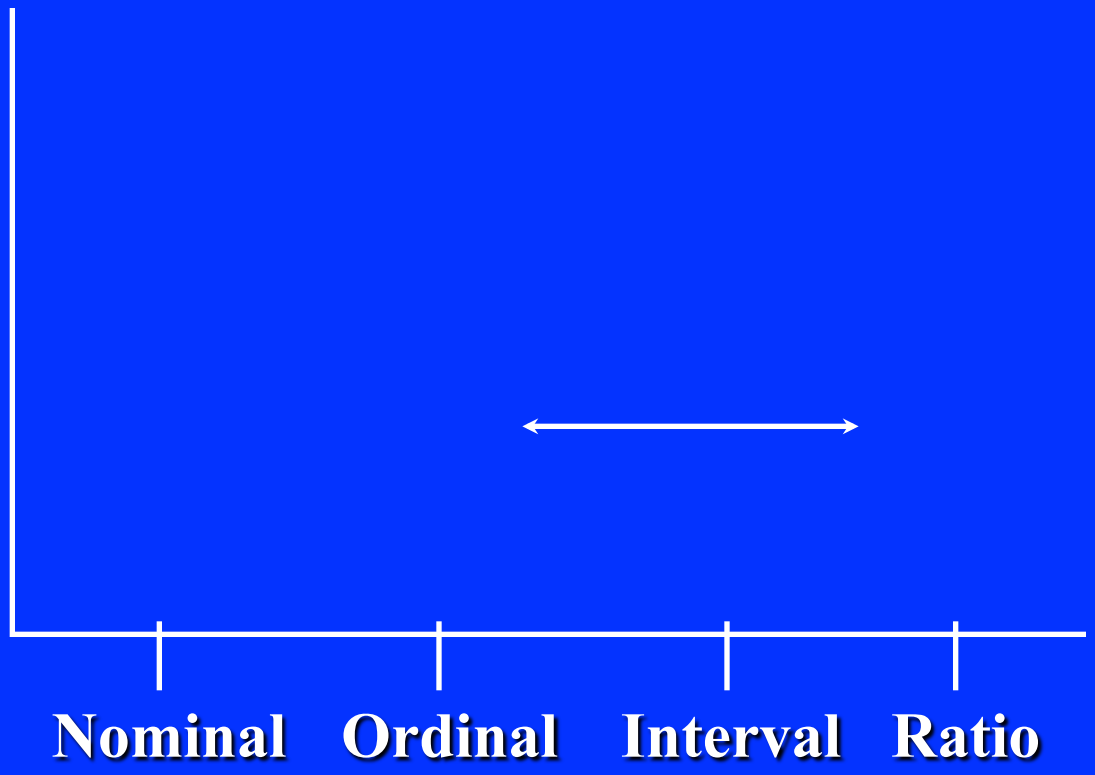
Quantitative Methods: Reliability, Validity, and Use of SEM to Assess Psychometric Equivalence

Ron D. Hays, Ph.D.
UCLA Department of Medicine
Los Angeles, CA

November 18, 2005

Measurement Range for Health Outcome Measures



Indicators of Acceptability

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

Variability

- All scale levels are represented
- Distribution approximates bell-shaped “normal”

Measurement Error

Observed = true + systematic + random
score error error

(bias)

Flavors of Reliability

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

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	$\frac{N(MS_{BMS} - MS_{EMS})}{NMS_{BMS} + MS_{JMS} - MS_{EMS}}$	$\frac{MS_{BMS} - MS_{EMS}}{MS_{BMS} + (K-1)MS_{EMS} + K(MS_{JMS} - MS_{EMS})/N}$

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*	

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

Reliability Minimum Standards

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

Construct Validity

- Does measure relate to other measures in ways consistent with hypothesis?
- Responsiveness to change

Responsiveness to Change and Minimally Important Difference

- HRQOL measures should be responsive to interventions that changes HRQOL

- Evaluating responsiveness requires assessment of HRQOL
  - » pre-post intervention of known efficacy
  - » at two times in tandem with anchor
  - » HRQOL change among people who changed on anchor

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

Clinical Anchor

- "changed" group = seizure free (100% reduction in seizure frequency)
- "unchanged" group = < 50% change in seizure frequency

Responsiveness Indices

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

(2) Standardized Response Mean (SRM) = D/SD†

(3) Guyatt responsiveness statistic (RS) = D/SD‡

D = raw score change in "changed" group;

SD = baseline SD;

†SD = SD of D;

‡ SD = SD of D among "unchanged"

Effect Size Benchmarks

- Small: 0.20->0.49
- Moderate: 0.50->0.79
- Large: 0.80 or above

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.

Confirmatory Factor Analysis

- Compares observed covariances with covariances generated by hypothesized model
- Statistical and practical tests of fit
- Factor loadings
- Correlations between factors
- Regression coefficients

Fit Indices

- Normed fit index: $\dfrac{\chi^2_{null} - \chi^2_{model}}{\chi^2_{null}}$
- Non-normed fit index: $\dfrac{\dfrac{\chi^2_{null}}{df_{null}} - \dfrac{\chi^2_{model}}{df_{model}}}{\left[\dfrac{\chi^2_{null}}{df_{null}} - 1\right]}$
- Comparative fit index: $1 - \left[\dfrac{\chi^2_{model} - df_{model}}{\chi^2_{null} - df_{null}}\right]$

Parameter Estimates (standardized) for Confirmatory Factor Analytic Model Estimated Separately Among English and Spanish (in parentheses) Respondents

	Physical Health	*Mental Health*	*Evaluations of Care*
Physical functioning	0.69 (0.25)		
Role–physical	0.81 (0.53)		
Pain	0.69 (0.88)	0.12 (0.02[a])	
General health perceptions	0.56 (0.39)	0.28 (0.34)	
Emotional well-being		0.86 (0.93)	
Role–emotional		0.62 (0.44)	
Energy/fatigue	0.38 (0.22)	0.56 (0.69)	
Social functioning	0.41 (0.45)	0.35 (0.36)	
Doctor			0.96 (0.97)
Overall			0.93 (0.92)
Access			0.81 (0.80)
Wait time			0.75 (0.79)
Choice			0.73 (0.78)
Plan			0.70 (0.70)

NOTE: Estimated correlations for English and Spanish respondents, respectively, were as follows: physical and mental health (r = .45, .55), mental health and patient evaluations of care (r = .23, .31), and physical health and patient evaluations of care (r = .05, .16[a]).
a. Not statistically significant.

Parameter Estimates (standardized) for Confirmatory Factor Analytic Model Constraining Factor Loadings and Correlations to Be Equal for English and Spanish Respondents

	Physical Health	*Mental Health*	*Evaluations of Care*
Physical functioning	0.69 (0.51)		
Role–physical	0.81 (0.78)		
Pain	0.69 (0.68)	0.12 (0.12)	
General health perceptions	0.56 (0.53)	0.28 (0.27)	
Emotional well-being		0.86 (0.80)	
Role–emotional		0.62 (0.57)	
Energy/fatigue	0.37 (0.37)	0.56 (0.56)	
Social functioning	0.41 (0.48)	0.35 (0.41)	
Doctor			0.96 (0.97)
Overall			0.93 (0.93)
Access			0.81 (0.81)
Wait time			0.75 (0.80)
Choice			0.73 (0.75)
Plan			0.70 (0.74)

NOTE: Estimated correlations for English and Spanish respondents, respectively, were as follows: physical and mental health ($r = .45$), mental health and patient evaluations of care ($r = .23$), and physical health and patient evaluations of care ($r = .06$).

Acknowledgments

- Supported in whole by UCLA Center for Health Improvement in Minority Elders/ Resource Centers for Minority Aging Research, National Institute on Aging (AG-02-004)