

Session 6

Minimally Important Differences

Contributors

- Dennis Revicki
- Ron Hays
- Jeff Sloan
- David Feeny
- David Cella

FDA Advice Mirrors Advice from Clinical Colleagues

- Body temperature
- Blood pressure
- Tumor size, etc.....
 - ... all understood in context of starting point and meaning of change, group and individual
 - ... This is how they become accepted in everyday use (and respected as interpretable outcomes)
- Pain, Fatigue, etc
- Physical / Social Functioning
- HRQL
 - ... Are they ready for the same?
 - Yes, if the standard is reliability or validity
 - No, if the standard is the meaningfulness of scores or changes in scores

From then to now

Minimally Important Difference (MID)

- “the smallest difference in score which patients perceive as beneficial and which would mandate, in the absence of troublesome side effects and excessive cost, a change in the patient’s management” (Jaeschke et al, 1989)
- amount of difference or change observed in a PRO measure between treatment groups in a clinical trial that will be interpreted as a treatment benefit (FDA, 2006)

In all views expressed along the way,
clinical significance ? statistical significance

Clinical Significance

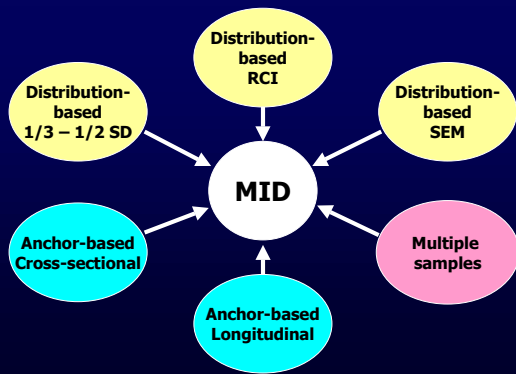
- critical for:
 - determining appropriate sample sizes
 - interpreting treatment group results
 - understanding change over time
- challenging because PRO measures lack a “gold standard” against which to quantify meaningful differences

Some Myths About MID

- MID = 0.5 Likert scale point / item
- MID = 0.5 SD
- MIDs are stable properties of instruments
- MIDs are stable across the range of a scale
- MIDs are symmetrical
- All anchors are equal

Key strategies when developing MIDs are
triangulation, range estimation and
accumulation of knowledge

A Comprehensive MID Approach



Distribution-based Approaches

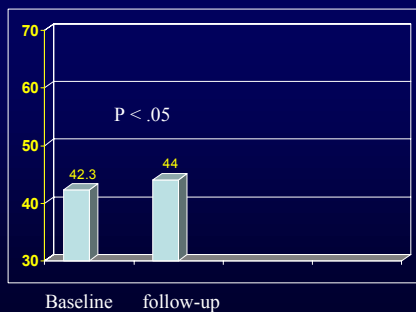
- 0.33 – 0.5 SD (effect size)
- 1 Standard Error of Measurement (SEM)

$$SEM = SD\sqrt{1-rel}$$

- Reliable Change Index

$$RCI = x_2 - x_1 / \sqrt{2}SEM$$

Group Change in SF-36 PF-10: Clinically meaningful?



Hays et al, 2005

Individual Classification using two MID Choices

PF-10	# Better	# Worse
SEM (5.9)	10	5
RCI (8.4)	7	1

Hays et al, 2005

Lung Cancer Symptoms: Distribution-based Estimates

FACT-L	1/3 SD	1/2 SD	SEM
Lung Cancer Subscale (LCS)			
Baseline	1.7	2.5	2.9
12 week	1.5	2.3	2.7
Baseline to 12-week change	1.7	2.6	NA

NA, not available

Cella et al, JCE, 2002

Anchor-based Approaches Cross-sectional

- Anchors are used to categorize patients into clinically distinct groups
- PRO score differences between adjacent categories are estimates of the MID
- Effect sizes for MID estimates:

$$(X_1 - X_2) / SD$$

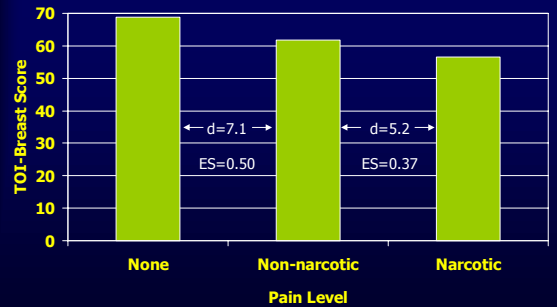
Lung Cancer Symptoms: Cross-sectional clinical anchors

Clinical Indicator		LCS	
Prior 6 months weight loss	<5%	19.4 (4.9)	0.5 SD and 1 SEM range = 2.5-3.0
	>5%	17.0 (5.0)	
	Difference	2.4	
Effect size	0.48		
ECOG PS	0	20.5 (4.7)	
	1	17.9 (5.0)	
	Difference	2.6	
Effect size	0.52		
Primary Dx symptoms	≤1	20.8 (4.5)	
	>1	17.2 (4.8)	
	Difference	3.6	
Effect size	0.72		

Values are mean (SD)

Cella et al, *JCE*, 2002

Example: Anchor-based Approaches, Cross-sectional Anchor



Eton, Cella, Yost et al.; 2004

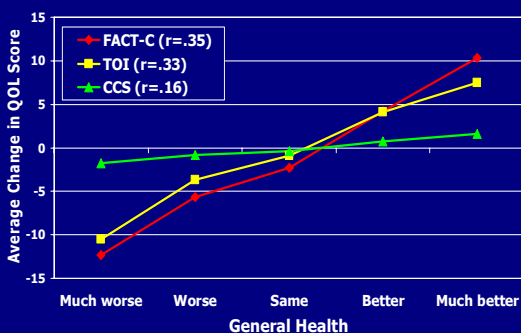
Not all Anchors are Equal

- Anchors are used to categorize patients in to clinically distinct groups
- PRO score differences between adjacent categories are estimates of the MID
- Effect sizes for MID estimates can help reconcile differences in MID estimates
 $-(X_1 - X_2) / SD$

Assessing the Usefulness of Anchors

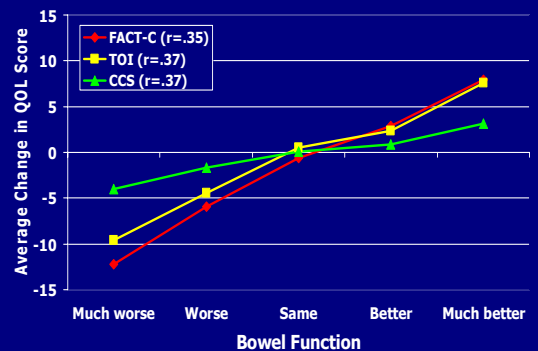
- Cross-sectional: Anchor and PRO should be linearly related and have at least a moderate correlation
- Longitudinal: Anchor change score and PRO change score should be linearly related and have at least a moderate correlation
- Small ($r=.1$), moderate ($r=.3$), large ($r=.5$)
- Weak correlations can yield MID estimates that are too small

Assessing the Usefulness of Anchors

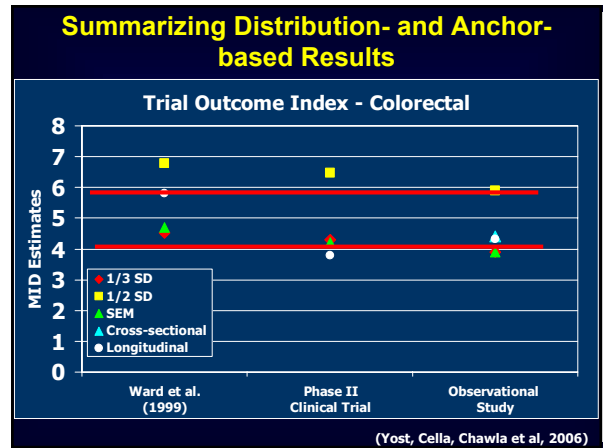
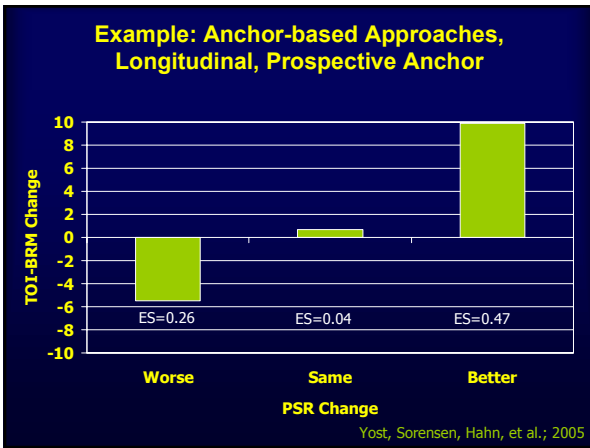
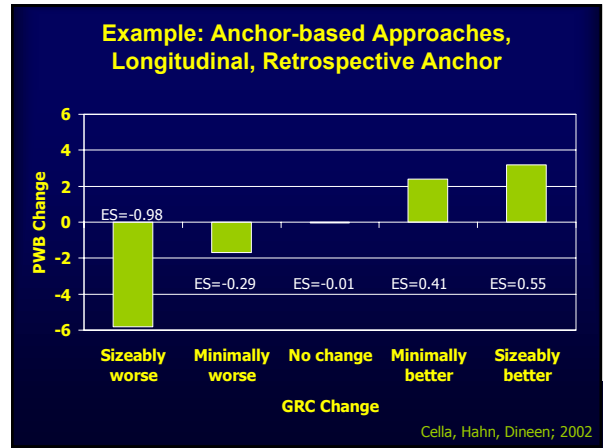
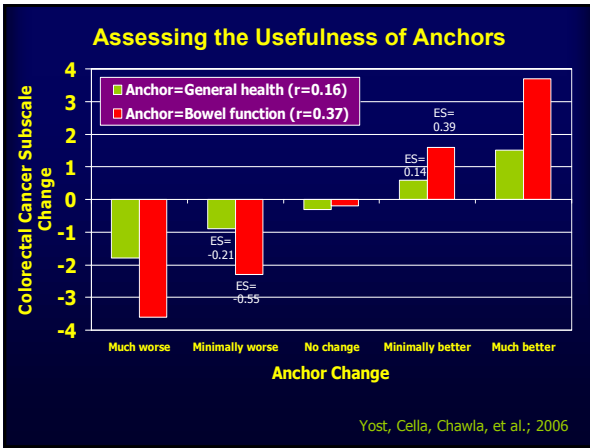


Yost and Eton, 2005

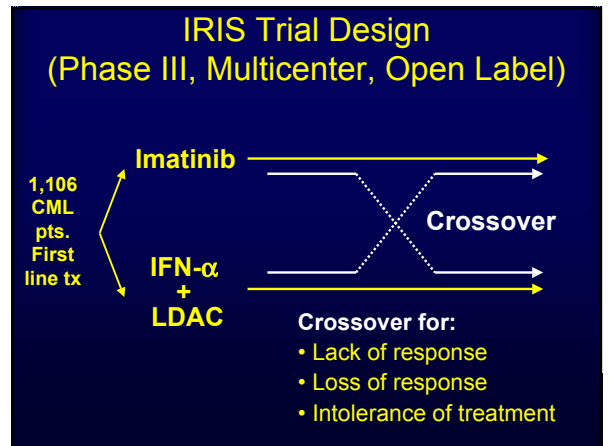
Assessing the Usefulness of Anchors



Yost and Eton, 2005



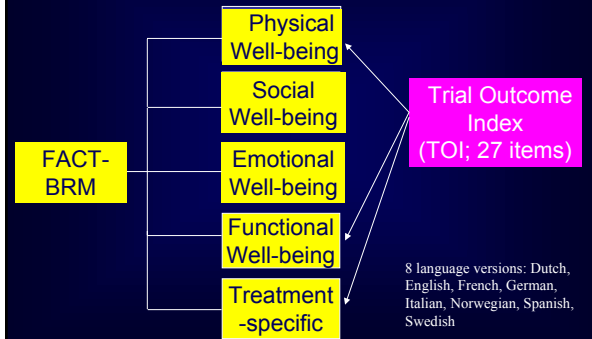
Clinical Trial Example



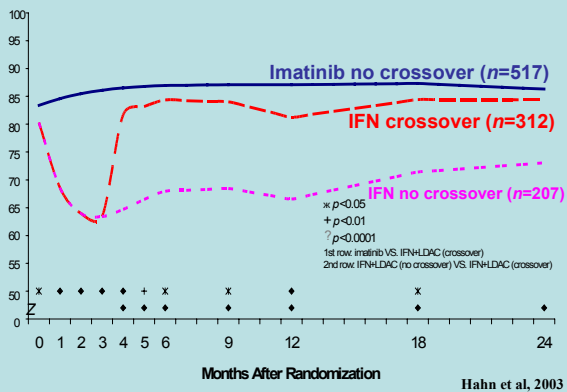
Study Endpoints

- **Primary Endpoint:** Time to Progression
- **Secondary Endpoints:**
 - Rate/duration of complete hematologic response & major cytogenetic response, safety, tolerability, molecular remission, pharmacogenomics, pharmacokinetics
 - **PRO** (measured at Baseline and Months 1-6, 9, 12, 18, 24)

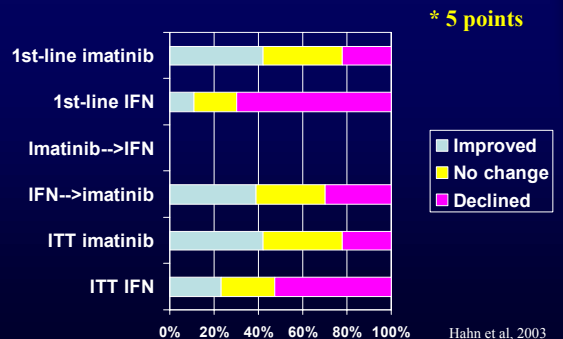
Functional Assessment of Cancer Therapy- Biologic Response Modifiers (FACT-BRM)



Estimated Mean TOI (Adjusted for crossover)



Meaningful TOI Change* from Baseline to Month 12 (n=736)



Not at all	A little bit	Some -what	Quite a bit	□ imatinib	□ IFN+Ara-C	□ =both
		○	■			} Better daily functioning & well-being on imatinib
		○	■			
		○	■			} Less fatigue on imatinib
		○	■			
		○	■			} Milder emotional/cognitive complaints on imatinib
		○	■			
		○	■			} Fewer side effects on imatinib
		○	■			
■	○					
■	○					
		□				

Group Difference versus Individual Change

- Between-group MID is likely smaller than within-person MID
- Most MID are derived using group data
 - Cross sectional anchors
 - Longitudinal anchors
- Anchors rarely have clearly-known “minimum thresholds” ...this can overestimate group MID

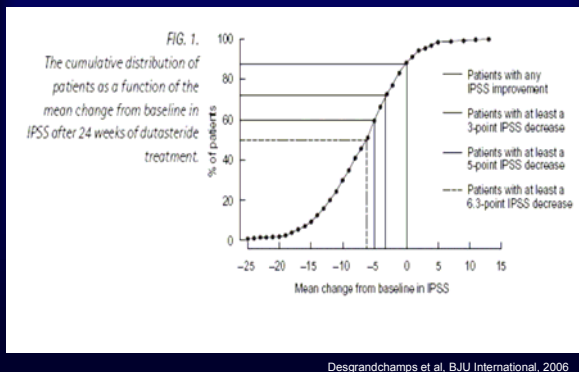
Applying Group-derived MID to Individual Classification

- Use high end of MID estimate
- Require confirmation of improvement
 - A challenge when tracking time to worsening
- Increase MID estimate for individual change to require a “substantial” change score.
- Keep MID estimate as is based on probability of correct classification and then ask if % difference between groups is “substantial”

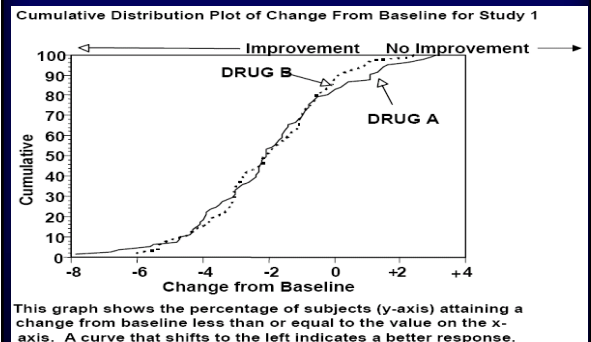
FDA 6-19-06: “Is there a better idea or more a more intuitive idea?”

- A comparison of the distribution of individual responses between treatment groups may be more informative
- But, deciding upon the cutoff for a response or a responder is not simple - may need evaluation of minimum change an individual can perceive as beneficial

Example of Cumulative Distribution Plot



FDA example: 6-19-06



General Recommendations

- Combined Approaches and Data Triangulation are recommended to generate likely MID range
- Use care in selection of anchors
- Use higher end of range for individual classification; lower end for group comparisons
- State MID and rationale up front
- Interpretation of benefit should rarely be based on PRO alone; use context of all data (e.g., fatigue benefit in context of Hb change)
- MID choice has implications for % cases improved or worsened; consider sensitivity analysis and even multiple cumulative distribution plots

Other Considerations

- MIDs for Preference-based measures have been estimated at 3-6% of total range.
- Item response theory (IRT) scores will require reconsideration of MID value across the range of the concept being measured

Future Research Needs

- New approaches to determining an optimal or “likely” MID within a given range for a specific application
 - Bayesian
 - Testing different MID choices against trusted anchors/outcomes
 - Meta-analysis of clinical trial data
- Effect of IRT score transformation upon MID estimates across the range
- Getting better anchors (qualitative research)