Health-Related Quality of Life as an Indicator of Quality of Care

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

HS216—Quality Assessment: Making the Business Case for Quality

April 12, 2012, 9am-11:30am

RAND Room 1232

http://gim.med.ucla.edu/FacultyPages/Hays/

Students

- Cantrock, Lisa
- Cassell, Charisse
- Danielson, Adam
- Hochman, Michael
- Jennings, Lee
- Moin, Tannaz
- Ray, Lhasa
- Scales, Charles
- Wilkes, Erin
- Zimmerman, Ruth

Patient-Reported Measures

Health-Related Quality of Life is …

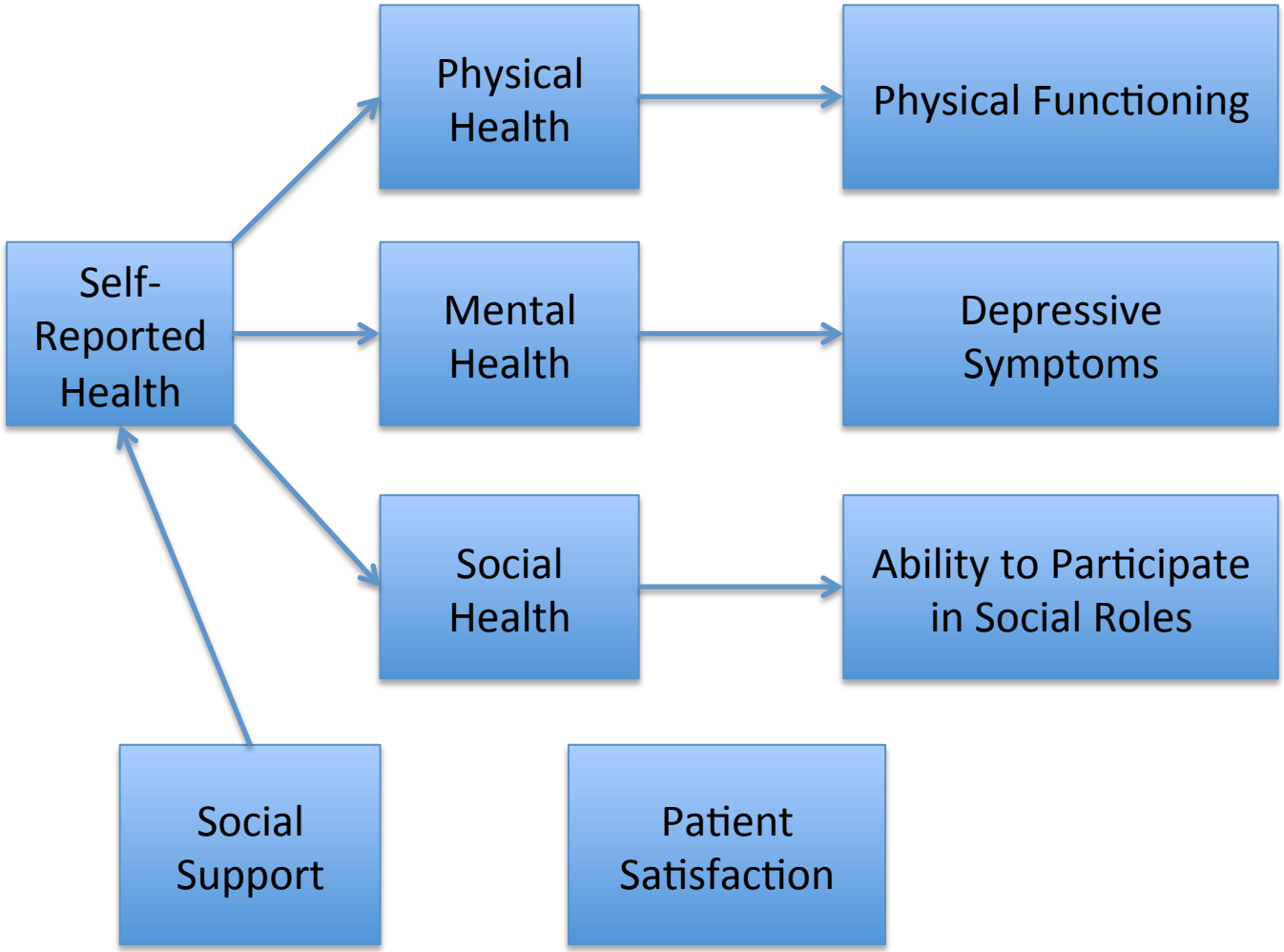
What you can do.

- Functioning
  - Self-care
  - Role
  - Social

How you feel about your life.

- Well-being
  - Emotional well-being
  - Pain
  - Energy

HRQOL Framework



SF-36® Generic Profile Measure

- Functioning
  - Physical functioning (10 items)
  - Role limitations/physical (4 items)
  - Role limitations/emotional (3 items)
  - Social functioning (2 items)
- Well-Being
  - Emotional well-being (5 items)
  - Energy/fatigue (4 items)
  - Pain (2 items)
  - General health perceptions (5 items)

6

Scoring HRQOL Profile Scales

- Average or sum all items in the same scale.

- Transform average or sum to
  - 0 (worse) to 100 (best) possible range
  - z-score (mean = 0, SD = 1)
  - T-score (mean = 50, SD = 10)

7

Formula for Transforming Scores

$$X = \frac{(\text{original score} - \text{minimum}) * 100}{(\text{maximum} - \text{minimum})}$$

$$Y = \text{target mean} + (\text{target SD} * Zx)$$

$$Z_X = \frac{(X - \bar{X})}{SD_X}$$

8

SF-36 online

- http://www.sf-36.org/demos/SF-36v2.html

Generic vs. Disease-Targeted

- ✓ In general, would you say your health is:

Excellent/ Very good/ Good/ Fair/ Poor?

- ✓ How much does kidney disease bother you in your ability to work around the house?

Not at all bothered/Somewhat bothered/
Moderately bothered/Very much bothered/
Extremely bothered

10

KDQOL-36 (24 targeted items)

- **Items 1-12:** SF-12
- **Items 13-16:** Burden of Kidney Disease (4)
- **Items 17-28:** Symptoms/Problems (12)
- **Items 29-36:** Effects of Kidney Disease (8)

Burden of Kidney Disease

- My kidney disease interferes too much with my life.
- Too much of my time is spent dealing with my kidney disease.
- I feel frustrated with my kidney disease
- I feel like a burden on my family.

Effects of Kidney Disease

How much does kidney disease bother you in …

- Fluid restrictions?
- Dietary restriction?
- Your ability to work around the house?
- Your ability to travel?
- Being dependent on doctors and other medical staff?
- Stress or worries caused by kidney disease?
- Your sex life?
- Your personal appearance?

Ultimate Use of HRQOL Measures- Helping to Ensure Access to Cost-Effective Care

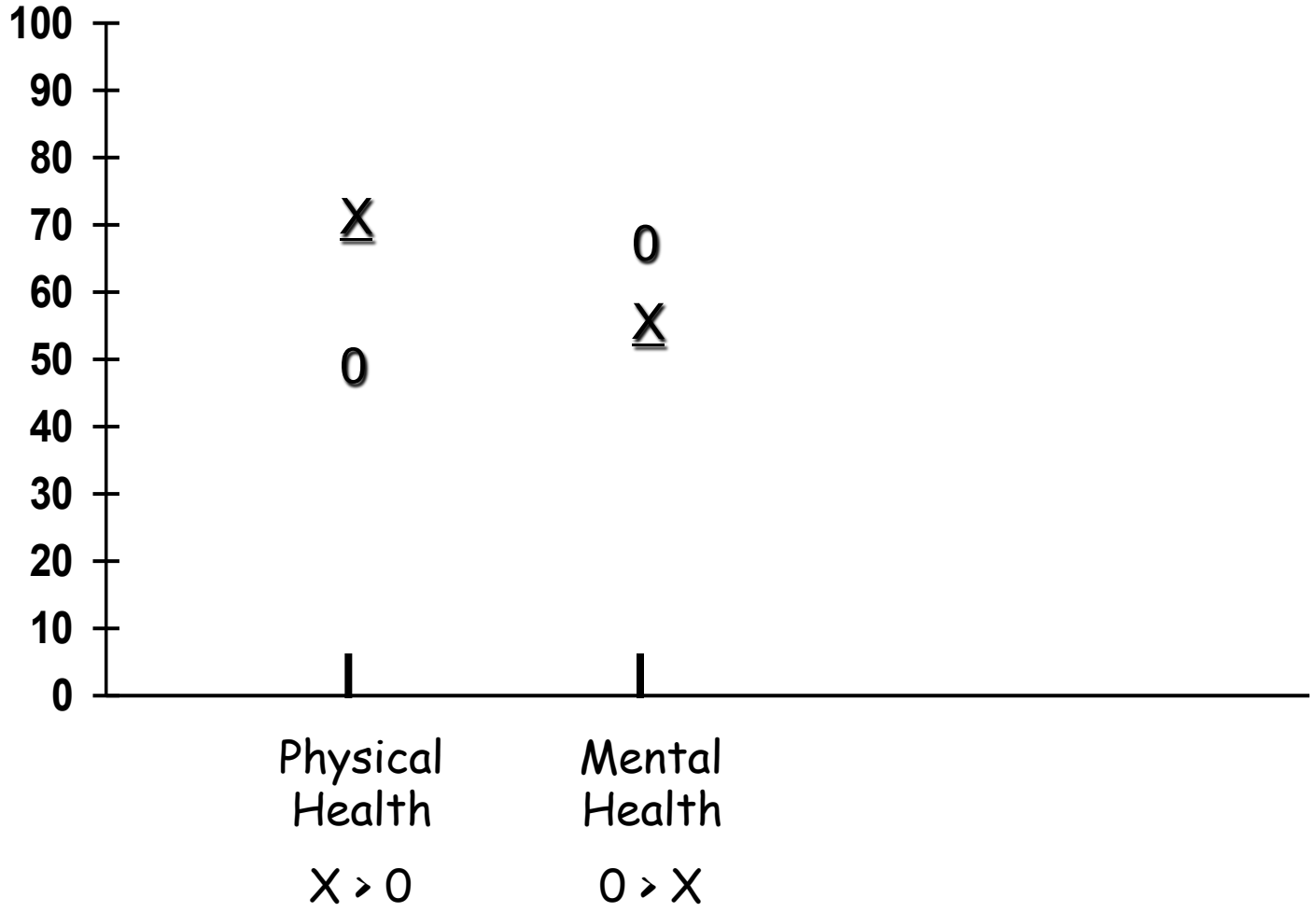
Cost ↓

———————

Effectiveness ↑

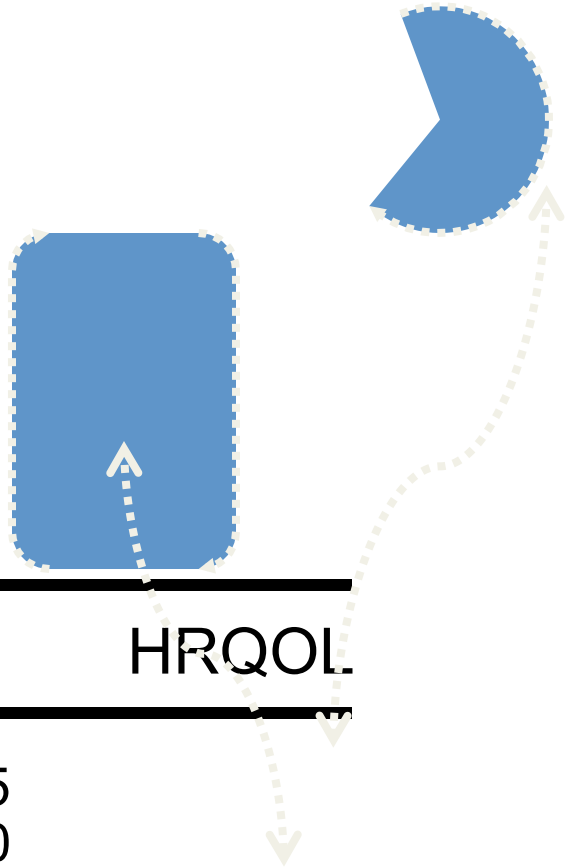
14

Is New Treatment (X) Better Than Standard Care (O)?



Is Medicine Related to Worse HRQOL?

Person	Medication Use	HRQOL (0-100)
	1	Nodead
	2	Nodead
	3 No	50
	4 No	75
	5 No	100
	6 Yes	0
	7 Yes	25
	8 Yes	50
	9 Yes	75
	10 Yes	100

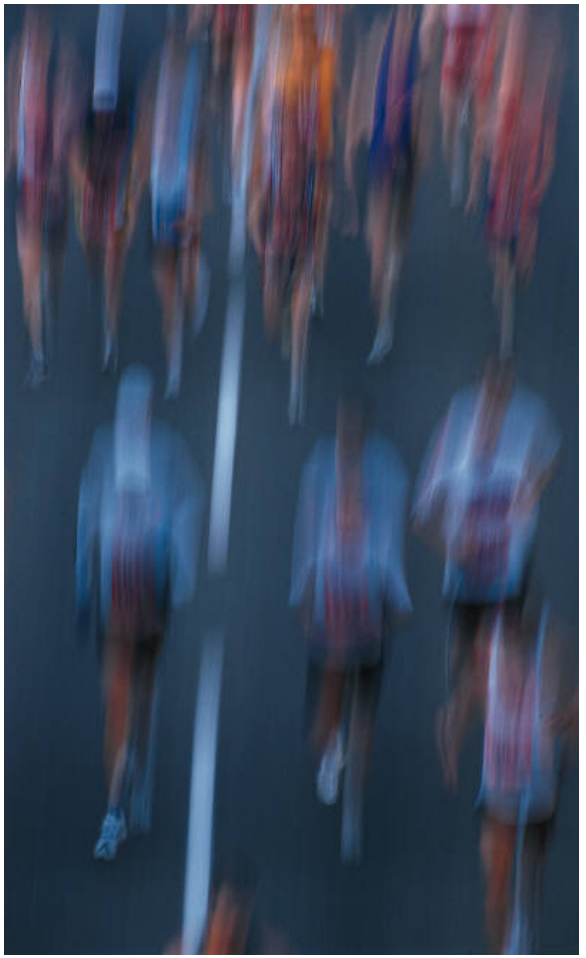


Group	n	HRQOL
No Medicine	375	
Yes Medicine	550	

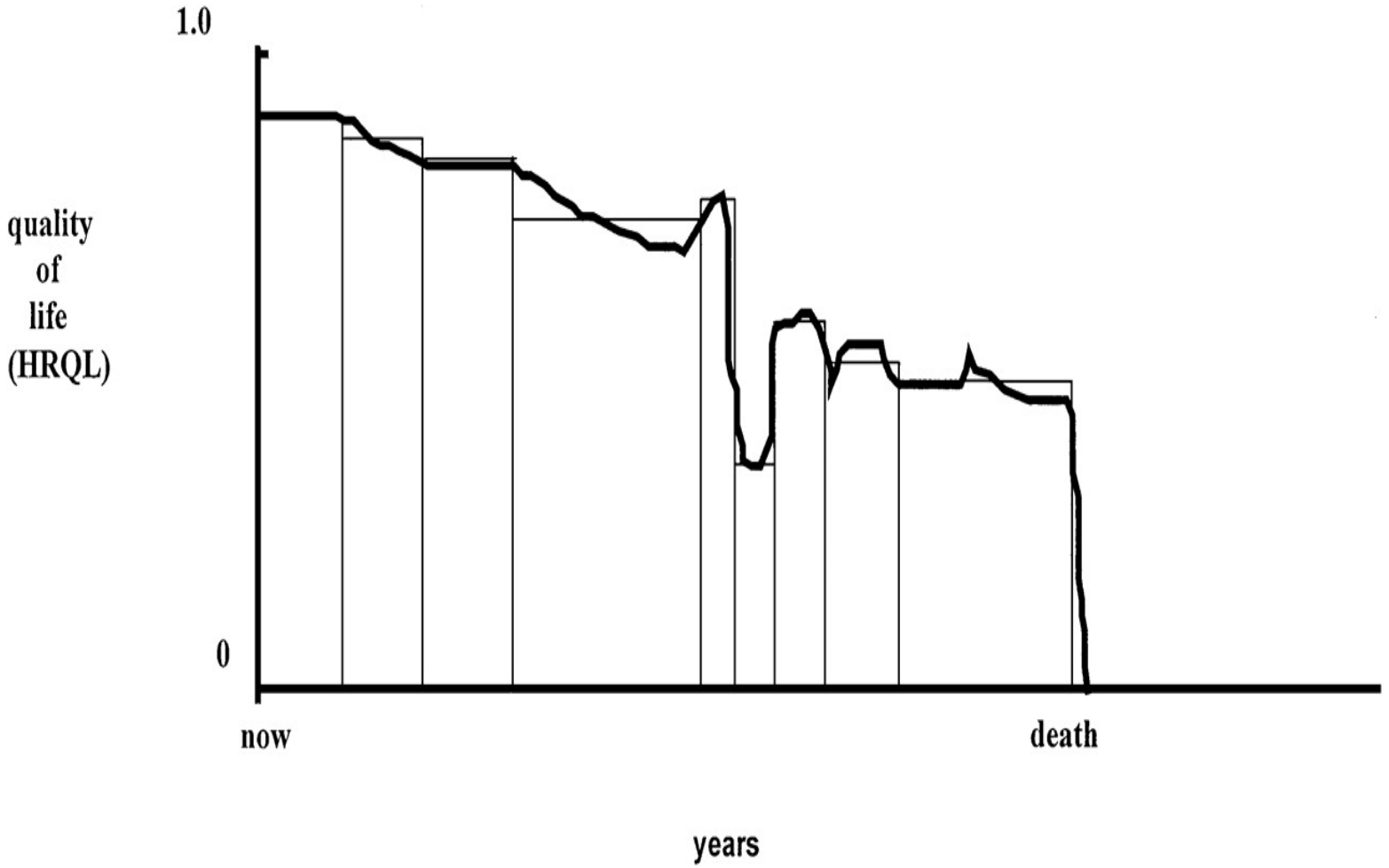
Survival Analysis

Dead	0.0
Alive	1.0

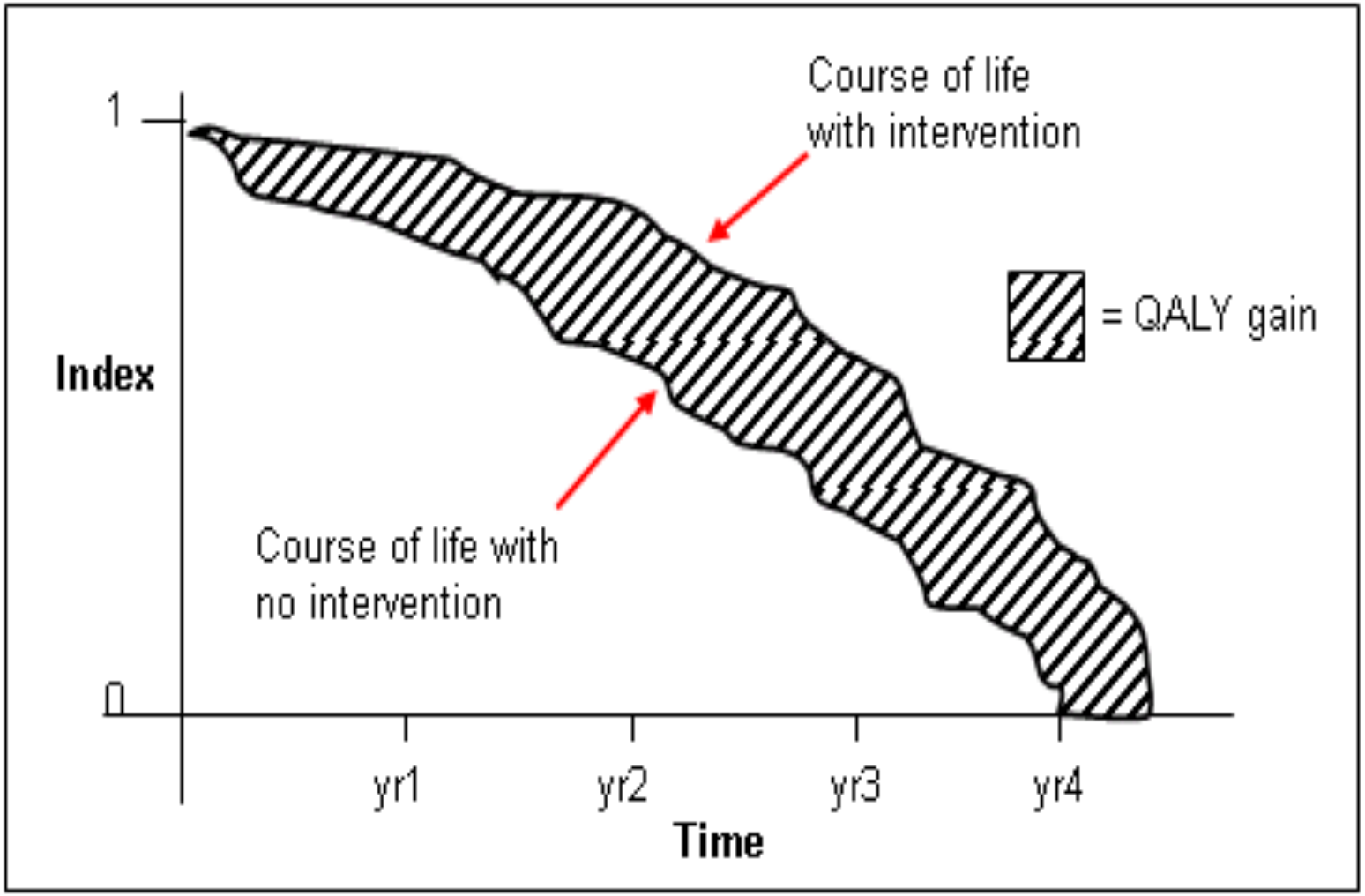
- Marathoner
- Person in coma



Quality of Life for Individual Over Time



http://www.ukmi.nhs.uk/Research/pharma_res.asp



Direct Preference Measures

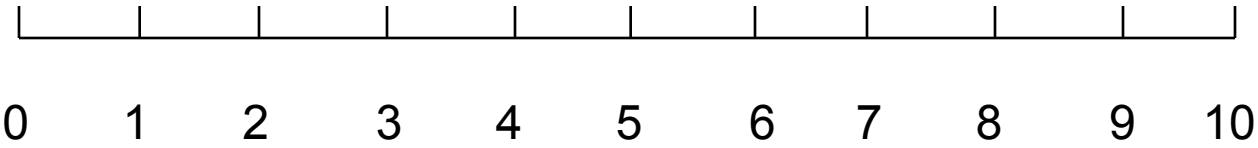
- Underlying attributes unknown
  - Rating Scale
  - Standard gamble
  - Time tradeoff

20

Rating Scale

Overall, how would you rate your current health?

(Circle One Number)



Worst possible health (as bad or worse than being dead)

Half-way between worst and best

Best possible health

Preference Assessment

- http://araw.mede.uic.edu/cgi-bin/utility.cgi

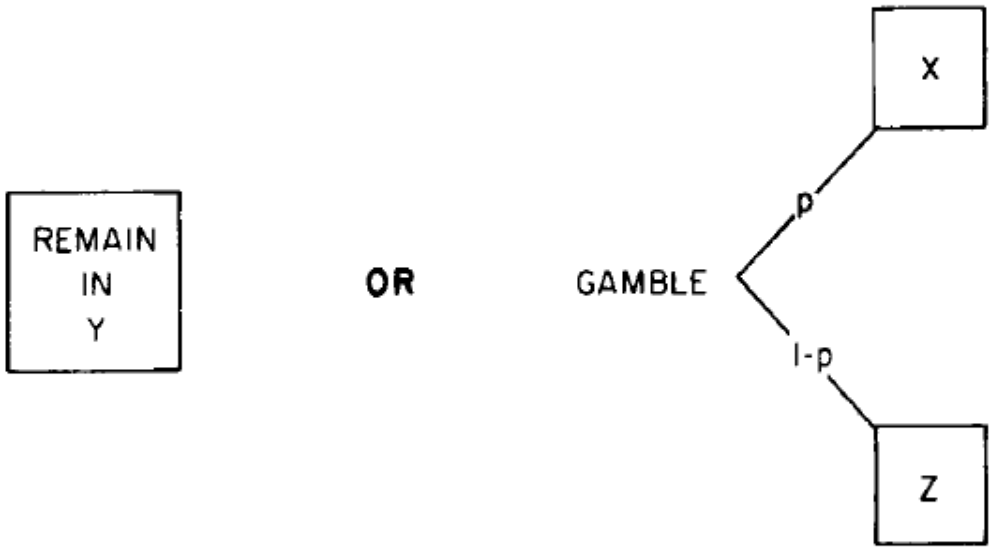
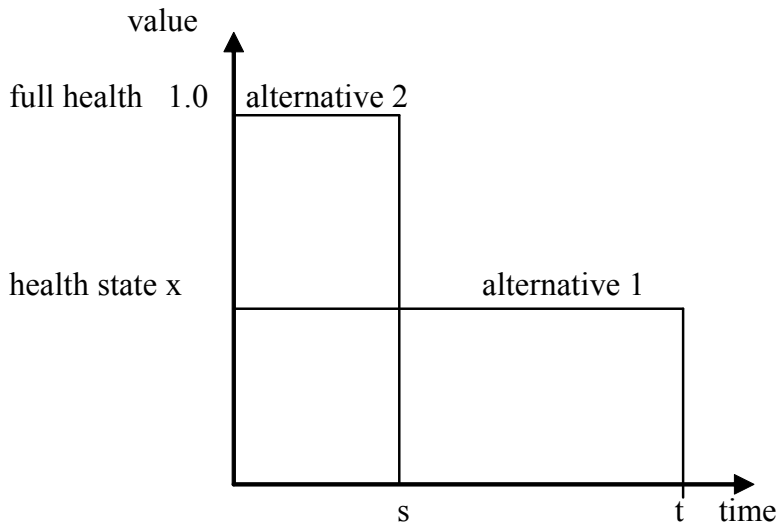


Figure 1. The choice in the standard gamble.

Alternative 1: Certainty of living in given health state y

Alternative 2: Probability of living in full health (x) or immediate death (z)

Time Trade-off approach:



Alternative 1: intermediate health state x, for time t, followed by death.
Alternative 2: full health for time s where s<t, followed by death.

Time t is given and the individual is asked to state s. The preference score is then worked out as s/t.

Indirect Preference Measures

- Estimate single score based on knowing health state for person and societal preferences for that health state
  - Quality of Well-Being (QWB) Scale
  - EQ-5D
  - HUI2 and HUI3
  - SF-6D

25

Health State 111111

Health state 111111

Your health does not limit you in **vigorous activities** (e.g. running, lifting heavy objects, participating in strenuous sports).

You have no problems with your work or other regular daily activities as a result of your **physical health or any emotional problems.**

Your health limits your **social activities** (like visiting friends or close relatives) a little or none of the time

You have no **pain**

You feel **tense or downhearted and low** a little or none of the time.

You have a lot of **energy** all of the time

Health state 424421 (0.59)

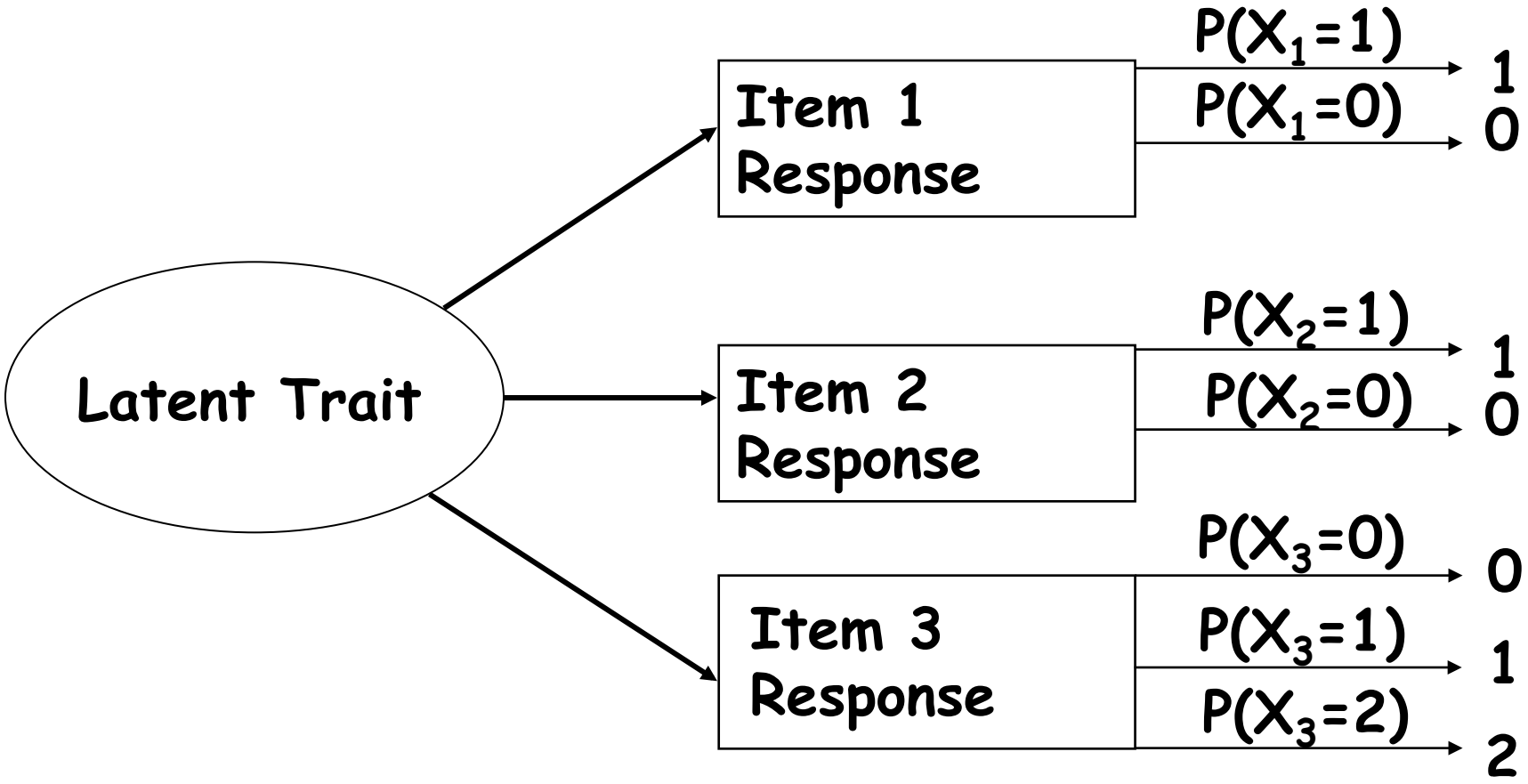
- Your health limits you a lot in moderate activities (such as moving a table, pushing a vacuum cleaner, bowling or playing golf)
- You are limited in the kind of work or other activities as a result of your physical health
- Your health limits your social activities (like visiting friends, relatives etc.) most of the time.
- You have pain that interferes with your normal work (both outside the home and housework) moderately
- You feel tense or downhearted and low a little of the time.
- You have a lot of energy all of the time

Correlations Among Indirect Preference-Based Measures

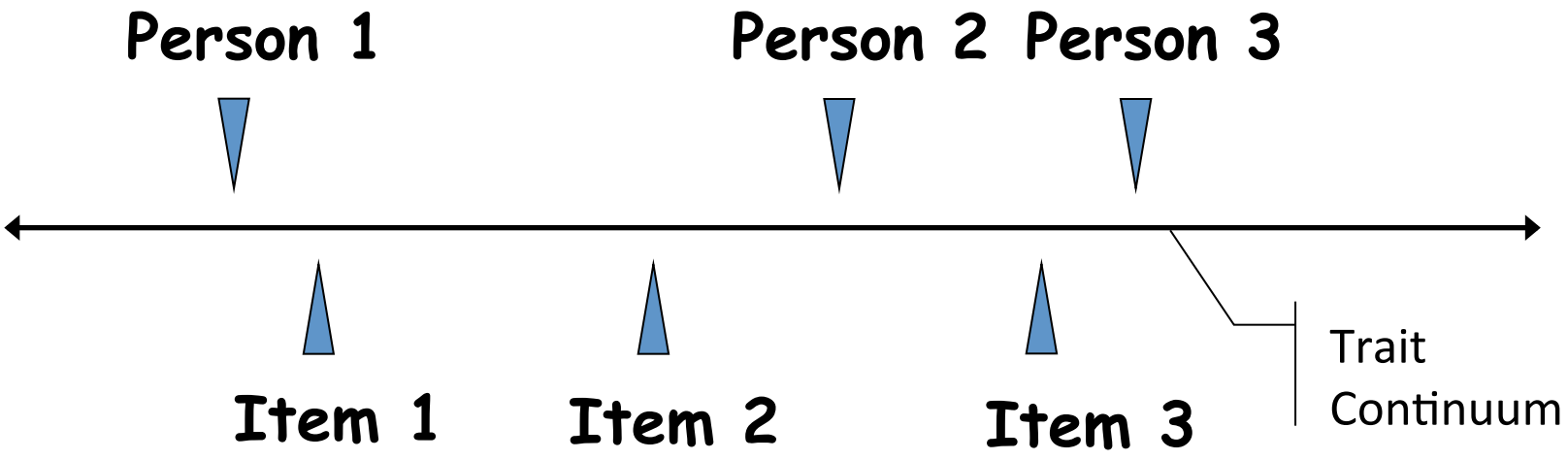
	EQ-5D	HUI2	HUI3	QWB-SA	SF-6D
EQ-5D	1.00				
HUI2	0.71	1.00			
HUI3	0.68	0.89	1.00		
QWB	0.64	0.66	0.66	1.00	
SF-6D	0.70	0.71	0.69	0.65	1.00

Fryback, D. G. et al., (2007). US Norms for Six Generic Health-Related Quality-of-Life Indexes from the National Health Measurement Study. Medical Care, 45, 1162- 1170.

Latent Trait and Item Responses



Item Responses and Trait Levels



PROMIS®
http://www.nihpromis.org/

- Patient-reported Outcomes Measurement Information System
- Item banks measuring patient-reported outcomes
    - Computer-adaptive testing (CAT)
    - Short-forms
- Reliability = $1 - SE^2$ (z-score)
    - SE = 0.32 for 0.90 reliability (z-score)
    - SE = 3.2 for T-score (mean= 50 & SD = 10)

Anger CAT[1] (In the past 7 days)

I was grouchy

- Never
- Rarely
- Sometimes
- Often
- Always

- Theta = 56.1 SE = 5.7 (rel. = 0.68)

[2]In the past 7 days …

I felt like I was ready to explode

- Never
- Rarely
- Sometimes
- Often
- Always

- Theta = 51.9 SE = 4.8 (rel. = 0.77)

[3]In the past 7 days …

I felt angry

- Never
- Rarely
- Sometimes
- Often
- Always

- Theta = 50.5 SE = 3.9 (rel. = 0.85)

[4]In the past 7 days …

I felt angrier than I thought I should

- Never
- Rarely
- Sometimes
- Often
- Always

- Theta = 48.8 SE = 3.6 (rel. = 0.87)

[5]In the past 7 days …

I felt annoyed

- Never
- Rarely
- Sometimes
- Often
- Always

- Theta = 50.1 SE = 3.2 (rel. = 0.90)

[6]In the past 7 days …

I made myself angry about something just by thinking about it.

- Never
- Rarely
- Sometimes
- Often
- Always

- Theta = 50.2 SE = 2.8 (rel. = 0.92)

Theta and SE estimates

- Item 1: 56 and 6
- Item 2: 52 and 5
- Item 3: 50 and 4
- Item 4: 49 and 4
- Item 5: 50 and 3
- Item 6: 50 and <3

PROMIS Banks

- Emotional Distress
  - Depression (28)
  - Anxiety (29)
  - Anger (29)
- Physical Function (124)
- Pain
  - Behavior (39)
  - Impact (41)
- Fatigue (95)
- Satisfaction with Participation in Discretionary Social Activities (12)
- Satisfaction with Participation in Social Roles (14)
- Sleep Disturbance (27)
- Wake Disturbance (16)

Item Response Theory (IRT)

IRT models the relationship between a person's response Y_i to the question (i) and his or her level of the latent construct θ being measured by positing

$$\Pr(Y_i \geq k) = \frac{1}{1 + \exp(-a_i\theta + b_{ik})}$$

b_{ik} estimates how difficult it is for the item (i) to have a score of k or more and the discrimination parameter a_i estimates the discriminatory power of the item.

IRT Features

- Information/reliability
- Category response curves
- Differential item functioning
- Person fit

Information/Reliability

- For z-scores (mean = 0 and SD = 1):
  - Reliability = $1 - SE^2$ = **0.90** (when SE = 0.32)
  - Information = $1/SE^2$ = **10** (when SE = 0.32)
  - Reliability = 1 – 1/information

- Lowering the SE requires adding or replacing existing items with more informative items at the target range of the continuum.

Posttraumatic Growth Inventory

Indicate for each of the statements below the degree to which this change occurred in your life as a result of your crisis.

(*Appreciating each day*)

(0) I did not experience this change as result of my crisis

(1) I experienced this change to a very small degree as a result of my crisis

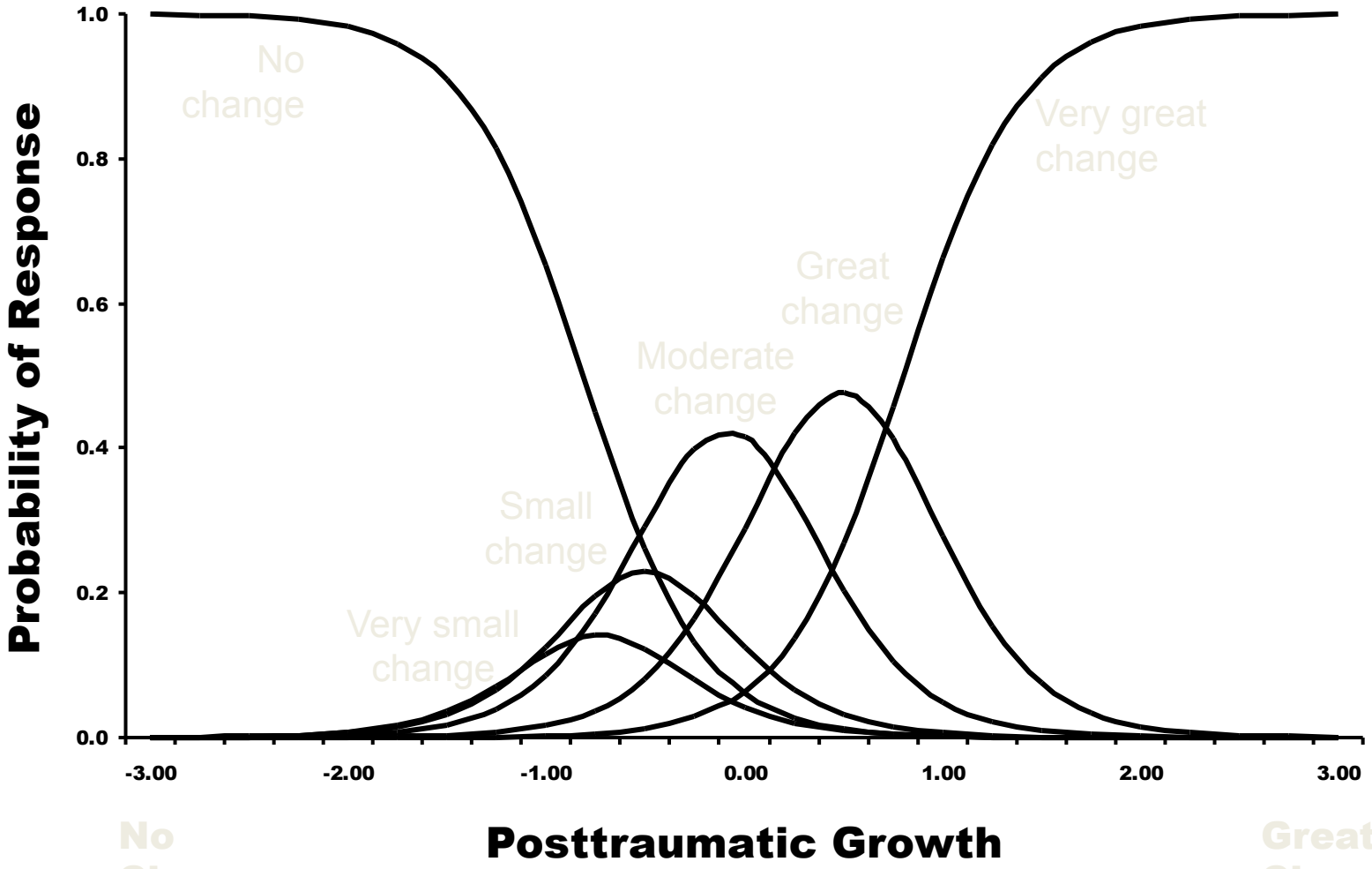
(2) I experienced this change to a small degree as a result of my crisis

(3) I experienced this change to a moderate degree as a result of my crisis

(4) I experienced this change to a great degree as a result of my crisis

(5) I experienced this change to a very great degree as a result of my crisis

Category Response Curves



Category Response Curves (CRCs)

- Figure shows that 2 of 6 response options are never most likely to be chosen
  - did not, **very small**, **small**, moderate, great, very great degree

- One or both of the response categories could be dropped or reworded to improve the response scale

Or have scoring implications

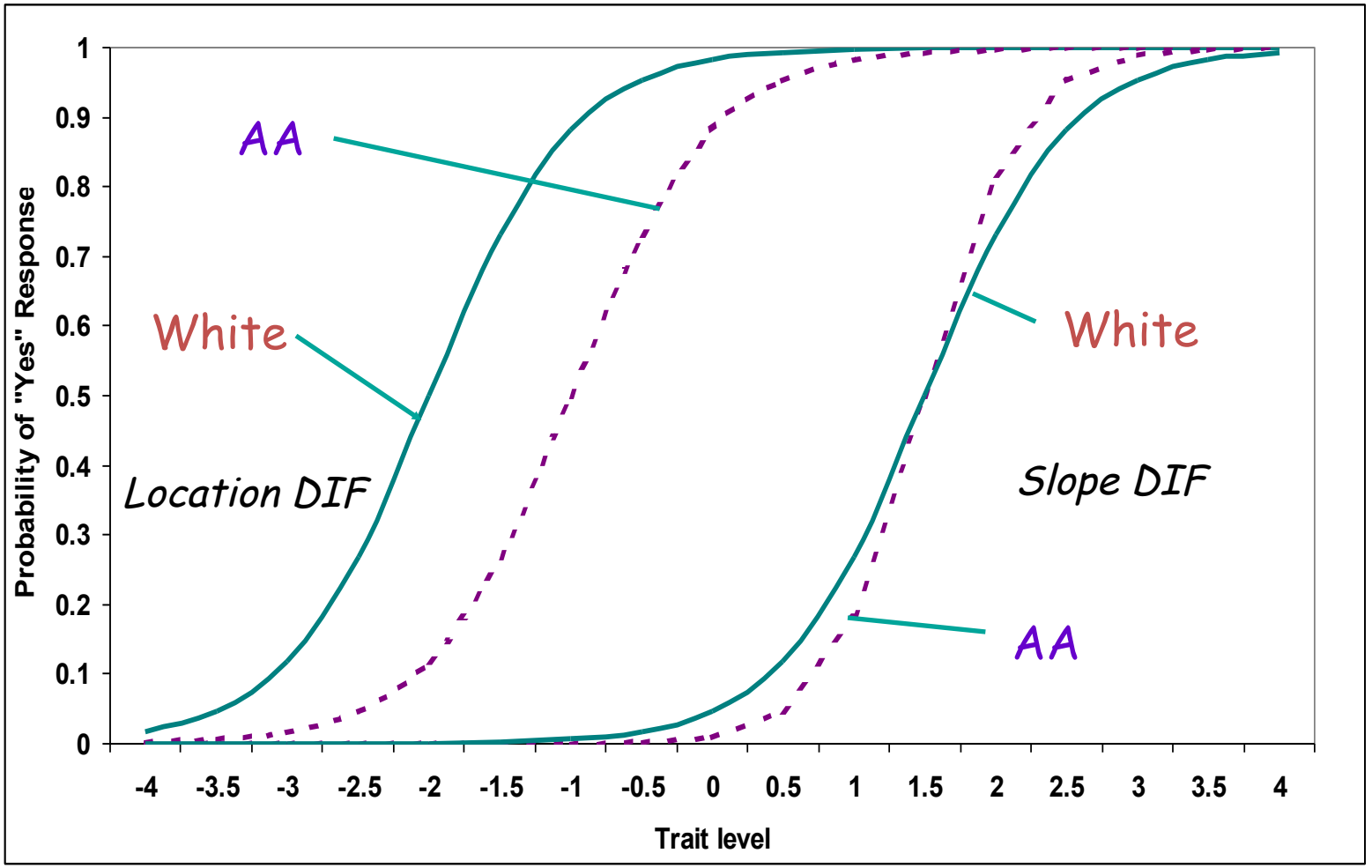
- CAHPS global rating items
  - 0 = worst possible
  - 10 = best possible
- 11 response categories capture about 3 levels of information.
  - 10/9/8-0 or 10-9/8/7-0
- Scale is administered as is and then collapsed in analysis

Differential Item Functioning (DIF)

- Probability of choosing each response category should be the same for those who have the same estimated scale score, regardless of their other characteristics

- Evaluation of DIF
  - Different subgroups
  - Mode differences

Differential Item Functioning (2-Parameter Model)



Location = uniform; Slope = non-uniform

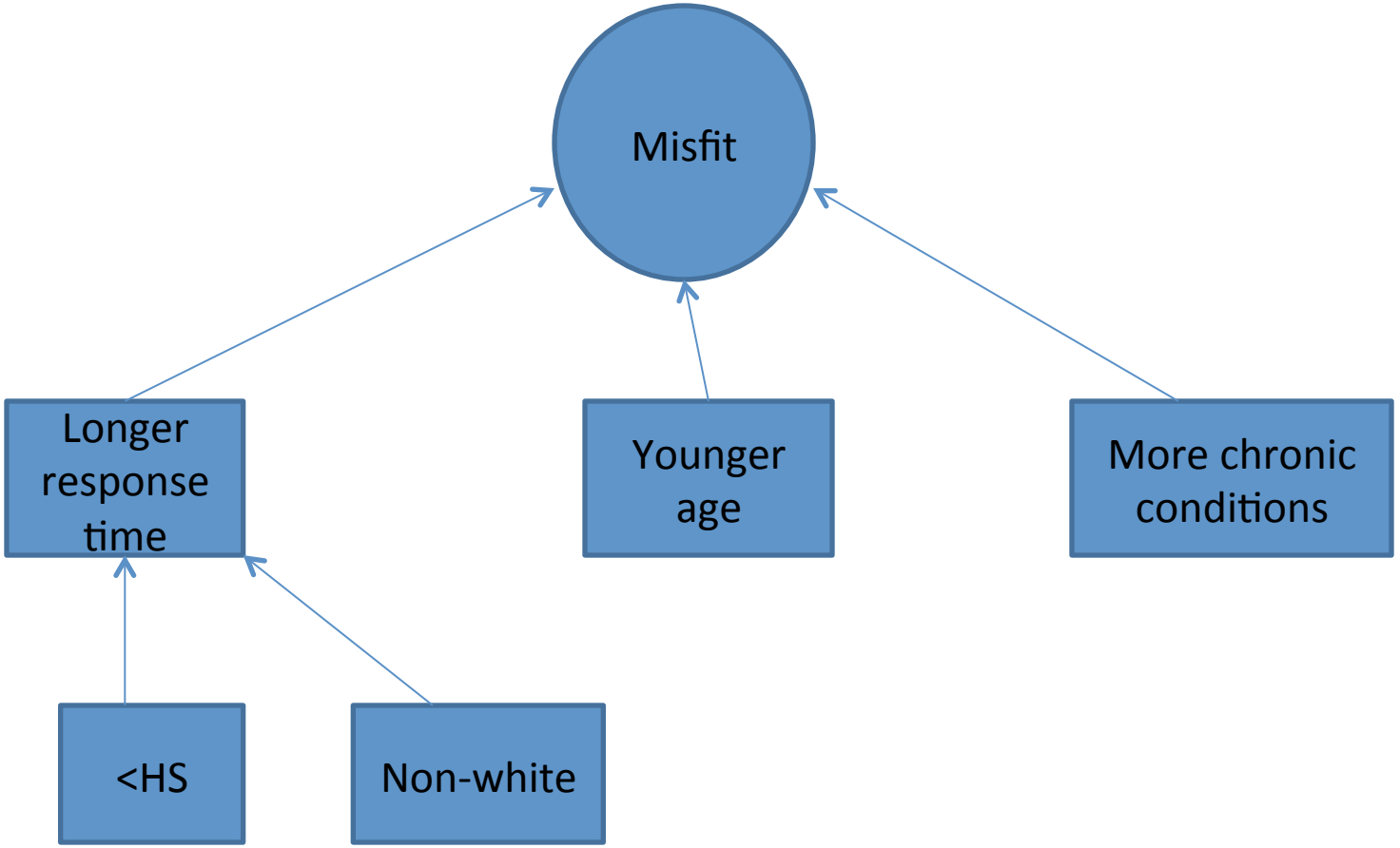
48

Person Fit

- Large negative Z_L values indicate misfit.

- Person responded to 14 items in physical functioning bank (Z_L = -3.13)
  - For 13 items the person could do the activity (including running 5 miles) without any difficulty.
  - However, this person reported *a little difficulty* being out of bed for most of the day.

Unique Associations with Person Misfit



Time to complete item

- 3-5 items per minute rule of thumb
  - 8 items per minute for dichotomous items
- Polimetrix panel sample
  - 12-13 items per minute (automatic advance)
  - 8-9 items per minute (next button)
- 6 items per minute among UCLA Scleroderma patients

Language DIF Example

- Ordinal logistic regression to evaluate differential item functioning
  - Purified IRT trait score as matching criterion
  - McFadden's pseudo R^2 >= 0.02
- Thetas estimated in Spanish data using
  - English calibrations
  - Linearly transformed Spanish calibrations (Stocking-Lord method of equating)

52

Lordif

http://CRAN.R-project.org/package=lordif

Model 1 : logit $P(u_i >= k) = \alpha_k + \mathbf{\beta_1 * ability}$

Model 2 : logit $P(u_i >= k) = \alpha_k + \beta_1 *$ ability $+ \mathbf{\beta_2 * group}$

Model 3 : logit $P(u_i >= k) = \alpha_k + \beta_1 *$ ability $+ \beta_2 *$ group $+ \mathbf{\beta_3 *}$ **ability * group**

DIFF assessment (log likelihood values compared):

- Overall: Model 3 versus Model 1
- Non-uniform: Model 3 versus Model 2
- Uniform: Model 2 versus Model 1

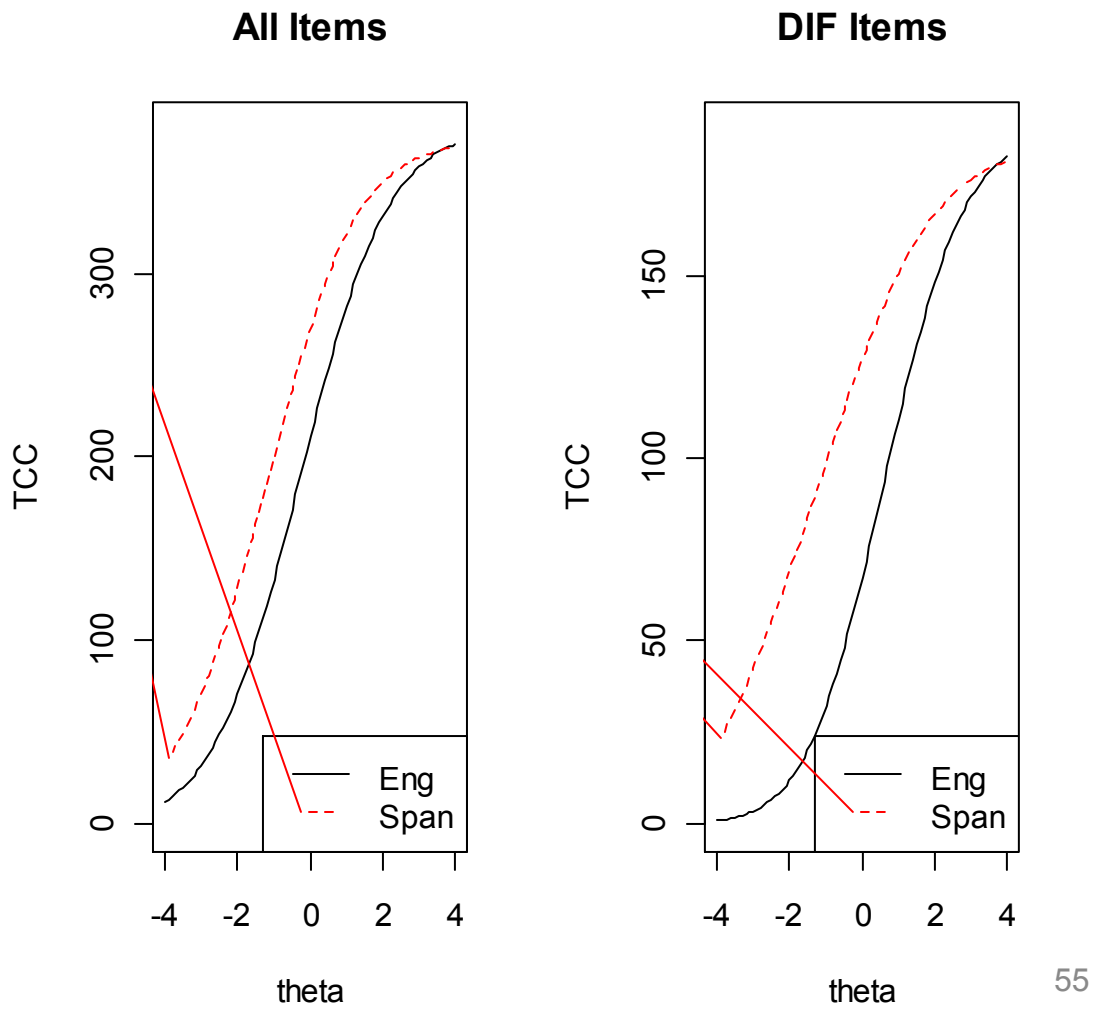
53

Results

- One-factor categorical model fit the data well (CFI=0.971, TLI=0.970, and RMSEA=0.052).
  - Large residual correlation of 0.67 between “Are you able to run ten miles” and “Are you able to run five miles?”
- 50 of the 114 items had language DIF
  - 16 uniform
  - 34 non-uniform

54

Impact of DIF on Test Characteristic Curves (TCCs)



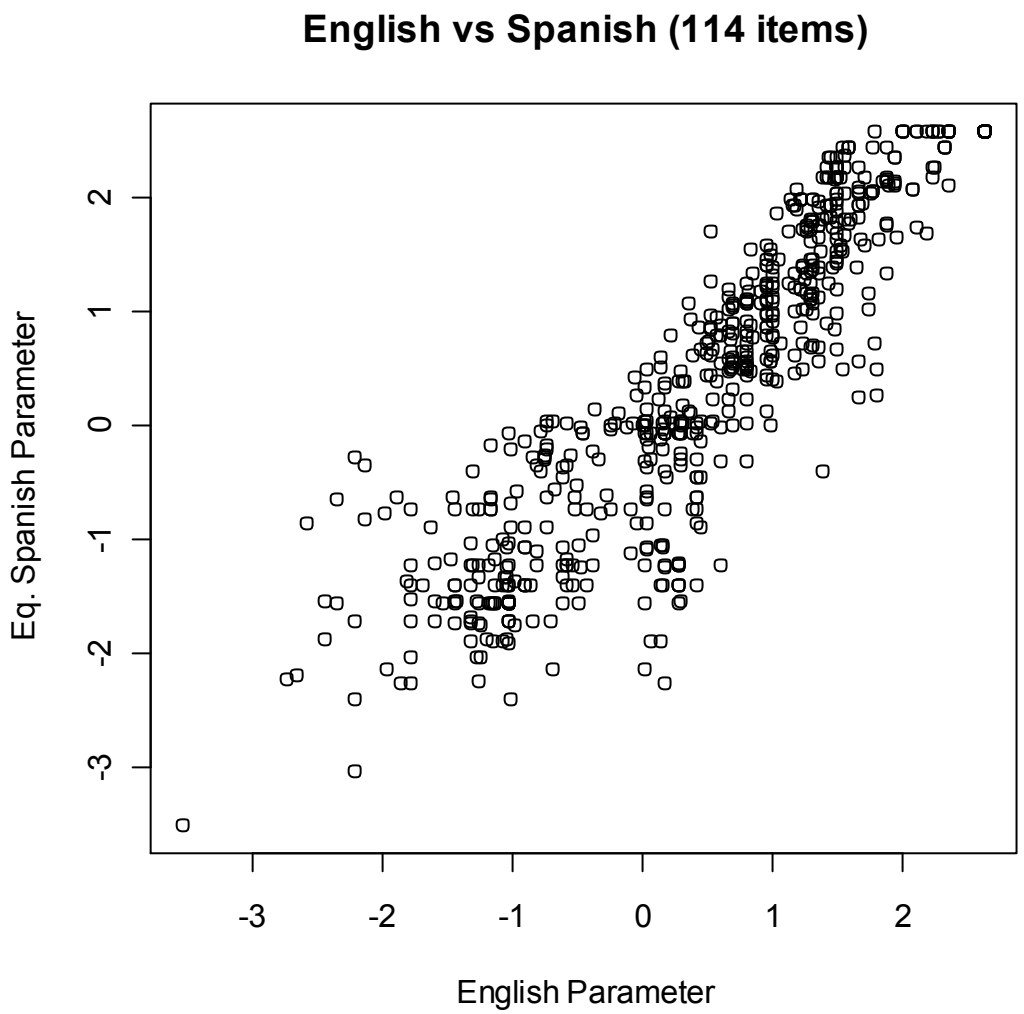
55

Stocking-Lord Method

- Spanish calibrations transformed so that their TCC most closely matches English TCC.
- $a^* = a/A$ and $b^* = A * b + B$
- Optimal values of A (slope) and B (intercept) transformation constants found through multivariate search to minimize weighted sum of squared distances between TCCs of English and Spanish transformed parameters
  - Stocking, M.L., & Lord, F.M. (1983). Developing a common metric in item response theory. *Applied Psychological Measurement, 7,* 201-210.

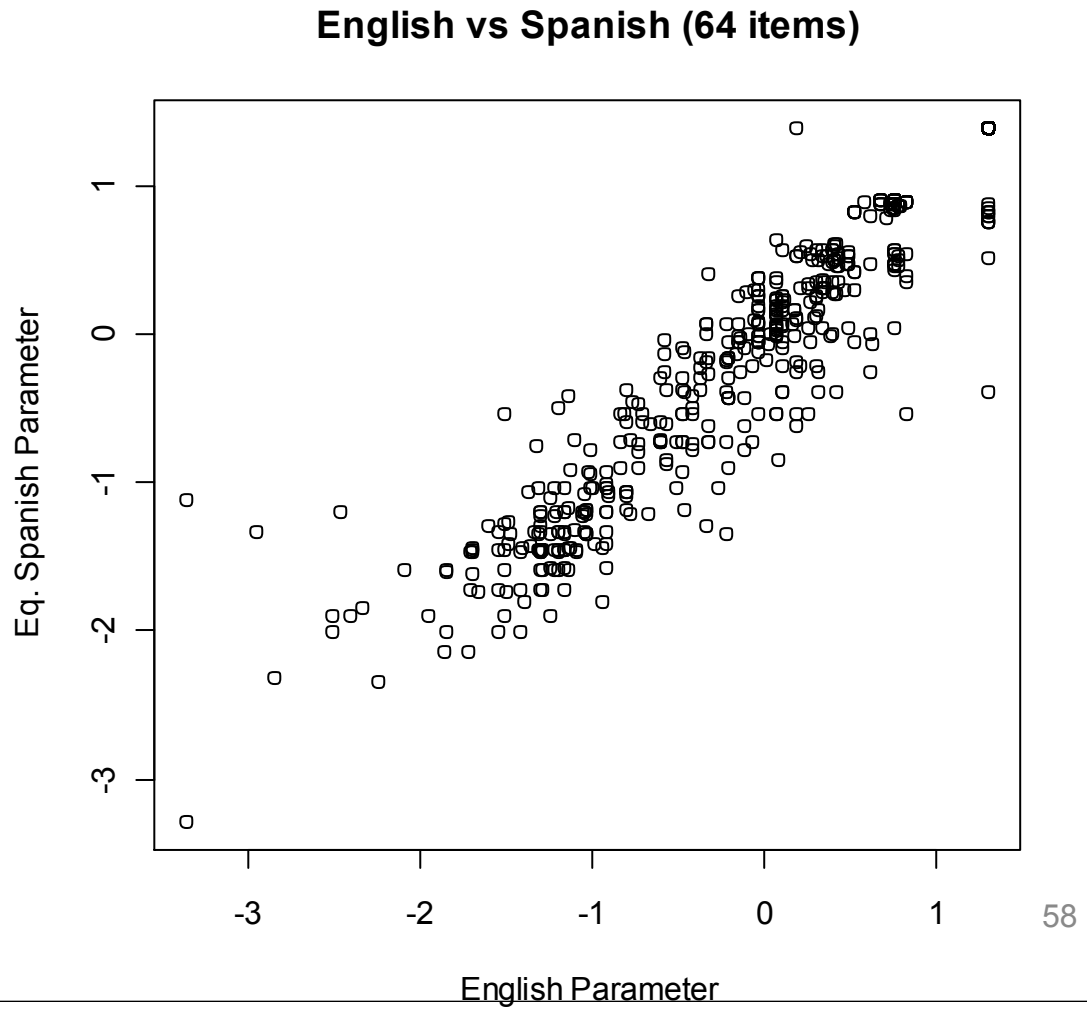
56

CAT-based Theta Estimates Using English (x-axis) and Spanish (y-axis) Parameters for 114 Items in Spanish Sample (n = 640, ICC = 0.89)



57

CAT-based Theta Estimates Using English (x-axis) and Spanish (y-axis) Parameters for 64 non-DIF Items in Spanish Sample (n = 640, ICC = 0.96)



58

Questions?