Evaluating IRT Assumptions

Ron D. Hays

November 14, 2012 (10:30-10:50am)

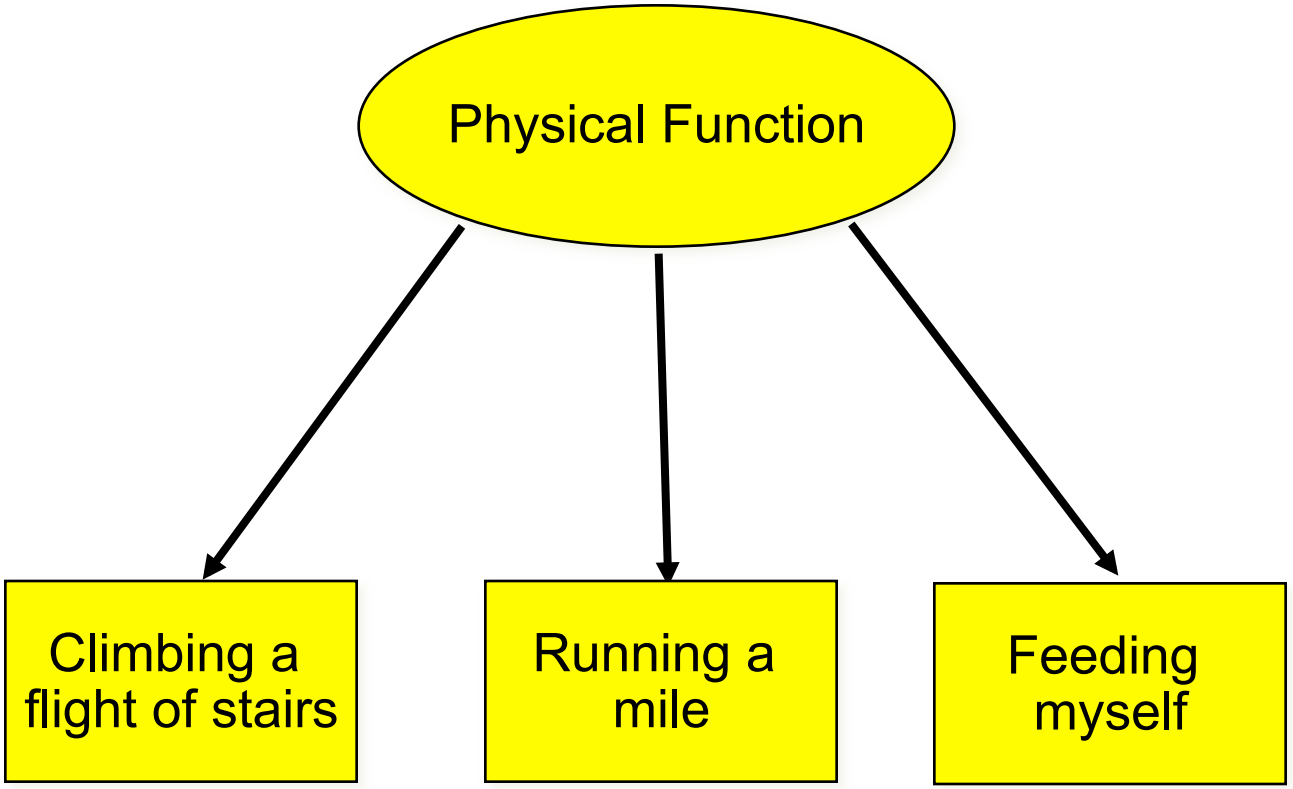
Geriatrics Society of America
Pre-Conference Workshop on
Patient-Reported Outcome Item Banks

San Diego Convention Center (Room 14-A)

IRT Assumptions

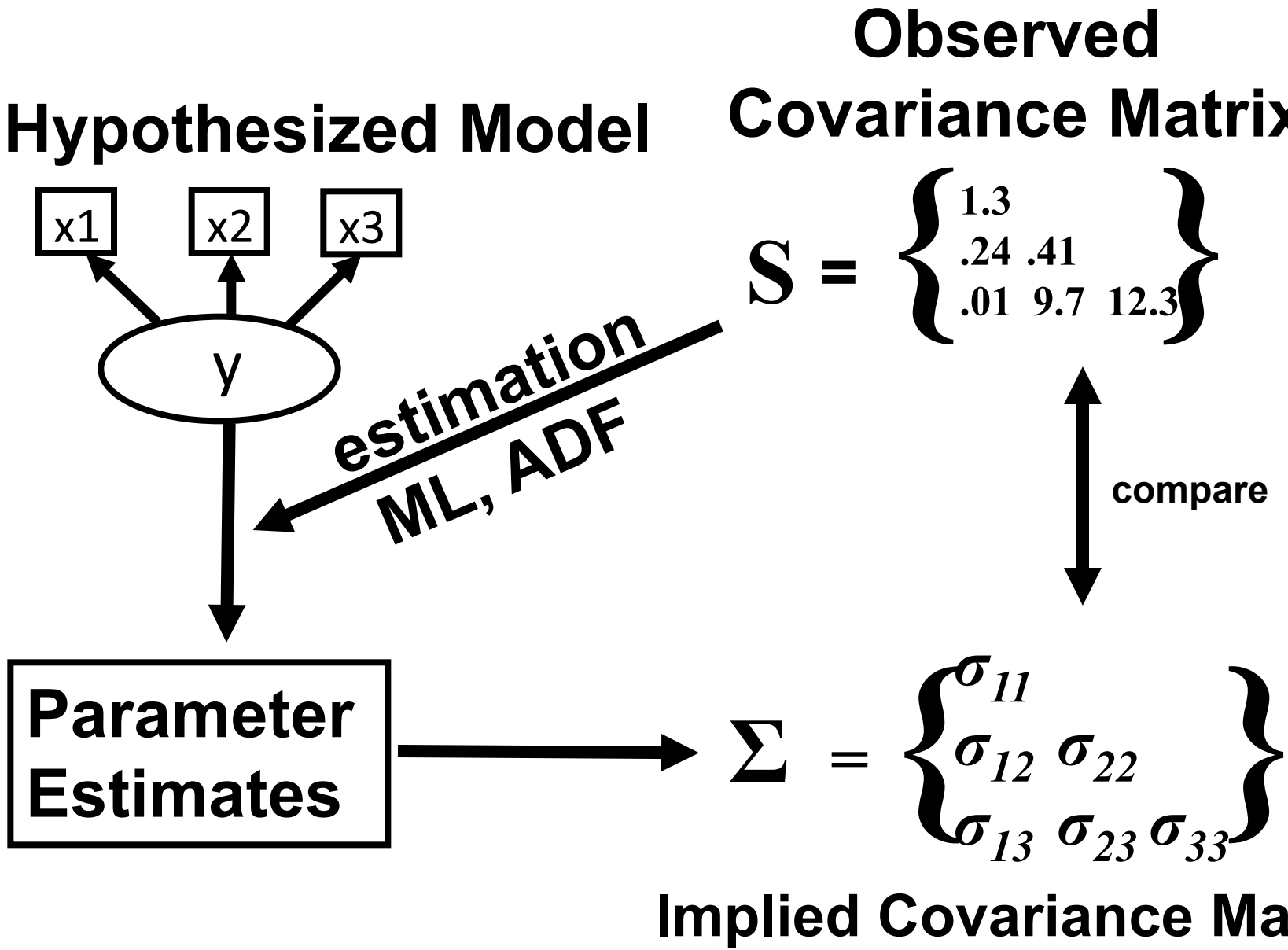
- Dimensionality
  - Unidimensionality for typical models
- Local Independence
- Monotonicity
- Person fit

Hypothesized One-Factor Model



Sufficient Unidimensionality

- One-Factor Categorical Confirmatory Factor Analytic Model (e.g., using Mplus)
  - Polychoric correlations
  - Weighted least squares with adjustments for mean and variance
- Bifactor Model
  - General factor and group-specific factors


Hypothesized Model
x1
x2
x3
y
estimation
ML, ADF
Observed
Covariance Matrix
S =
1.3
.24 .41
.01 9.7 12.3
compare
Parameter
Estimates
Σ =
σ11
σ12 σ22
σ13 σ23 σ33
Implied Covariance Matrix

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: $1 - \left[\dfrac{\chi^2_{model} - df_{model}}{\chi^2_{null} - df_{null}}\right]$

Root Mean Square Error of Approximation (RMSEA)

- Lack of fit per degrees of freedom, controlling for sample size
  - $Q = (s - \sigma(\theta))' W(s - \sigma(\theta))$
  - SQR of (Q/df) – (1/(N – G))
    - RMSEA = SQRT (λ^2 – df)/SQRT (df (N – 1))

- RMSEA < 0.06 desirable
  - Standardized root mean residuals < 0.08
  - Average absolute residual correlations < 0.10

Local Independence

- After controlling for dominant factor(s), item pairs should not be associated.
- Evaluated by looking at size of residual correlations from one-factor model
  - Look for residual correlations > 0.20
- Avoid asking the same item multiple times.
  - "I' m generally sad about my life."
  - "My life is generally sad."

Graded Response Model Parameters for Global Physical Health

Item	a	b1	b2	b3	b4
Global01	7.37 (na)	-1.98 (na)	-0.97 (na)	0.03 (na)	1.13 (na)
Global03	7.65 (2.31)	-1.89 (-2.11)	-0.86 (-0.89)	0.15 (0.29)	1.20 (1.54)
Global06	1.86 (2.99)	-3.57 (-2.80)	-2.24 (-1.78)	-1.35 (-1.04)	-0.58 (-0.40)
Global07	1.13 (1.74)	-5.39 (-3.87)	-2.45 (-1.81)	-0.98 (-0.67)	1.18 (1.00)
Global08	1.35 (1.90)	-4.16 (-3.24)	-2.39 (-1.88)	-0.54 (-0.36)	1.31 (1.17)

Note: Parameter estimates for 5-item scale are shown first, followed by estimates for 4-item scale (in parentheses). na = not applicable

a = discrimination parameter; b1 = 1st threshold; b2 = 2nd threshold; b3 = 3rd threshold; b4 = 4th threshold

Global01: In general, would you say your health is …?

Global03: In general, how would you rate your physical health?

Global06: To what extent are you able to carry out your everyday physical activities?

Global07: How would you rate your pain on average?

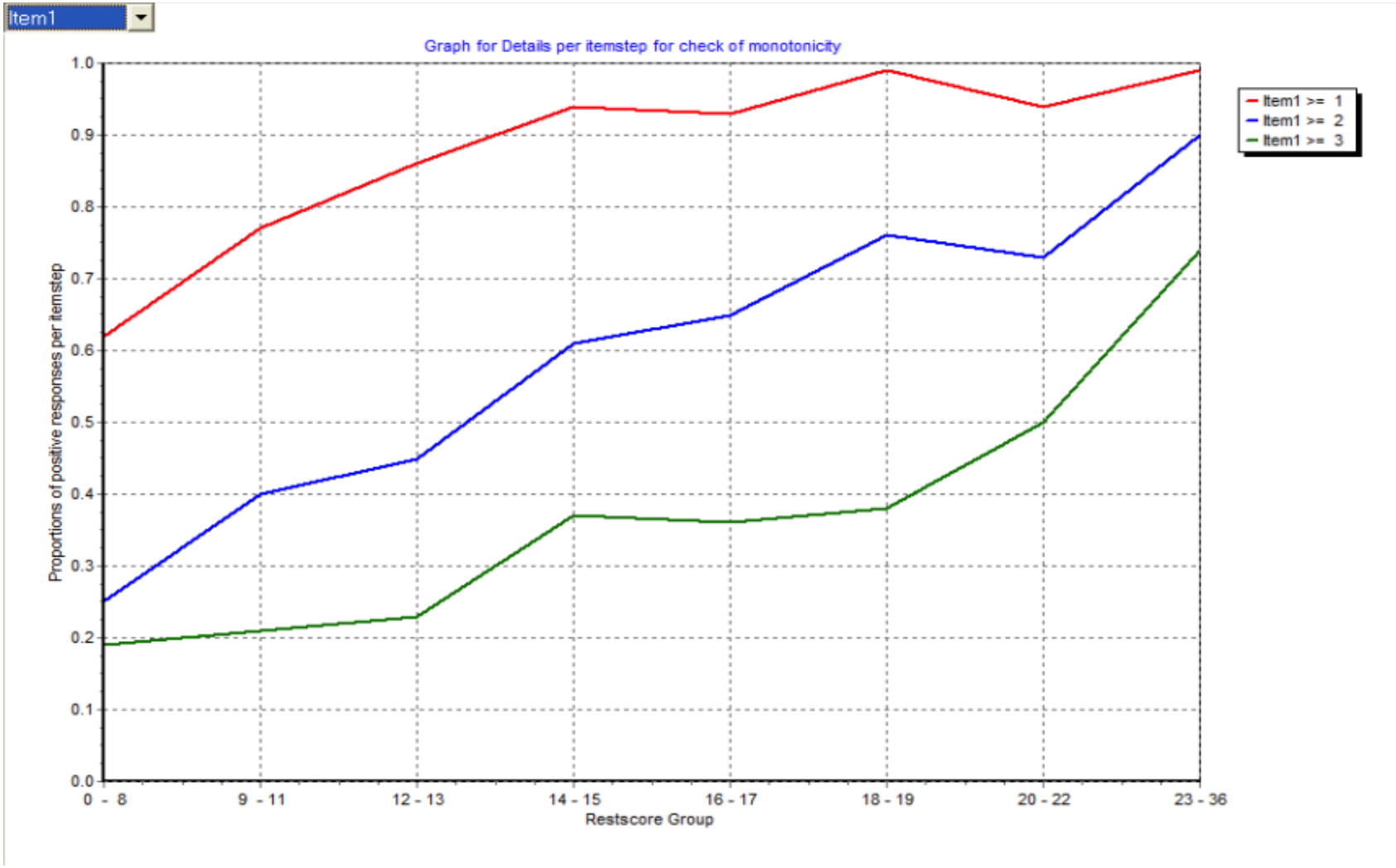
Global08: How would you rate your fatigue on average?

Monotonicity

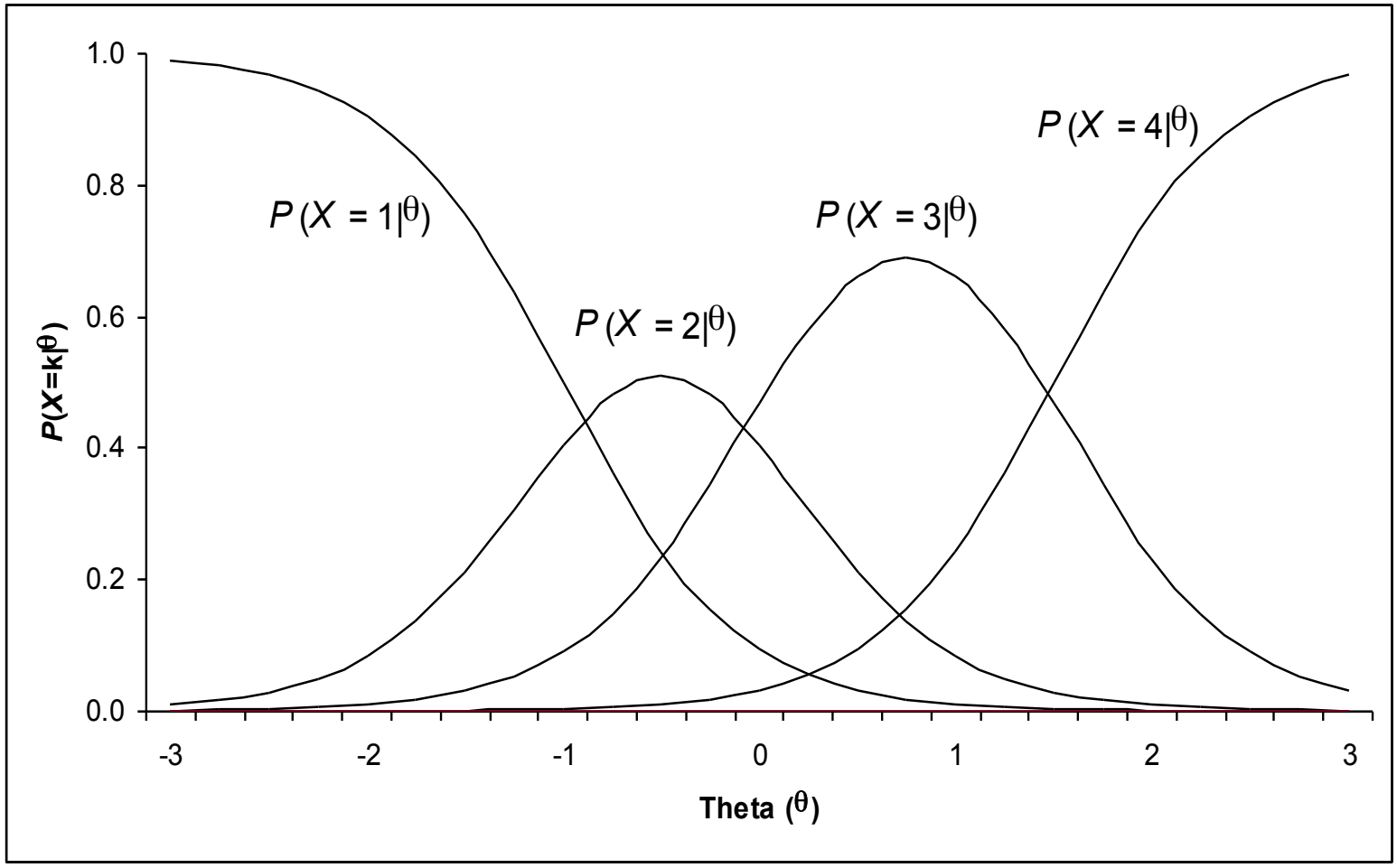
- Probability of selecting a response category indicative of better health should increase as underlying health increases.

- Item response function graphs with
  - y-axis: proportion positive for item step
  - x-axis: raw scale score minus item score

Check of Monotonicity



Samejima's Graded Response Model (Category Response Curves)



IRT Model Fit

- Compare observed and expected response frequencies by item and response category
- Items that do not fit and less discriminating items identified and reviewed by content experts

Person Fit

- Large negative Z_L values indicate misfit.
  - one person who responded to 14 of the PROMIS physical functioning items had a Z_L = -3.13
  - For 13 items the person could do the activity (including running 5 miles) *without any difficulty*.
    - But this person reported *a little difficulty* being out of bed for most of the day.

Person Fit

Item misfit significantly associated with

- *Less than high school education*
- *More chronic conditions*
- *Non-white*

- Including response time in the model lead to significant associations for:
  - *More chronic conditions*
  - *Longer response time*
  - *Younger age*

Acknowledgment of Support

Ron D. Hays (drhays@ucla.edu) received support from the University of California, Los Angeles, Resource Center for Minority Aging Research (RCMAR)/ Center for Health Improvement in Minority Elderly (CHIME), under NIH/NIA Grant P30-AG021684. The content of this presentation does not necessarily represent the official or unofficial views of the NIA or the NIH.