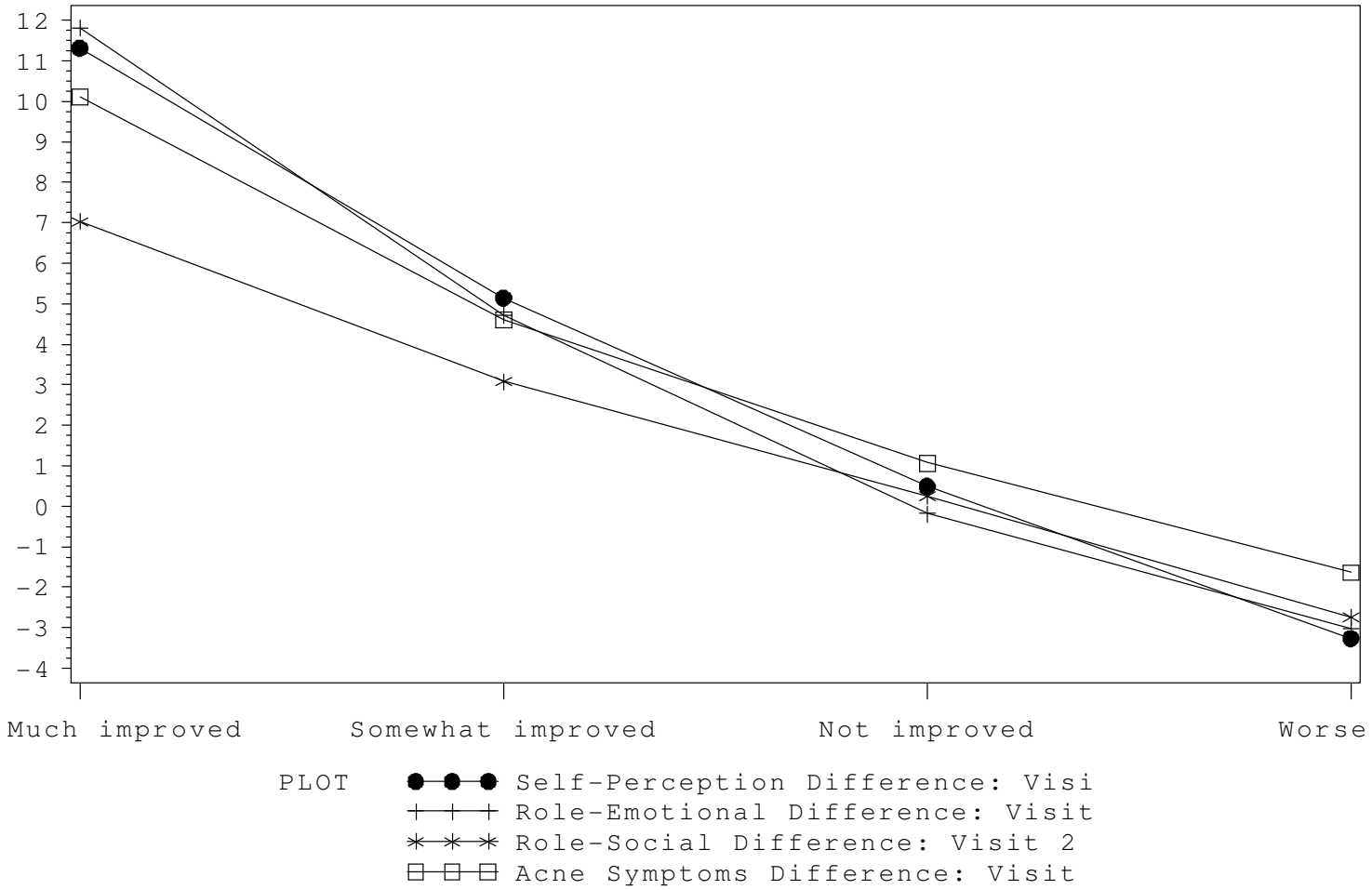
McLeod et al., 2003

Example: Acne-QoL

- Patient-based anchor: “How would you rate your acne now compared to how it was before you started the study medication?”
  - Much improved
  - Somewhat improved
  - Not improved
  - Worse
  - Much worse

- Patients responding “somewhat improved” were defined as those who had experienced a response in acne appearance

Example: Acne-QoL Score Change by Patient Global



Example: Acne-QoL

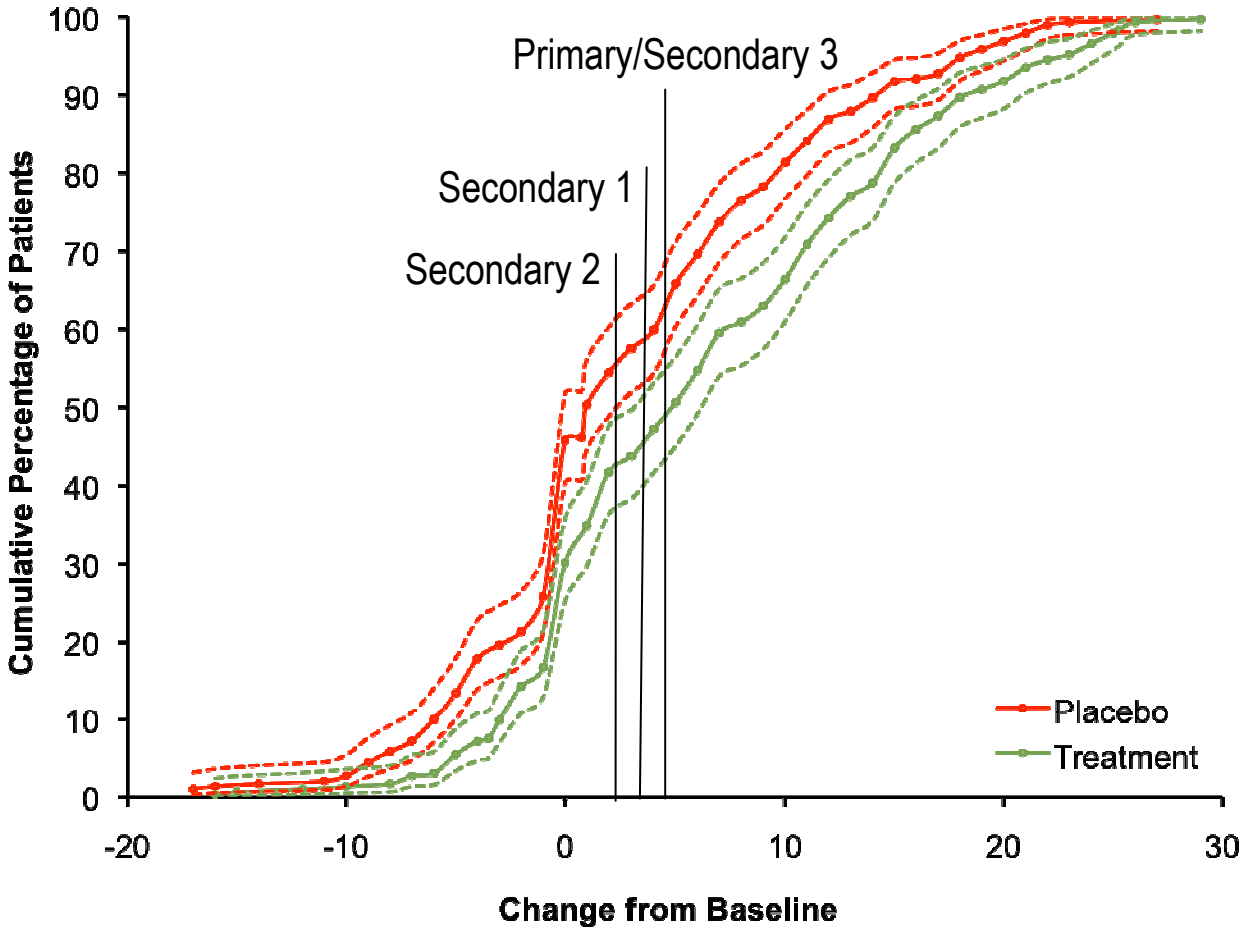
- Secondary methods:
  - Physician-based anchor. The Facial Acne Global Assessment (FAGA) ratings categorized the patient's acne as one of the following:
    - Absent
    - Minimal
    - Mild
    - Mild to moderate
    - Moderate
    - Marked
    - Severe

- Physician FAGA ratings at mid-study were compared to their responses at end-of-study
- Responder defined as the average subscale value for the patients that moved up two classifications

Example: Acne-QoL Responder Cutpoints

	Self-perception	Role-emotional	Role-social	Acne Symptoms
Primary Patient global	5.2	4.7	3.1	4.6
Secondary/Supportive 1 Physician global	4.5	4.5	2.7	3.6
Secondary/Supportive 2 0.5 SD at Baseline	4.1	4.2	3.3	2.9
Secondary/Supportive 3 RCI	5.2	7.7	5.3	7.0

Example: Acne-QoL - CDF for Self-perception



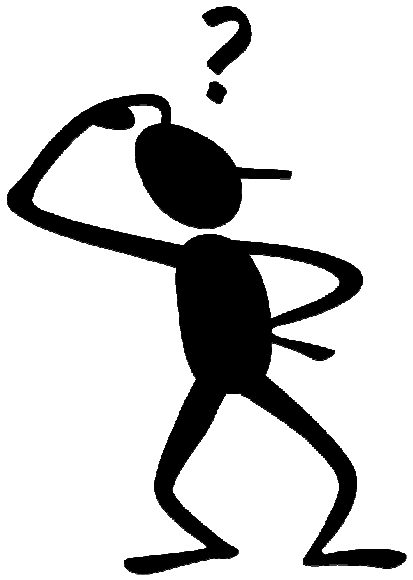
Summary

- Primary method: Based on a relevant anchor
  - How to select an appropriate anchor?
    - "The anchors chosen should be easier to interpret than the PRO measure itself" (Final FDA PRO Guidance, 2009)
    - "The external anchor chosen must itself be a valid measure of clinical change" (Eurich et al., 2006)

Summary

- If there is no appropriate anchor?
  - Include one next time!
  - “Distribution-based methods for determining clinical significance of particular score changes should be considered as supportive and are not appropriate as the sole basis for determining a responder definition” (Final FDA PRO Guidance, 2009)
  - “In actuality, only anchor-based methods estimate whether group change is big enough to be regarded as minimally or clinically important. The so-called distribution-based indices are simply a way of expressing the observed change in a standardized metric” (Hays et al., 2005)
  - CDF does not require a specified anchor and can be used to assess differences in treatment groups across a relevant range of change scores

Questions?



References

- Diaz-Ramos S, Stevens, Jr. DL, Olsen AR. EMAP Statistical Methods Manual. EPA/620/R-96/. Corvallis, OR: U.S. Environmental Protection Agency, Office of Research and Development, National Health Effects and Environmental Research Laboratory, Western Ecology Division, 1996.
- Eurich DT, Johnson JA, Reid KJ, Spertus JA. Assessing responsiveness of generic and specific health related quality of life measures in heart failure. Health and Quality of Life Outcomes 2006;4:89-102.
- Juniper ER, Guyatt GH, Willan A, Griffith LE. Determining a minimal important change in a disease-specific quality of life questionnaire. J Clin Epidemiol 1994 Jan;47(1):81-7.
- Hays RD, Brodsky M, Johnston MF, Spritzer KL, Hui K-K. Evaluating the statistical significance of health-related quality-of-life change in individual patients. Evaluation & the Health Professions 2005;28(2):160-171.
- Hays RD, Cunningham WE, Sherbourne CD, Wilson IB, Wu AW, Cleary PD, et al. Health-related quality of life in patients with human immunodeficiency virus infection in the United States: results from the HIV Cost and Services Utilization Study. AM J Med 2000;108:714-722.
- Farrar JT, Dworkin RH, Max MB. Use of the cumulative proportion of responders analysis graph to present pain data over a range of cut-off points: making clinical trial data more understandable. Journal of Pain and Symptom Management 2006;31(4):369-377.
- McLeod LD, Fehnel SE, Brandman J, Symonds T. Evaluating minimal clinically important differences (MCID) for the acne-specific quality of life questionnaire (Acne-QoL). Pharmacoeconomics 2003;21(15):1069-79.
- Martin S, Stanford R, Dale P and Fehnel S. What represents a meaningful improvement in SFD and RFD? The patient's perspective - Submitted to the 2010 European Respiratory Society Meeting
- Riffenburgh, R.H.. Statistics in medicine. NY: Academic Press, 1999
- US Department of Health and Human Services. Food and Drug Administration. CDER2002193, docket no. 2006D-0044, Guidance for industry patient-reported outcome measures: use in medical product development to support labeling claims draft guidance. February 2006.
- US Department of Health and Human Services. Food and Drug Administration. Guidance for industry patient-reported outcome measures: use in medical product development to support labeling claims. December 2009.

Aricept Label

- T-test of mean change at week 24

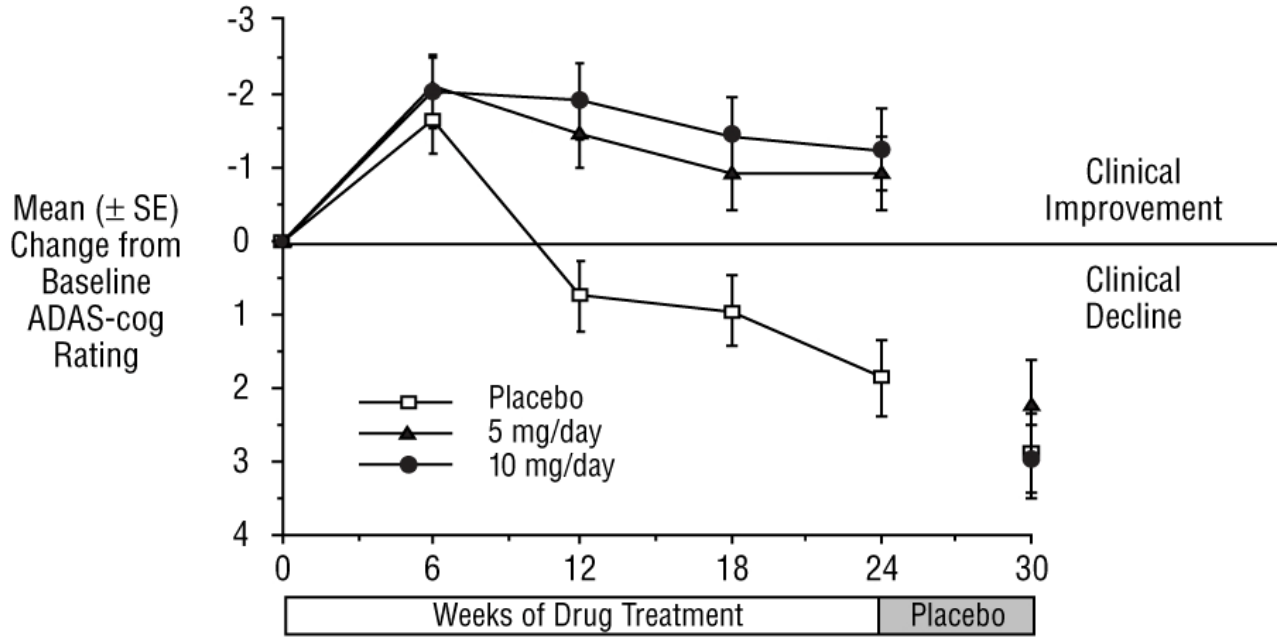


Figure 1. Time-course of the Change from Baseline in ADAS-cog Score for Patients Completing 24 Weeks of Treatment.

Aricept, 2006

Aricept Label

- CDF showing separation between treatment and placebo for multiple responder cutpoints

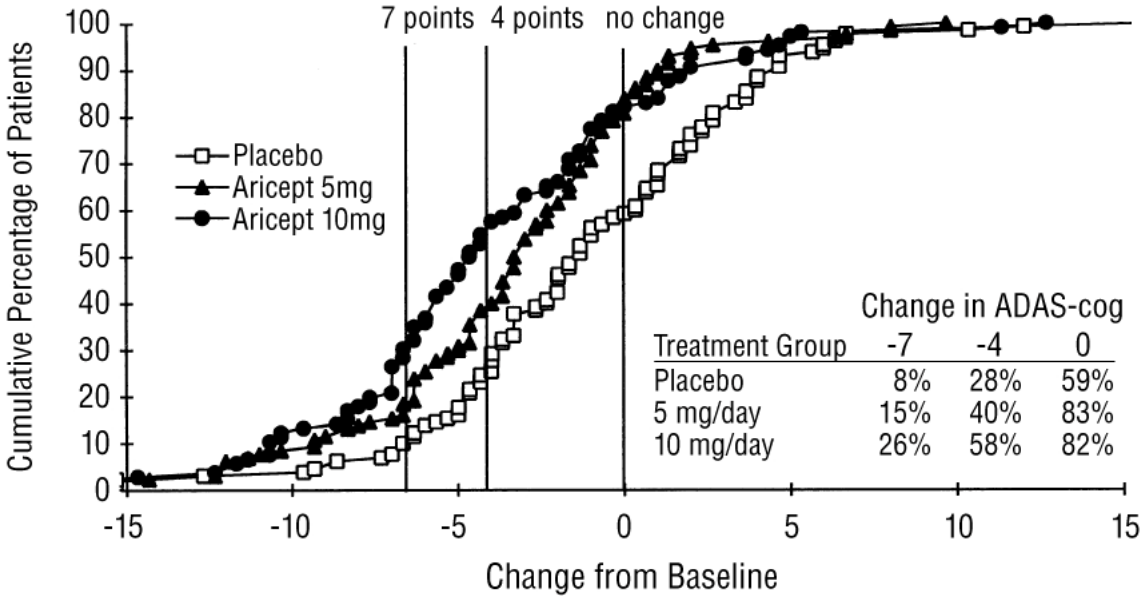
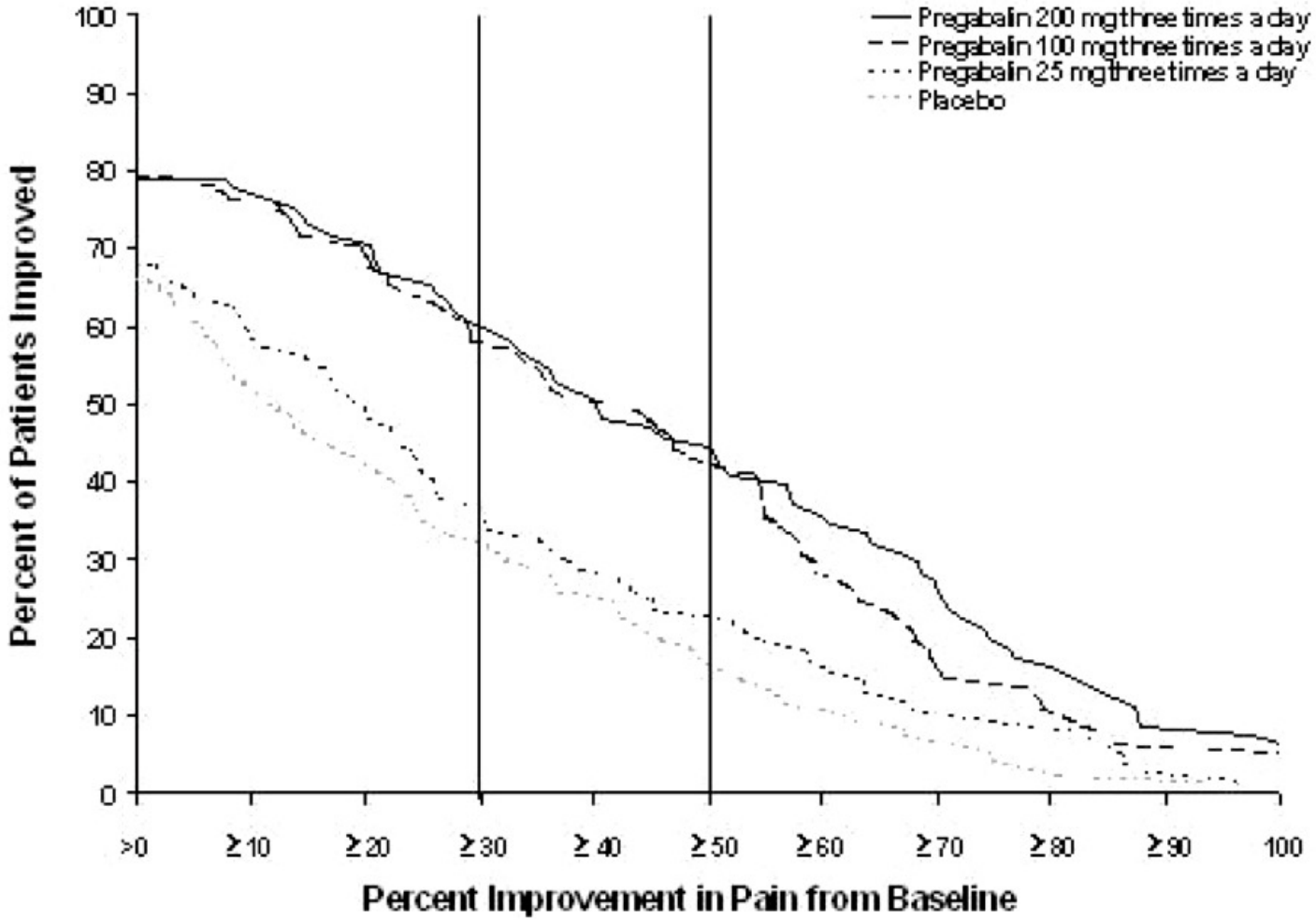


Figure 2. Cumulative Percentage of Patients Completing 24 Weeks of Double-blind Treatment with Specified Changes from Baseline ADAS-cog Scores. The Percentages of Randomized Patients who Completed the Study were: Placebo 80%, 5 mg/day 85% and 10 mg/day 68%.

Aricept, 2006

Clinical Improvement: reduction in score/negative change

Pregabalin Label



Contact Information

Lori McLeod, PhD
(919) 541-6741
lmcleod@rti.org

Ron D. Hays, PhD
(310) 794-2294
drhays@ucla.edu

Susan Martin, MSPH
(734) 213-5469
smartin@rti.org

Sheri Fehnel, PhD
(919) 541-7454
sfehnel@rti.org