

Estimates of Minimally Important Differences Should Reflect Uncertainty

Ron D. Hays
(hays@rand.org)

November 15, 2003, 4:15-5:45 pm

ISOQOL 2004, Prague

Important Terminology

Minimally Important Difference (MID) or Minimal difference (MD)

Some differences between groups or over time may be so small in magnitude that they are not important. Smallest difference in score that is worth caring about.

-> Minimally Detectable Difference (MDD)

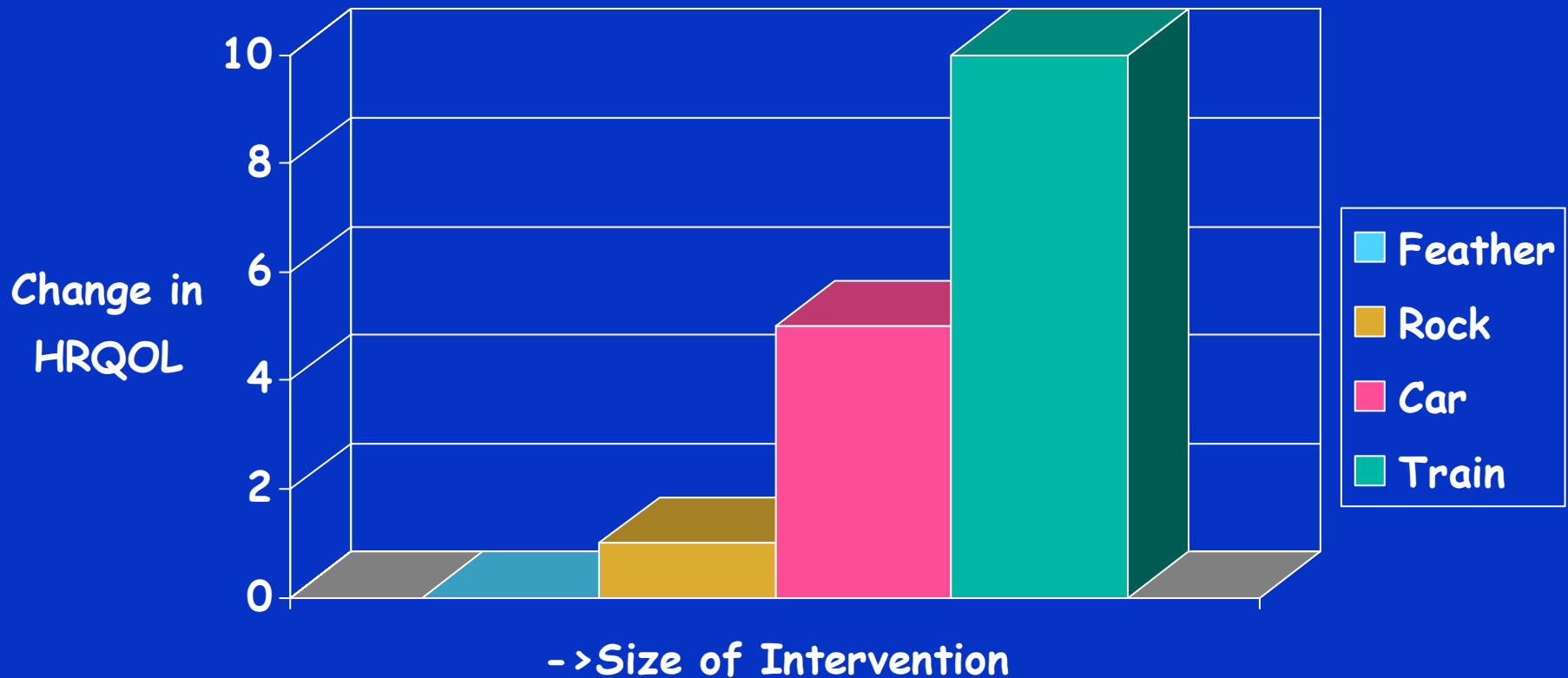
-> Clinically Important Difference (CID)

Obviously Important Difference (OID)

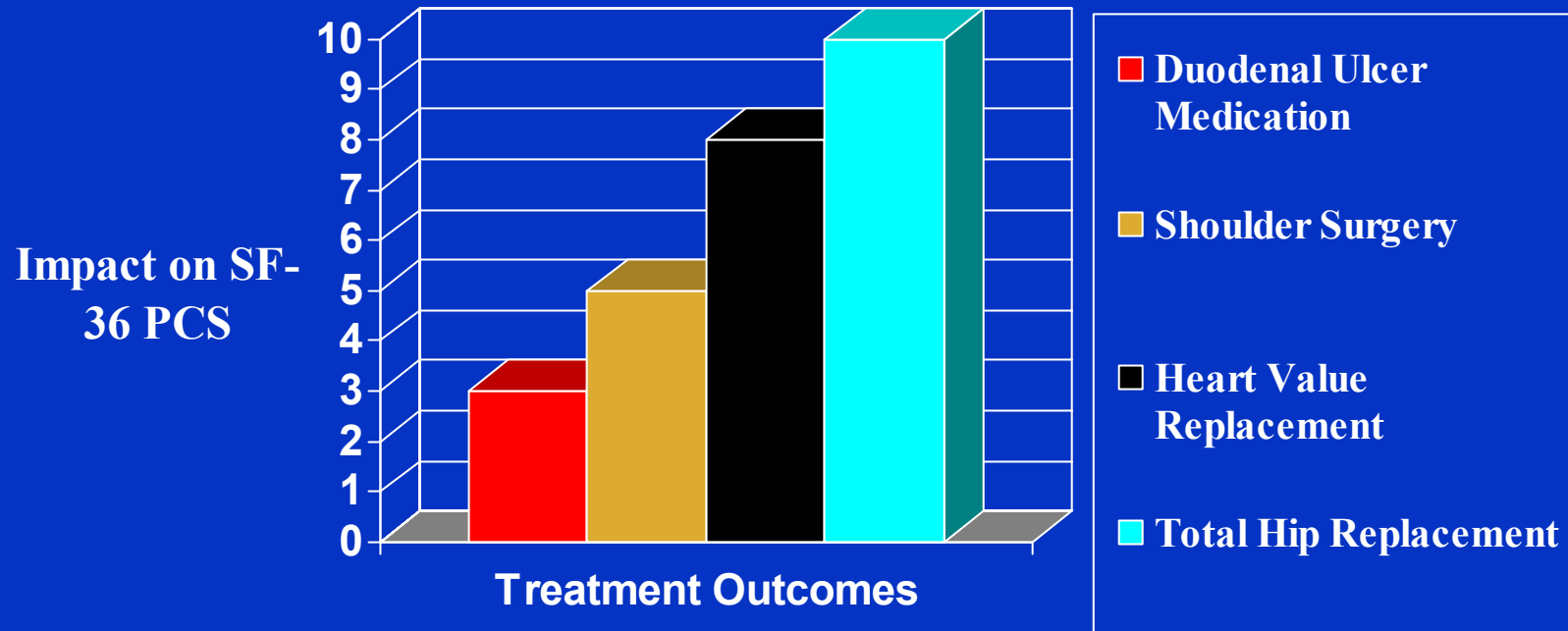
MID is Subset of Responsiveness to Change

- Evaluated by comparing HRQOL change to external indicators of change (anchors)
 - > Self-report
 - > Clinician or other report
 - > Clinical parameter
 - > Clinical intervention
- MID estimation requires an "anchor"

Magnitude of HRQOL Change Should Parallel Underlying Change

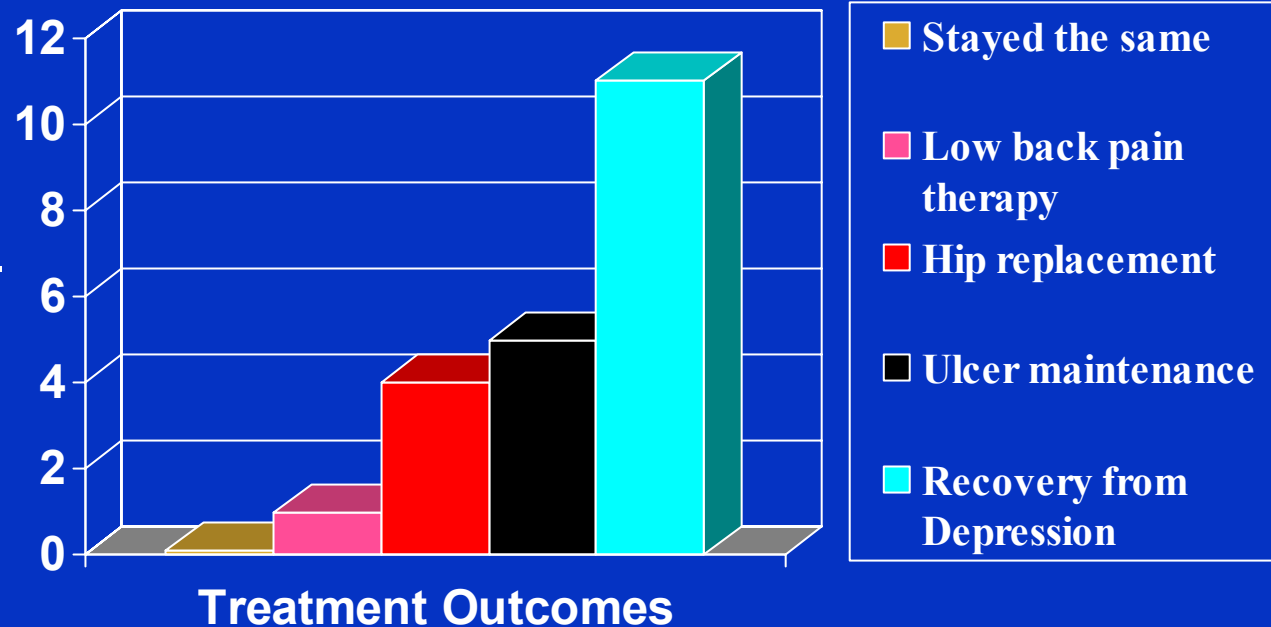


Change and Responsiveness in PCS Depends on Treatment

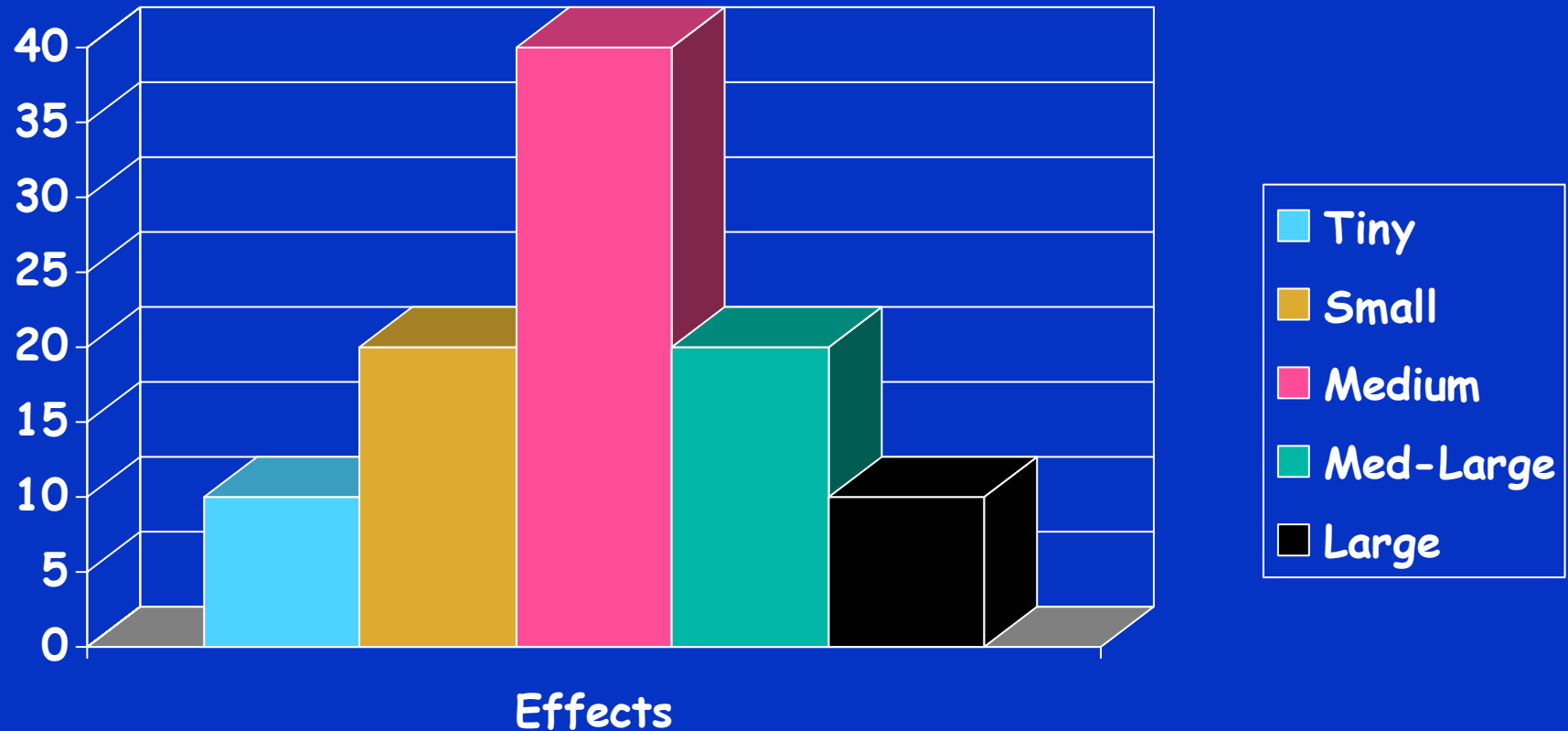


Change and Responsiveness in MCS Depends on Treatment

Impact on SF-36 MCS



Average of Observed Important Changes Should be a Medium Effect



Factors associated with variation in MID estimates

Size of change anchor represents

Type of anchor (self-report, clinically, etc.)

Measure or scale (in part due to fit of anchor to it)

Starting Point (low or high)

Cumulative Effects Over Time

Norman, Sloan, Wyrwich (2003)

Interpretation of Changes in Health-related Quality of Life: The remarkable universality of half a standard deviation.

Table 1 reports estimates of MIDs for 33 published articles.

“For all but 6 studies, the MID estimates were close to one half a SD (mean = 0.495, SD = 0.155)” (p. 582).

I have reasonable doubts about this conclusion.

MID Varies by Anchor

693 RA clinical trial participants evaluated at baseline and 6-weeks post-treatment.

Five anchors: 1) patient global self-report; 2) physician global report; 3) pain self-report; 4) joint swelling; 5) joint tenderness

Kosinski, M. et al. (2000). Determining minimally important changes in generic and disease-specific health-related quality of life questionnaires in clinical trials of rheumatoid arthritis. Arthritis and Rheumatism, 43, 1478-1487.

Effect Sizes (mean = 0.34) for SF-36 Changes Linked to Minimal Change in Anchors

Scale	Self-R	Clin.-R	Pain	Swell	Tender	Mean
PF	<u>.35</u>	.33	.34	<u>.26</u>	<u>.32</u>	.32
Role-P	<u>.56</u>	.52	<u>.29</u>	.35	.36	.42
Pain	<u>.83</u>	.70	.47	.69	<u>.42</u>	.62
GH	<u>.20</u>	.12	.09	.12	<u>.04</u>	.12
EWB	<u>.39</u>	.26	.25	.18	<u>.05</u>	.23
Role-E	<u>.41</u>	.28	<u>.18</u>	.38	.26	.30
SF	<u>.43</u>	.34	<u>.28</u>	.29	.38	.34
EF	<u>.50</u>	.47	<u>.22</u>	.22	.35	.35
PCS	<u>.49</u> < -	.48	<u>.34</u>	.29	.36	.39
MCS	<u>.42</u> < -	.27	<u>.19</u>	.27	.20	.27 ^A

Changes in SF-36 Scores Associated with Minimal Change in Anchors

Scale	Mean (ES)	Range	SD	Range/SD
PF	08 (.32)	2 (6 - 8)	20	.10
Role-P	16 (.42)	10 (11-21)	40	.25
Pain	11 (.62)	8 (7-15)	20	.40
GH	02 (.12)	3 (1- 4)	20	.15
EWB	04 (.23)	6 (1- 7)	20	.30
Role-E	13 (.30)	10 (8-18)	40	.25
SF	09 (.34)	4 (8-12)	20	.20
EF	08 (.35)	6 (5-11)	20	.30
PCS	03 (.39)	1 (3- 4)	10	.10
MCS	03 (.27)	3 (2- 5)	10	.30

Norman et al.

Redelmeier et al. (1997) studied 112 people with stable COPD. Norman et al. reported an ES of 0.43 (0.31-0.54) for 6-minute walk test based on a comparison of “about the same” versus “a little better” and “about the same” with “a little bit worse.”

The 6-minute walk test is not a measure of HRQOL.

Norman et al. continued

Badia et al. (2000) looked at change on MOS-HIV and MQOL-HIV in 296 persons with AIDS who improved (better and much better)--not an estimate of MID.

Bagenstowe and Bernstein estimated change in RQLQ (ES = 0.48) for 19 new rhinitis patients before and after seeing an allergist who prescribed a new medication regimen. Without knowing the magnitude of the intervention, it is unclear whether it serves as a good anchor for estimating the MID. In addition, SF-36 data not included. Average ES for SF-36 was 0.29.

Goldstein et al. (1994) examined change in CRQ among 89 patients in respiratory rehabilitation (medically supervised multidisciplinary). Without knowing the magnitude of the intervention, it is unclear whether it serves as a good anchor for estimating the MID.

Norman et al. continued

Jones and Bosh (1997) compared changes between baseline and 16 weeks for COPD patients treated with salmeterol. Effect sizes reported for SGRQ for 50 ug group but not 100 ug group. SF-36 not reported but ES was 0.11 (0.02-0.25) for 50 ug group.

Jowett et al. (2001) examined 123 ulcerative colitis patients with cross-sectional comparisons between mild vs. moderate and mild versus remission subgroups according to SEO index of disease activity. Are these cross-sectional differences indicative of minimally important differences?
Norman et al. reported SIBDQ ES of 1.06 (1.03-1.08).

Norman et al. continued

Miller et al. (2000) looked at differences between adjacent EDSS categories among 300 MS patients. Are these minimally important differences?

SF-36 ES for PCS and MCS = 0.49 (0.09-0.89) vs. 0.42 (0.09-0.89) and MCS goes in opposite direction of physical health. SIP weighted total score reported (ES = 0.89), but not SIP Physical, Psychosocial, FIS total or FIS Physical scores.

Norman et al. continued

Osman et al. (1997) studied 266 COPD patients. Compared a) no further COPD admission and survived vs. further COPD admission or died (ES = 0.32); b) no nebulizer vs. nebulizer (ES = 0.64); and c) no domiciliary O₂ versus O₂ (ES = 0.09). Do these things tell us about MID?

SGRQ ES = 0.49 (0.07-1.5) versus 0.33 (0.24-0.37)

Norman et al. continued

Bessette et al. (1998) pre-post comparison of 40 carpal tunnel patients who were "somewhat satisfied" with surgery.

SF-36 ES 0.52 (0.24-0.62) vs. 0.43 (0.35-0.49)

Global QOL ES = 0.34 vs. 0.23

Norman et al. continued

Cella et al. (1995) study of 135 lung cancer patients and used change on Eastern Cooperative Group Performance Status Self-Report (0 = ambulatory without symptoms; 1 = ambulatory with symptoms; 2 = requires bed rest less than half of waking day; 3 = requires bed rest more than half of waking day; 4 = bedridden) Is change on this measure a "minimal" change? ES of 0.50, 0.55 and 0.43 reported for LCS, TOI, and FACT-G, respectively.

Norman et al. do not report FACT-G PWB (ES = 0.57), SWB (ES = wrong direction for improvers), EWB (ES = 0.24), or FWB (ES = 0.37)

Norman et al. continued

Cella et al. (2002) study of 573 lung cancer patients used global rating of change to categorize into: sizably worse, minimally worse, no change, minimally better, and sizably better.

FACT-G ES = 0.48 (0.05-0.51) vs. 0.35 (0.03-0.62). 0.48 is average for FACT-G total, but average of 4 subscales and total score is 0.35.

Norman et al. continued (final)

Ringash et al. (2000) study of 98 patients with laryngeal cancer is cited in Norman et al. reference list but Table 1 refers to 2002 rather than 2000. The 2000 article does not appear to report effect sizes.

Norman et al. continued

Wyrwich et al. (1999) studied 605 CAD/CHF patients and Wyrwich et al. (1999) evaluated 417 COPD patients. Norman et al. cited ES of 0.36 and 0.35 for the CHQ and CRQ, respectively.

No anchors were used in these studies. The study referred to ES that would be needed to demonstrate equivalency to MCID standards reported previously.

Ceiling and Floor Effects Influence Amount of Observed Change

Same retrospective report of change associated with bigger prospective change for those with more room to change

- Among those who said their physical health was *somewhat worse*, change was -26 points for those with high (81-100) compared to +3 points for those with low (0-20) baseline physical health (Baker et al., 1997, Medical Care).

Cumulative Change Over time

Baseline	42
Year 4	36

Note: 4-year decline in PCS among US seniors, 1990-94.

-> 1.5 points per year (0.15 SD)



Summary

MID may vary by type of anchor used to estimate it, starting position, and other variables.

Bounded estimates (range or CIs) are better than a single point estimate given inherent uncertainty.

OID is 0.50; MID is likely to be considerably smaller.



Coming June 24-25, 2004!!



Advances in Health Outcomes Measurement:

Exploring the Current State and the Future Applications of Item Response Theory, Item Banks and Computer-Adaptive Testing

Conference co-sponsored by the

National Cancer Institute

Drug Information Association



Resource Centers for
Minority Aging Research



This presentation was supported in part by the UCLA/DREW Project EXPORT, National Institutes of Health, National Center on Minority Health & Health Disparities, (P20-MD00148-01) and the UCLA Center for Health Improvement in Minority Elders / Resource Centers for Minority Aging Research, National Institutes of Health, National Institute of Aging, (AG-02-004).

Appendix (Back-Pocket) Slides Follow

Changes in SF-36 Scores Associated with Minimal Change in Anchors

Scale	Self-R	Clin.-R	Pain	Swell	Tender	Mean
PF	8	8	8	<u>6</u>	<u>8</u>	8
Role-P	<u>21</u>	20	<u>11</u>	13	13	16
Pain	<u>15</u>	12	8	12	<u>7</u>	11
GH	<u>4</u>	2	2	3	<u>1</u>	2
EWB	<u>7</u>	5	5	3	<u>1</u>	4
Role-E	<u>18</u>	12	<u>8</u>	16	11	13*
SF	<u>12</u>	9	<u>8</u>	8	10	9
EF	<u>11</u>	10	<u>5</u>	5	8	8
PCS	<u>4</u>	4	<u>3</u>	3	3	3.5*
MCS	<u>5</u>	3	<u>2</u>	3	2	3 ²⁹

Self-Report Anchor

Overall has there been any change in your asthma since the beginning of the study?

Much improved; Moderately improved; Minimally improved

No change

Much worse; Moderately worse; Minimally worse

Self-Report Anchor

People who report a “minimal” change

How is your physical health now compared to 4 weeks ago?

Much improved; Moderately Improved;

Minimally Improved;

No Change;

Minimally Worse;

Moderately Worse; Much Worse

Examples of Other Anchors

Clinician report

- How is Jan's physical health now compared to 4 weeks ago?

Clinical parameter

- Change from CDC Stage A to B
- Became seizure free

Clinical intervention

- Before and after Prozac

Group Average is Different from Individual Change

Average change collapses across individual responses.

Is inference about minimum amount of change that is important for individuals based on a group average reasonable?

What if scale score improved by 4 points for half the people and 0 points for the other half? Is the MID = 2 or 4?