#### Estimating Minimally Important Differences (MIDs) of Health-Related Quality of Life Measures? WORK IN PROGRESS

#### Ron D. Hays

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

http://twitter.com/rondhays



#### UCLA GIM/HSR Seminar Series, Los Angeles, CA

October 15, 2010

											11
1	2	3	4	5	6	7	8	9	10	11	

### **Take Away Points**

- External indicators of change ("anchors") are needed to estimate minimally important differences (MIDs) in health-related quality of life measures
  - MID estimate is useful only if anchor is positively correlated with the measure being evaluated
  - Multiple anchors are desirable
  - -What the MID is may be unclear
- MID is a group-level average change that is not applicable to individual change ("responder")

### **Minimally Important Difference (MID) Rationale**

- Differences in group means can be statistically significant but small (with large sample size)
- MID is the smallest difference to "care about"
  Clinically important or implies non-trivial treatment benefit

### "Anchor-based" Estimates of MID

- Anchor used to classify respondents in terms of change
  - Self-report, provider report, clinical measure
- Estimate change ("delta") on measure for subgroup that changed by a minimally important amount (via anchor)

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



### Hypothetical Change in Physical Function (T-score units) by magnitude of intervention

 $\mathcal{T}\mathcal{T}$ 



	Change #1 MID = ?	Change #2 MID = ?	Change #3 <b>MID = 4</b>
Minimal Change on Anchor	0	+ 2	+4
No Change on Anchor	Doesn' t matter	+ 2	0, +1, or +2

# Despite "The truly remarkable universality of half a standard deviation"

 Norman, Sloan, & Wyrwich, 2004, "The truly remarkable universality of half a standard deviation" <u>Expert Rev</u> <u>Pharmacoecon Outcome Res</u>

### Effect Sizes (mean = 0.34) for SF-36 Changes Linked to Minimal Change in Anchors (Kosinski et al., <u>Arth Rheu</u>, 2000)

Scale	Self-R	ClinR	Pain	Swell	Tender	Mean
PF	<u>.35</u>	.33	.34	<u>.26</u>	.32	.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	.42	.27	<u>.19</u>	.27	.20	.27

### Health Assessment Questionnaire (HAQ)

- 20 physical functioning questions
- Are you able to dress yourself, including tying shoelaces and doing buttons?
- 0 = Without any difficulty; 1 = With some difficulty; 2 = With much difficulty; 3 = Unable to do
- *Mean* = 1.06 *and* SD = 0.753 *in* 1,079 *arthritis patients*
- MID estimates:

0.22 (ES = 0.29) in clinical trials

0.10-0.15 (ES = 0.13-0.20) in observational studies (e.g., Kwok & Pope, <u>J Rheumatology</u>, 2010; 0.10 in this study)

# PROMIS Physical Function Scale in Rheumatoid Arthritis" (Fries et al.)

- 124-item physical functioning bank/20-item short form
- T-score mean of 50 and SD of 10 in general U.S. pop.
  - Liu et al., J Clin Epidemiology, 2010.
  - -www.nihpromis.org
- Three waves of data
  - -Baseline (n = 521)
  - -6 months post-baseline (n = 483)
  - One year post-baseline (n = 472)

### **Retrospective Ratings of Change in RA Study**

- Change in activity (CHG\_ACTIVITY)
- Change in fatigue (CHG\_FATIGUE)
- Change in pain (CHG\_PAIN)
  - Got a lot better
  - Got a little better <----
  - Stayed the same
  - ---- Got a little worse ----
  - Got a lot worse

### **SF-36 Retrospective Change Item**

- Compared to one year ago, how would you rate your health in general now? (HT)
  - Much better now than one year ago
  - Somewhat better now than one year ago <---</p>
  - About the same as one year ago
  - Somewhat worse now than one year ago <---</p>
  - Much worse now than one year ago

### **Global Rating of Physical Functioning**

- To what extent are you able to carry out your everyday physical activities such as walking, climbing stairs, carrying groceries, or moving a chair? (GLOBAL06)
  - Completely
  - -Mostly
  - Moderately
  - -A little
  - Not at all

### **Minutes Spent Exercising**

- Vigorous exercise
- Aerobics
- Biking
- Racquet sports
- Swimming
- Other exercises

# Correlations of Anchor with Change in PROMIS Physical Functioning Domain (goal: r = 0.371 or higher)

PF <sub>2</sub> - PF <sub>1</sub>	$PF_3 - PF_2$	$PF_3 - PF_1$
0.26 (0.29)	0.34 (0.33)	
0.20 (0.21)	0.23 (0.24)	
0.27 (0.28)	0.30 (0.30)	
		0.21(0.23)
0.17 (0.19)	0.20 (0.20)	
0.10 (0.07)	0.08 (0.06)	
	PF <sub>2</sub> - PF <sub>1</sub> 0.26 (0.29) 0.20 (0.21) 0.27 (0.28)  0.17 (0.19) 0.10 (0.07)	$PF_2 - PF_1$ $PF_3 - PF_2$ 0.26 (0.29)0.34 (0.33)0.20 (0.21)0.23 (0.24)0.27 (0.28)0.30 (0.30)0.17 (0.19)0.20 (0.20)0.10 (0.07)0.08 (0.06)

Note: Product moment correlations presented, followed by Spearman rank-order correlations (in parentheses)

### Change in PROMIS Physical Functioning (W2 - W1) by Retrospective Ratings of Change (n = 463)

 $PF_1 = 40.19 (SD = 9.18); PF_2 = 39.81 (SD = 9.44)$ F (4, 458 dfs) = 9.41, p < .0001 (Activity) r = 0.26 (0.29)F (4, 457 dfs) = 4.68, p=.0010(Fatigue) r = 0.20 (0.21) F(4, 457 dfs) = 9.81, p < .0001(Pain) r = 0.27 (0.28)Activity Fatigue Pain n – Got a lot better: + 0.94<sup>a</sup> 0.94<sup>a</sup> 1.25<sup>a</sup> (19-21) 0.50<sup>a</sup> (41-61) – Got a little better: + 0.65<sup>a</sup> 0.54<sup>a</sup> 0.02<sup>a,b</sup> (224-258) - Stayed the same: - 0.04<sup>a,b</sup> - 0.16<sup>a,b</sup> – Got a little worse: - 1.31<sup>b</sup> - 1.16<sup>b</sup> (107-126) - 1.06<sup>b,c</sup> - Got a lot worse: - 3.19<sup>c</sup> - 3.12<sup>c</sup> (28- 32) - 2.06<sup>c</sup>

### Change in PROMIS Physical Functioning (W2 - W1) by Change in Global PF (n = 465)

F(4, 460 dfs) = 3.86, p = .0043 r = 0.17 (0.19)

	Giodal PF	Π
-2+ levels better:	0.53 <sup>a</sup>	22
– <u>1 level better</u> :	0.32 <sup>a</sup>	68
- Stayed the same:	- 0.31 <sup>a,b</sup>	273
- <u>1 level worse</u> :	- 1.52 <sup>b</sup>	60
-2+ levels worse:	- 1.39 <sup>b</sup>	42

### Change in PROMIS Physical Functioning (W3 - W2) by Retrospective Rating of Change in Activity (n = 443)

 $PF_2 = 39.95 (SD = 9.24); PF_3 = 40.07 (SD = 9.60)$ 

F (4, 438 dfs) = 14.98, p<.0001 (Activity) r = 0.34 (0.33)

F (4, 438 dfs) = 6.32, p<.0001 (Fatigue) r = 0.23 (0.24)

F(4, 437 dfs) = 11.34, p<.0001 (Pain) r = 0.30 (0.30)

	Activity	Fatigue	Pain n
– Got a lot better:	+ 3.26 <sup>a</sup>	2.24 <sup>a</sup>	3.37 <sup>a</sup> (16-20)
– Got a little better:	+ 1.96 <sup>a,b</sup>	1.67 <sup>a,b</sup>	1.31 <sup>b</sup> (33-55)
– Stayed the same:	0.43 <sup>b,c</sup>	0.38 <sup>b,c</sup>	0.40 <sup>b,c</sup> (211-245)
– Got a little worse:	- 0.82 <sup>c</sup>	- 0.48 <sup>c,d</sup>	- 0.79 <sup>c,d</sup> (114-138)
– Got a lot worse:	- 3.16 <sup>d</sup>	- 1.94 <sup>d</sup>	-2.28 <sup>d</sup> (29-31)
n A			

### Change in PROMIS Physical Functioning (W3 - W2) by Change in Global PF (n = 439)

 $PF_2 = 39.95 (SD = 9.24); PF_3 = 40.07 (SD = 9.60)$ F (4, 434 dfs) = 4.70, p = 0.0010 r = 0.20 (0.20)

	Global PF	n
-2+ levels better:	1.84 <sup>a</sup>	27
– <u>1 level better</u> :	0.54 <sup>a,b</sup>	74
– Stayed the same:	0.25 <sup>b</sup>	235
– <u>1 level worse</u> :	- 0.86 <sup>b,c</sup>	77
-2+ levels worse:	- 1.67 <sup>c</sup>	26

### Change in PROMIS Physical Functioning (W3 - W1) by **Retrospective Rating of Change, Overall Health (n = 451)**

 $PF_1 = 40.18 (SD = 9.03); PF_3 = 39.91 (SD = 9.54)$ F(4, 446 dfs) = 13.34, p < .0001 r = 0.21 (0.23)

- Much better (n = 38): 1.26<sup>a</sup> - Somewhat better (n = 221): 0.29<sup>a</sup> - About the same (n = 39): - 2.57<sup>b</sup> - Somewhat worse (n = 34): **1.45**<sup>a</sup> - 1.51<sup>b</sup>
- Much worse (n = 119):

### Change in PROMIS Physical Functioning (W3 - W1) by Retrospective Rating of Change, Overall Health (n = 451)

 $PF_1 = 40.18 (SD = 9.03); PF_3 = 39.91 (SD = 9.54)$ F (4, 446 dfs) = 13.34, p<.0001 r = 0.21 (0.23)



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

- "Responder"
  - Change in score over a predetermined time period that should be interpreted as a treatment benefit.
  - Empiric evidence for any responder definition is *derived using anchor-based method*.

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

### Standard Error of Measurement (SEM) and Reliable Change Index (RCI)

- SEM = SD \* SQRT (1-reliability)
  - 95% CI = Estimated true score +/- 1.96 \* SEM
- RCI= (X<sub>2</sub> X<sub>1</sub>)/ (SEM \* SQRT (2))

Note: The change needed to be significant for an individual is about 0.50 SD for the SEM and 0.70 SD for the RCI when reliability is 0.94.

Expert Review of Pharmacoeconomics & Outcomes

### Individual Significant Change for SF-36 Scales and Summary Scores

	%	%	
	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%

### Thank you.

- Farivar, S. S., Liu, H., & Hays, R. D. (2004). Half standard deviation estimate of the minimally important difference in HRQOL scores?. <u>Expert Review of Pharmacoeconomics and Outcomes</u> <u>Research.</u>, <u>4</u> (5), 515-523.
- Hays, R. D., Brodsky, M., Johnston, M. F., Spritzer, K. L., & Hui, K. (2005). Evaluating the statistical significance of health-related quality of life change in individual patients. <u>Evaluation</u> and the Health Professions, 28, 160-171.
- Hays, R. D., Farivar, S. S., & Liu, H. (2005). Approaches and recommendations for estimating minimally important differences for health-related quality of life measures. <u>COPD: Journal of Chronic Obstructive Pulmonary Disease</u>, 2, 63-67.
- Hays, R. D., & Woolley, J. M. (2000). The concept of clinically meaningful difference in healthrelated quality-of-life research: How meaningful is it? <u>PharmacoEconomics</u>, <u>18</u>, 419-423.
- Revicki, D. A., Cella, D., Hays, R. D., Sloan, J. A., Lenderking, W. R., & Aaronson, N. K. (2006). Responsiveness and minimal important differences for patient reported outcomes. <u>Health and</u> <u>Quality of Life Outcomes</u>, <u>4</u>: 70.
- Revicki, D., Hays, R. D., Cella, D., & Sloan, J. (2008). Recommended methods for determining responsiveness and minimally important differences for patient reported outcomes. <u>Journal of Clinical Epidemiology</u>., <u>61</u>, 102-109



### **Appendix A: "Distribution-Based" Methods**

 Change in PROMIS domain score theta that is equal to a "prior" for the minimally important change

$$-0.5^*SD_b = 5$$
  
 $-0.2^*SD_b = 2$ 

- Standard error of measurement (SEM) =  $SD_b \times \sqrt{1 r_{xx}}$ 
  - $SD_b$  = standard deviation at baseline  $r_{xx}$  = reliability

### Appendix B: Change in HAQ (W2 - W1) by Retrospective Ratings of Change (n = 461) in Rheumatoid Arthritis

 $HAQ_1 = 0.88 (SD = 0.72); HAQ_2 = 0.92 (SD = 0.74)$ 

F (4, 456 dfs) = 18.72, p<.0001(Activity) r = 0.34 (0.30)

F (4, 455 dfs) = 10.52, p<.0001(Fatigue) r = 0.28 (0.25)

F (4, 455 dfs) = 16.94, p < .0001(Pain) r = 0.35 (0.35)

	Activity	Fatigue	Pain	n
-Got a lot better:	+ 0.11 <sup>a</sup>	0.10 <sup>a</sup>	0.16 <sup>a</sup>	(19-21)
– Got a little better:	+ 0.10 <sup>a</sup>	0.06 <sup>a</sup>	0.08 <sup>a,b</sup>	(41-61)
- Stayed the same:	0.00 <sup>a,b</sup>	0.00 <sup>a,b</sup>	0.00 <sup>b,c</sup>	(223-257)
– Got a little worse:	- 0.09 <sup>b</sup>	- 0.10 <sup>b</sup>	- 0.10 <sup>c</sup>	(107-125)
-Got a lot worse:	- 0.37 <sup>c</sup>	- 0.26 <sup>c</sup>	- 0.31 <sup>d</sup>	(28-35)