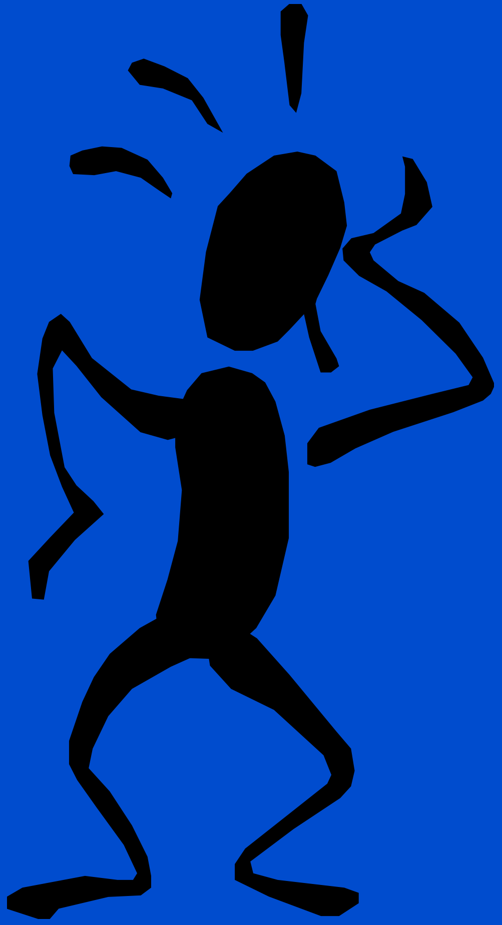
Introduction to Psychometric Analysis of Survey Data

Ron D. Hays, Ph.D. (hays@rand.org)

July 10, 2003 (10:30-12:30pm)

Lori Shiotani



What kind of data collection errors are possible?

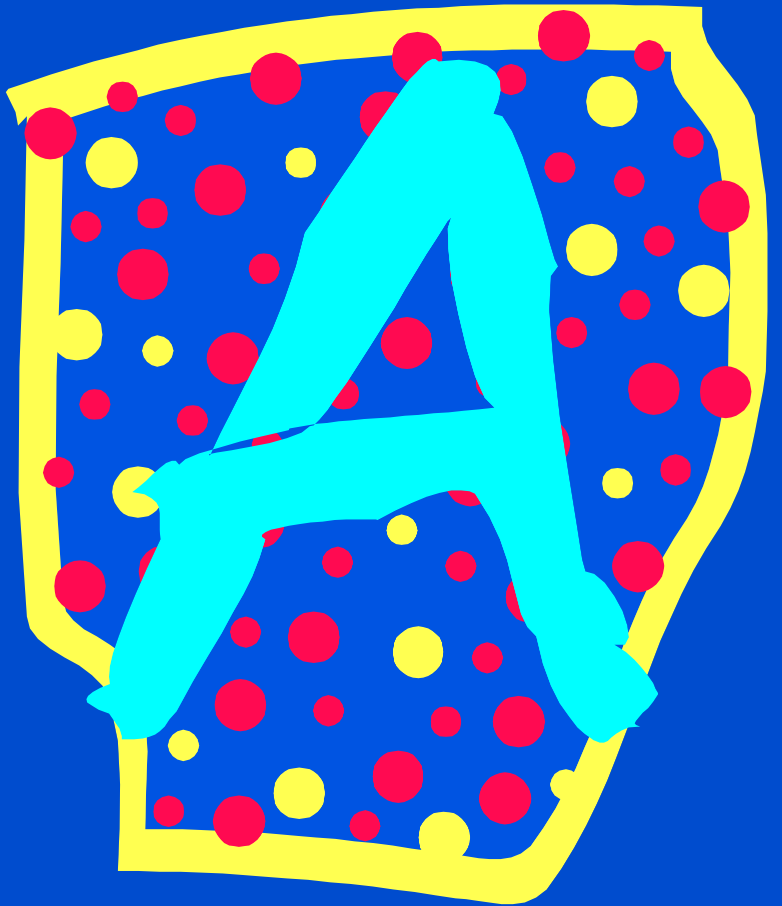
Data Collection Errors

- Do respondents represent underlying population?
  - Coverage Error (every person in population is not included in the sampling frame)
  - Sampling Error (only some members of the population are sampled)
  - Non-response error (those who response are different from those who do)
- Are inaccurate answers given to survey questions?
  - Measurement error

What's a Good Measure?

- It is practical to use (feasibility)
- Same person gets same score (reliability)
- Different people get different scores (validity)
- People get scores you expect (validity)

Peter Chin



How are good measures developed?

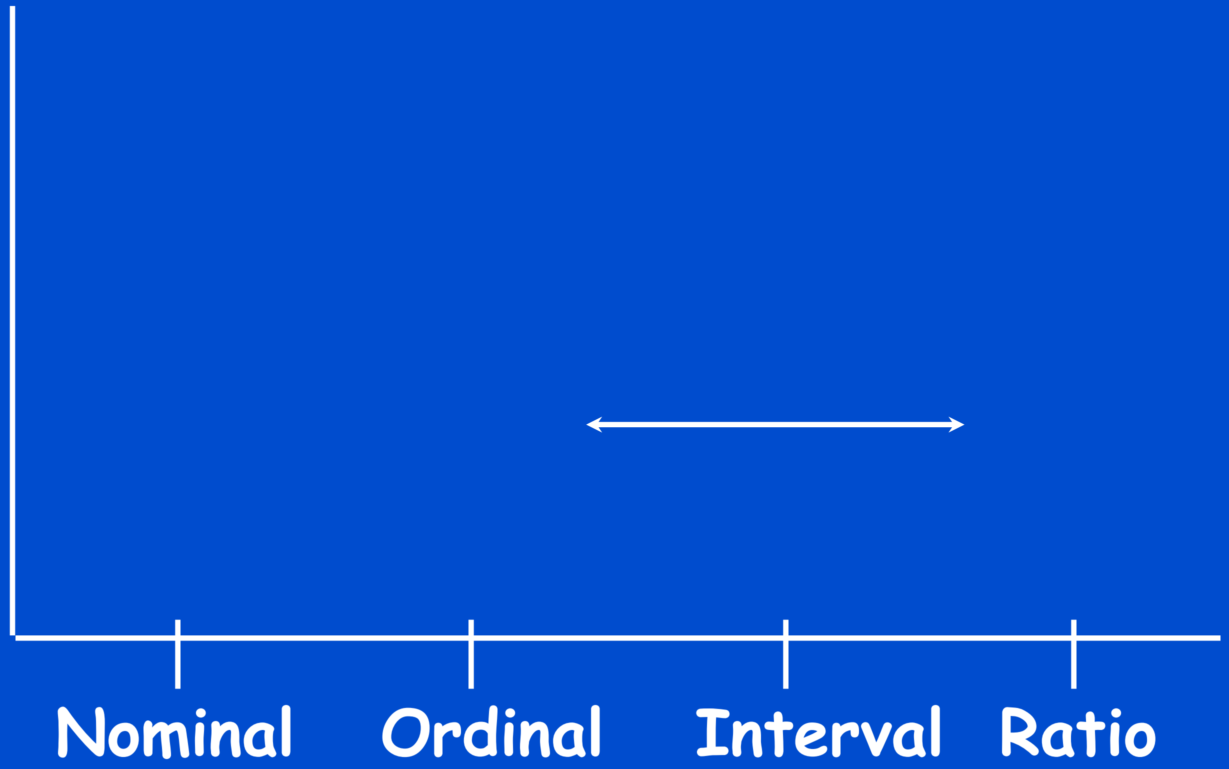
How Are Good Measures Developed?

- Review literature
- Expert input (patients and clinicians)
- Define constructs you are interested in
- Draft items (item generation)
- Pretest
    - Cognitive interviews
    - Field and pilot testing
- Revise and test again
- Translate/harmonize across languages

Scales of Measurement and Their Properties

	Property of Numbers		
Type of Scale	Rank Order	Equal Interval	Absolute 0
Nominal	No	No	No
Ordinal	Yes	No	No
Interval	Yes	Yes	No
Ratio	Yes	Yes	Yes

Measurement Range for Health Services Measures

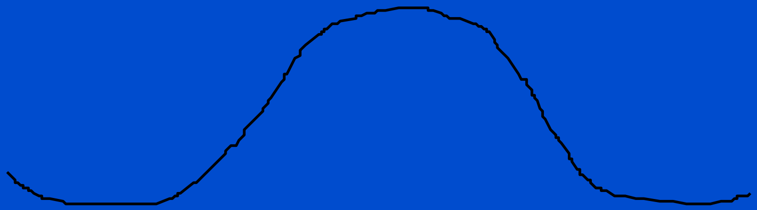


Indicators of Acceptability

- Response rate
- Missing data (item, scale)
- Administration time

Variability

- All scale levels are represented
- Distribution approximates bell-shaped "normal"



Measurement Error

observed = true score + systematic error + random error

(bias)

Flavors of Reliability

- Test-retest (administrations)
- Intra-rater (raters)
- Internal consistency (items)

Test-retest Reliability of MMPI 317-362

		MMPI 317		
		True	False	
MMPI 362	True	169	15	184
	False	21	95	116
		190	110	

I am more sensitive than most other people. (r = 0.75)

Kappa Coefficient of Agreement (Corrects for Chance)

$$\text{kappa} = \frac{(\text{observed - chance})}{(1 - \text{chance})}$$

Example of Computing KAPPA

				Rater A			Row
		1	2	3	4	5	Sum
	1	1	1				2
	2		2				2
Rater B	3			2			2
	4				2		2
	5					2	2
Column Sum		1	3	2	2	2	10

Example of Computing KAPPA (Continued)

$$P_c = \frac{(1 \times 2) + (3 \times 2) + (2 \times 2) + (2 \times 2) + (2 \times 2)}{(10 \times 10)} = \boxed{0.20}$$

$$P_{obs.} = \frac{9}{10} = \boxed{0.90}$$

$$\text{Kappa} = \frac{0.90 - 0.20}{1 - 0.20} = \boxed{0.87}$$

Guidelines for Interpreting Kappa

Conclusion	Kappa
Poor	< .40
Fair	.40 - .59
Good	.60 - .74
Excellent	> .74

Fleiss (1981)

Conclusion	Kappa
Poor	< 0.0
Slight	.00 - .20
Fair	.21 - .40
Moderate	.41 - .60
Substantial	.61 - .80
Almost perfect	.81 - 1.00

Landis and Koch (1977)

Ratings of Height of Houseplants

Plant		Baseline Height	Follow-up Height	Experimental Condition
A1				
	R1	120	121	1
	R2	118	120	
A2				
	R1	084	085	2
	R2	096	088	
B1				
	R1	107	108	2
	R2	105	104	
B2				
	R1	094	100	1
	R2	097	104	
C1				
	R1	085	088	2
	R2	091	096	

Ratings of Height of Houseplants (Cont.)

Plant		Baseline Height	Follow-up Height	Experimental Condition
C2				
	R1	079	086	1
	R2	078	092	
D1				
	R1	070	076	1
	R2	072	080	
D2				
	R1	054	056	2
	R2	056	060	
E1				
	R1	085	101	1
	R2	097	108	
E2				
	R1	090	084	2
	R2	092	096	

Reliability of Baseline Houseplant Ratings

Ratings of Height of Plants: 10 plants, 2 raters

Baseline Results

Source	DF	SS	MS	F
Plants	9	5658	628.667	35.52
Within	10	177	17.700	
Raters	1	57.8	57.800	
Raters x Plants	9	119.2	13.244	
Total	19	5835		

Sources of Variance in Baseline Houseplant Height

Source	dfs	MS	
Plants (N)	9	628.67	(BMS)
Within	10	17.70	(WMS)
Raters (K)	1	57.80	(JMS)
Raters x Plants	9	13.24	(EMS)
Total	19		

Intraclass Correlation and Reliability

Model	Reliability	Intraclass Correlation
One-Way	$\frac{MS_{BMS} - MS_{WMS}}{MS_{BMS}}$	$\frac{MS_{BMS} - MS_{WMS}}{MS_{BMS} + (K-1)MS_{WMS}}$
Two-Way Fixed	$\frac{MS_{BMS} - MS_{EMS}}{MS_{BMS}}$	$\frac{MS_{BMS} - MS_{EMS}}{MS_{EMS} + (K-1)MS_{EMS}}$
Two-Way Random	$\frac{N(MS_{BMS} - MS_{EMS})}{NMS_{BMS} + MS_{JMS} - MS_{EMS}}$	$\frac{MS_{BMS} - MS_{EMS}}{MS_{BMS} + (K-1)MS_{EMS} + K(MS_{JMS} - MS_{EMS})/N}$

Summary of Reliability of Plant Ratings

	Baseline		Follow-up	
	R_{TT}	R_{II}	R_{TT}	R_{II}
One-Way Anova	0.97	0.95	0.97	0.94
Two-Way Random Effects	0.97	0.95	0.97	0.94
Two-Way Fixed Effects	0.98	0.96	0.98	0.97

Source	Label	Baseline MS
Plants	BMS	628.667
Within	WMS	17.700
Raters	JMS	57.800
Raters X Plants	EMS	13.244

Cronbach's Alpha

Source	df	SS	MS
Respondents (BMS)	4	11.6	2.9
Items (JMS)	1	0.1	0.1
Resp. x Items (EMS)	4	4.4	1.1
Total	9	16.1	

$$\text{Alpha} = \frac{2.9 - 1.1}{2.9} = \frac{1.8}{2.9} = \boxed{0.62}$$

Alpha by Number of Items and Inter-item Correlations

$$\text{alpha}_{st} = \frac{K\,\bar{r}}{1 + (K - 1)\,\bar{r}}$$

K = number of items in scale

Alpha for Different Numbers of Items and Homogeneity

	Average Inter-item Correlation ($\bar{r}$)					
Number of Items (K)	.0	.2	.4	.6	.8	1.0
2	.000	.333	.572	.750	.889	1.000
4	.000	.500	.727	.857	.941	1.000
6	.000	.600	.800	.900	.960	1.000
8	.000	.666	.842	.924	.970	1.000

Spearman-Brown Prophecy Formula

$$\text{alpha}_y = \left(\frac{N \cdot \text{alpha}_x}{1 + (N - 1) * \text{alpha}_x} \right)$$

N = how much longer scale y is than scale x

Reliability Minimum Standards

- 0.70 or above (for group comparisons)
- 0.90 or higher (for individual assessment)
  - SEM = SD $(1- \text{reliability})^{1/2}$

Reliability of a Composite Score

$$\text{Mosier} = 1 - \frac{\Sigma(w_j^2)(S_j^2) - \Sigma(w_j^2)(S_j^2)(\alpha_j)}{\Sigma(w_j^2)(S_j^2) + 2\Sigma(w_j)(w_K)(S_j)(S_K)(r_{jK})}$$

w_j = weight given to component J

w_K = weight given to component K

S_j = standard deviation of J

α_j = reliability of J

r_{jK} = correlation between J and K

Hypothetical Multitrait/Multi-Item Correlation Matrix

	Trait #1	Trait #2	Trait #3
Item #1	0.80*	0.20	0.20
Item #2	0.80*	0.20	0.20
Item #3	0.80*	0.20	0.20
Item #4	0.20	0.80*	0.20
Item #5	0.20	0.80*	0.20
Item #6	0.20	0.80*	0.20
Item #7	0.20	0.20	0.80*
Item #8	0.20	0.20	0.80*
Item #9	0.20	0.20	0.80*

*Item-scale correlation, corrected for overlap.

Multitrait/Multi-Item Correlation Matrix for Patient Satisfaction Ratings

	Technical	Interpersonal	Communication	Financial
Technical				
1	0.66*	0.63†	0.67†	0.28
2	0.55*	0.54†	0.50†	0.25
3	0.48*	0.41	0.44†	0.26
4	0.59*	0.53	0.56†	0.26
5	0.55*	0.60†	0.56†	0.16
6	0.59*	0.58†	0.57†	0.23
Interpersonal				
1	0.58	0.68*	0.63†	0.24
2	0.59†	0.58*	0.61†	0.18
3	0.62†	0.65*	0.67†	0.19
4	0.53†	0.57*	0.60†	0.32
5	0.54	0.62*	0.58†	0.18
6	0.48†	0.48*	0.46†	0.24

Note – Standard error of correlation is 0.03. Technical = satisfaction with technical quality. Interpersonal = satisfaction with the interpersonal aspects. Communication = satisfaction with communication. Financial = satisfaction with financial arrangements. *Item-scale correlations for hypothesized scales (corrected for item overlap). †Correlation within two standard errors of the correlation of the item with its hypothesized scale.

Forms of Validity

- Content
- Criterion
- Construct Validity
  - Measure’s relationships with other things are consistent with hypotheses/theory.
  - Includes responsiveness to change

Relative Validity Example

	Severity of Heart Disease				
	None	Mild	Severe	F-ratio	Relative Validity
Scale #1	87	90	91	2	---
Scale #2	74	78	88	10	5
Scale #3	77	87	95	20	10

Responsiveness to Change

- Measures should reflect true change
- Evaluating responsiveness requires an external indicator of change (anchor)

Responsiveness Indices

(1) Effect size (ES) = D/SD

(2) Standardized Response Mean (SRM) = $D/SD^{\dagger}$

(3) Guyatt responsiveness statistic (RS) = $D/SD^{\ddagger}$

D = raw score change in “changed” group;
SD = baseline SD;
$SD^{\dagger}$ = SD of D;
$SD^{\ddagger}$ = SD of D among “unchanged”

Kinds of Anchors

- Self-report
- Clinician or other report
- Clinical parameter
- Clinical intervention

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

Examples of Other Anchors

Clinician report

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

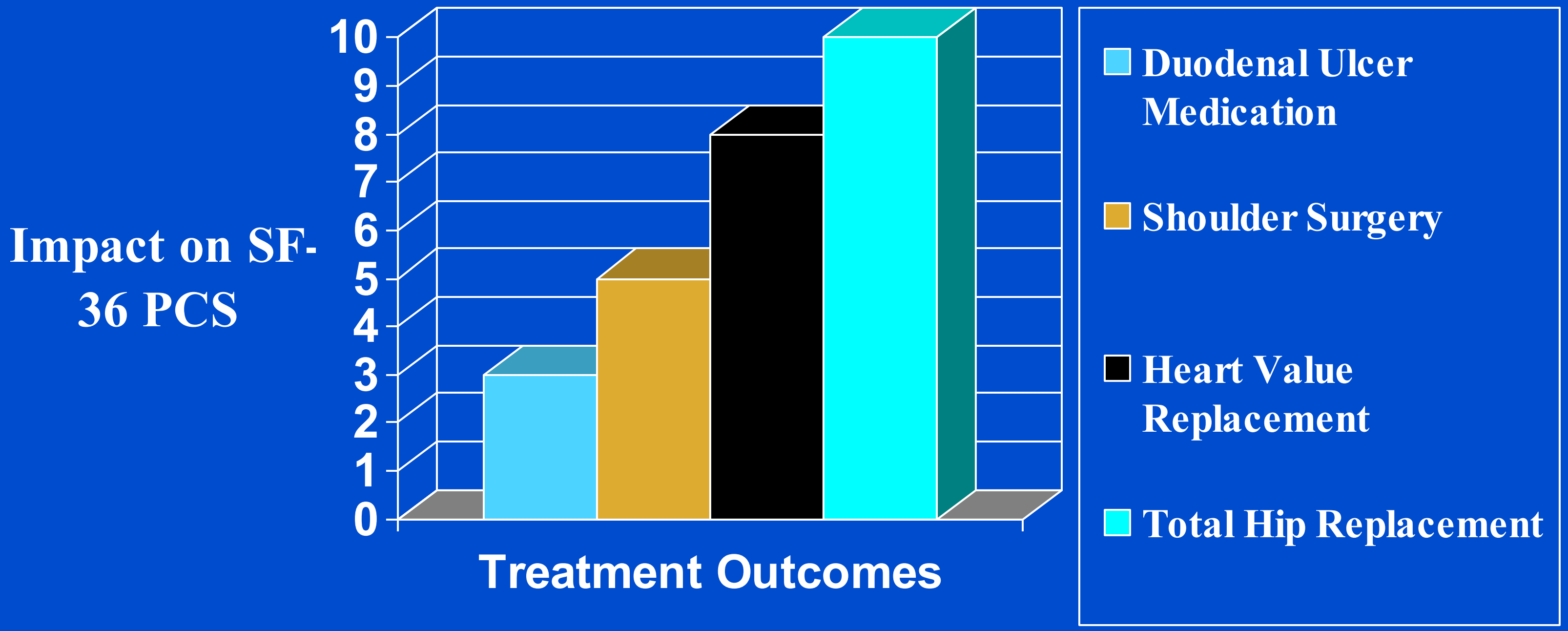
Clinical parameter

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

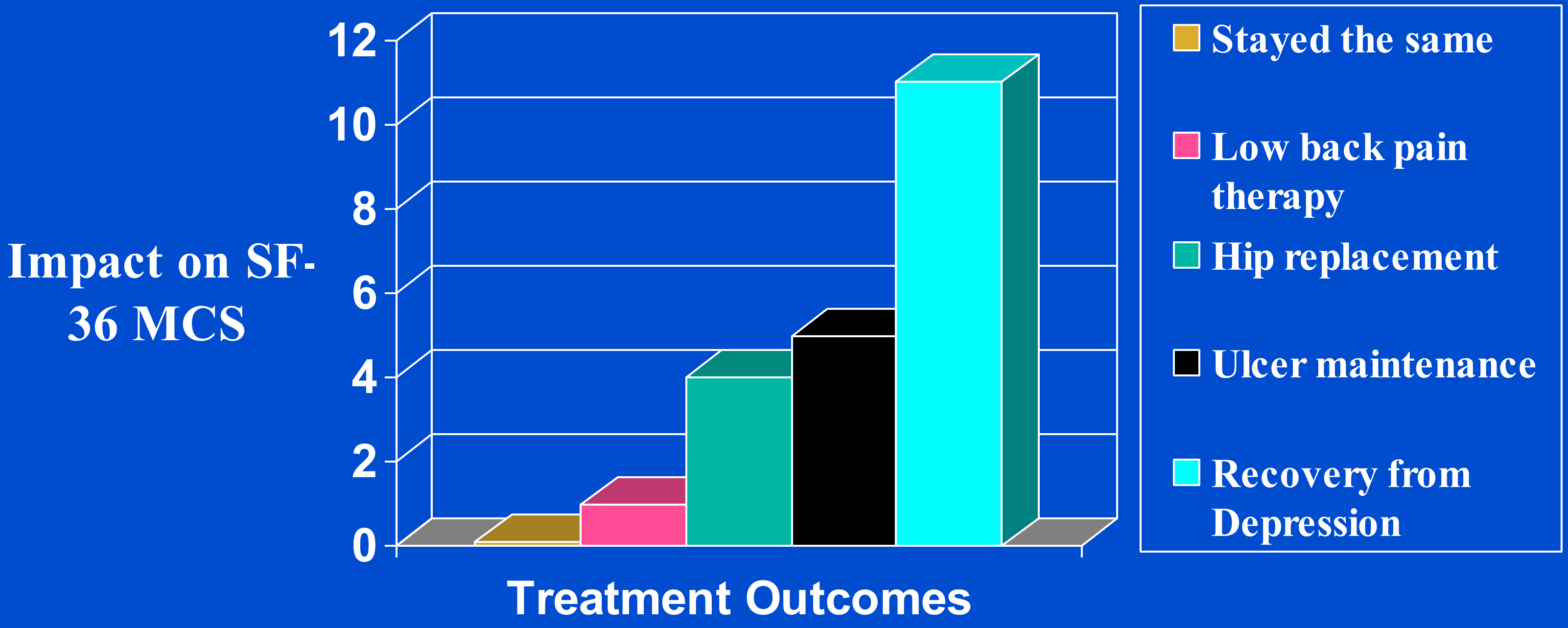
Clinical intervention

- Before and after Prozac

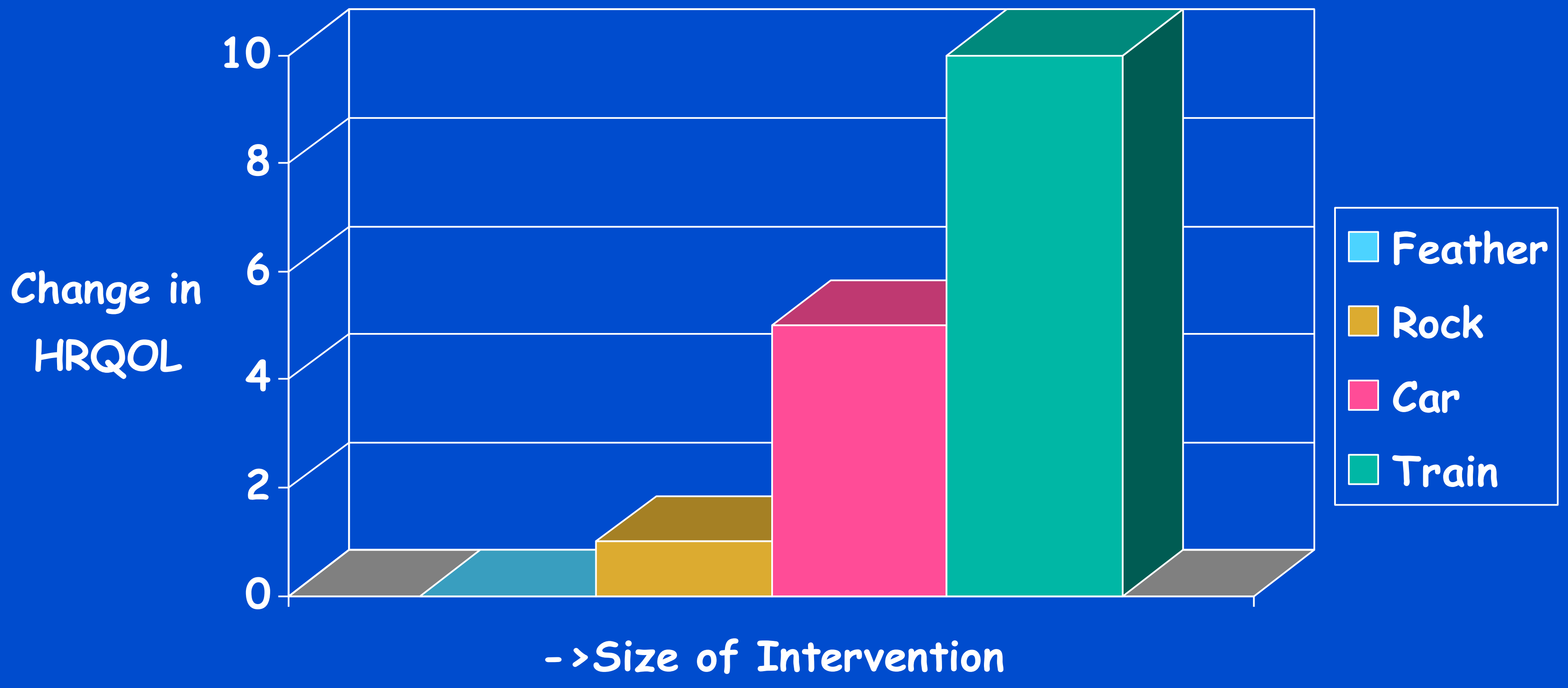
Change and Responsiveness in PCS Depends on Treatment



Change and Responsiveness in MCS Depends on Treatment



Magnitude of HRQOL Change Should Parallel Underlying Change



Minimally Important Difference (MID)

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 (important).

Change large enough for a clinician to base treatment decisions upon it.

Identifying the MID

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

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.

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

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

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

Scale	Mean (ES)	Range	SD	Range/SD
PF	8 (.4)	2 (6 - 8)	20	.10
Role-P	16 (.4)	10 (11-21)	40	.25
Pain	11 (.5)	8 (7-15)	20	.40
GH	2 (.1)	3 (1- 4)	20	.15
EWB	4 (.2)	6 (1- 7)	20	.30
Role-E	13 (.2)	10 (8-18)	40	.25
SF	9 (.5)	4 (8-12)	20	.20
EF	8 (.4)	6 (5-11)	20	.30
PCS	3 (.3)	1 (3- 4)	10	.10
MCS	3 (.3)	3 (2- 5)	10	.30

Resource Centers for
Minority Aging Research

RCMAR

This class 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).