Evaluating Multi-Item Scales

Health Services Research Design (HS 225A)

November 15, 2013, 10:00-11:45 am

51-279 (Public Health)

Listed below are a few statements about your relationships with others. How much is each statement TRUE or FALSE for you

	Definitely True	Mostly True	Don't Know	Mostly False	Definitely False
1. I am always courteous even to people who are disagreeable.	1	2	3	4	5
2. There have been occasions when I took advantage of someone.	1	2	3	4	5
3. I sometimes try to get even rather than forgive and forget.	1	2	3	4	5
4. I sometimes feel resentful when I don't get my way.	1	2	3	4	5
5. No matter who I'm talking to, I'm always a good listener.	1	2	3	4	5

Scoring Multi-Item 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)

Linear Transformations

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

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

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

Listed below are a few statements about your relationships with others. How much is each statement TRUE or FALSE for you

	Definitely True	Mostly True	Don't Know	Mostly False	Definitely False
1. I am always courteous even to people who are disagreeable.	100	75	50	25	0
2. There have been occasions when I took advantage of someone.	0	25	50	75	100
3. I sometimes try to get even rather than forgive and forget.	0	25	50	75	100
4. I sometimes feel resentful when I don't get my way.	0	25	50	75	100
5. No matter who I'm talking to, I'm always a good listener.	100	75	50	25	0

Create T-score

z-score = (score – 36)/31

T-score = (10 * z-score) + 50

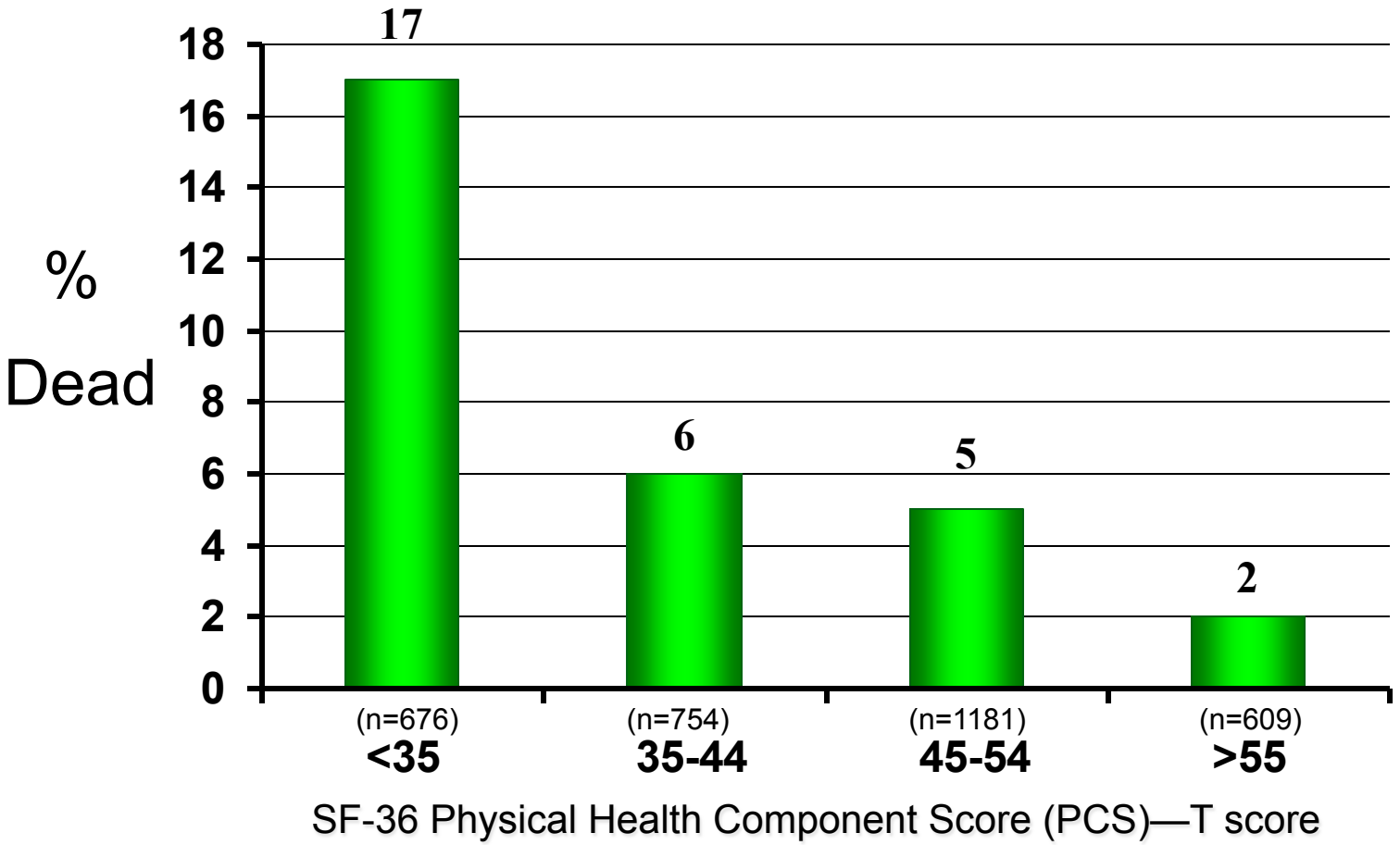
z-score = (100- 36)/31 = 2.06

T-score = 71

Validity

- Content validity
  - Patients and/or experts judge the items to be representing the intended concept adequately
- Construct validity
  - Extent to which associations with other variables are consistent with prior hypotheses

Self-Reports of Physical Health Predict Five-Year Mortality



Ware et al. (1994). SF-36 Physical and Mental Health Summary Scales: A User's Manual.

Evaluating Construct Validity

Scale	Age	Obesity	ESRD	Nursing Home Resident
Physical Functioning	Medium (-).	Small (-)	Large (-)	Large (-)
Depressive Symptoms	?	Small (+)	?	Small (+)

Cohen effect size rules of thumb (d = 0.2, 0.5, and 0.8):
Small correlation = 0.100
Medium correlation = 0.243
Large correlation = 0.371
$r = d / [(d^2 + 4)^{.5}] = 0.8 / [(0.8^2 + 4)^{.5}] = 0.8 / [(0.64 + 4)^{.5}] = 0.8 / [(4.64)^{.5}] = 0.8 / 2.154 = 0.371$

Beware: r's of 0.10, 0.30 and 0.50 are often cited as small, medium, and large.

Responsiveness to Change

- Valid measures should be responsive to interventions that change the thing being measured.

- Compare change on measure to change indicated on external indicator (anchor)

Self-Reported Change Anchor

We would like to know about any changes in how you are feeling now compared with how you were feeling 6 months ago. Has your ability to carry out your everyday physical activities such as walking, climbing stairs, carrying groceries, or moving a chair …

got a lot better?

got a little better?

stayed the same?

got a little worse?

got a lot worse?'

Change on PROMIS® Physical Functioning Scale (T-score) by Change on Anchor

	Lot Better	Little Better	Same	Little Worse	Lot Worse
	(n = 21)	(n = 35)	(n = 252)	(n = 113)	(n = 30)
Wave 3 – Wave 1	1.94[a]	1.63[a,b]	0.27[b]	-1.68[c]	-3.20[d]
Wave 3 – Wave 2	3.26[a]	1.96[a,b]	0.43[b,c]	-0.82[c]	-3.16[d]

Wave 3 is 12 months after wave 1. Wave 2 is 6 months after wave 1.

Reliability

- Extent to which measure yields similar result when the thing being measured hasn't changed

- Ranges from 0-1
  - Standard is 0.70 or above for research or other group comparisons

Two Raters' Ratings of GOP Debate Performance on *Excellent* to *Poor* Scale

[1 = Poor; 2 = Fair; 3 = Good; 4 = Very good; 5 = Excellent]

1=Bachman Turner Overdrive (Good, Very Good)
2=Ging Rich (Very Good, Excellent)
3=Rue Paul (Good, Good)
4=Gaylord Perry (Fair, Poor)
5=Romulus Aurelius (Excellent, Very Good)
6=Sanatorium (Fair, Fair)

(*Target = 6 candidates; assessed by 2 raters*)

Cross-Tab of Ratings

Rater 2 \ Rater 1	P	F	G	VG	E	Total
P	0	1				1
F		1				1
G			1			1
VG			1	0	1	2
E				1	0	1
Total	0	2	2	1	1	6

Calculating KAPPA

$$\mathbf{P_C} = \frac{(0 \times 1) + (2 \times 1) + (2 \times 1) + (1 \times 2) + (1 \times 1)}{(6 \times 6)} = 0.19$$

$$\mathbf{P_{obs.}} = \frac{2}{6} = 0.33$$

$$\mathbf{Kappa} = \frac{0.33 - 0.19}{1 - 0.19} = \mathbf{0.17}$$

Weighted Kappa Linear (Quadratic)

	P	F	G	VG	E
P	1	.75 (.937)	.50 (.750)	.25 (.437)	0
F	.75 (.937)	1	.75 (.937)	.50 (.750)	.25 (.437)
G	.50 (.750)	.75 (.937)	1	.75 (.937)	.50 (.750)
VG	.25 (.437)	.50 (.750)	.75 (.937)	1	.75 (.937)
E	0	.25 (.437)	.5 (.750)	.75 (.937)	1

$W_l = 1 - (i/ (k - 1))$

$W_q = 1 - (i^2 / (k - 1)^2)$

i = number of categories ratings differ by
k = n of categories

All Kappas

$$\mathbf{P_C} = \frac{(0 \times 1) + (2 \times 1) + (2 \times 1) + (1 \times 2) + (1 \times 1)}{(6 \times 6)} = 0.19$$

$$\mathbf{P_{obs.}} = \frac{2}{6} = 0.33$$

$$\mathbf{Kappa} = \frac{0.33 - 0.19}{1 - 0.19} = \mathbf{0.17}$$

Linear weighted kappa = 0.52
Quadratic weighted kappa = 0.77

Reliability and Intraclass Correlation

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	$\frac{MS_{BMS} - MS_{EMS}}{MS_{BMS}}$	$\frac{MS_{BMS} - MS_{EMS}}{MS_{BMS} + (k-1)MS_{EMS}}$
Two-way random	$\frac{N(MS_{BMS} - MS_{EMS})}{NMS_{BMS} + MS_{JMS} - MS_{EMS}}$	$\frac{MS_{BMS} - MS_{EMS}}{MS_{BMS} + (k-1)MS_{EMS} + k(MS_{JMS} - MS_{EMS})/N}$

BMS = Between Ratee Mean Square
WMS = Within Mean Square
JMS = Item or Rater Mean Square
EMS = Ratee x Item (Rater) Mean Square

N = n of ratees
k = n of items or raters

01 34
02 45
03 33
04 21
05 54
06 22

Reliability of Performance Ratings

Source	df	SS	MS
Candidates (BMS)	5	15.67	3.13
Raters (JMS)	1	0.00	0.00
Cand. x Raters (EMS)	5	2.00	0.40
Total	11	17.67	

$$\text{2-way R} = \frac{6\,(3.13 - 0.40)}{6\,(3.13) + 0.00 - 0.40} = \boxed{0.89}$$

ICC = 0.80

GOP Presidential Candidates Responses to Two Questions about Their Health

1. Bachman Turner Overdrive (Good, Very Good)
2. Ging Rich (Very Good, Excellent)
3. Rue Paul (Good, Good)
4. Gaylord Perry (Fair, Poor)
5. Romulus Aurelius (Excellent, Very Good)
6. Sanatorium (Fair, Fair)

(*Target = 6 candidates; assessed by 2 items*)

01 34
02 45
03 33
04 21
05 54
06 22

Two-Way Fixed Effects (Cronbach's Alpha)

Source	df	SS	MS
Respondents (BMS)	5	15.67	3.13
Items (JMS)	1	0.00	0.00
Resp. x Items (EMS)	5	2.00	0.40
Total	11	17.67	

$$\text{Alpha} = \frac{3.13 - 0.40}{3.13} = \frac{2.93}{3.13} = \boxed{0.87}$$

ICC = 0.77

Overall Satisfaction of 12 Patients with 6 Doctors (2 patients per doctor)

1. Dr. Overdrive (p1: Good, p2: Very Good)
2. Dr. Rich (p3: Very Good, p4: Excellent)
3. Dr. Paul (p5: Good, p6: Good)
4. Dr. Perry (p7: Fair, p8: Poor)
5. Dr. Aurelius (p9: Excellent, p10: Very Good)
6. Dr. Sanatorium (p11: Fair, p12: Fair)

(*Target = 6 doctors; assessed by 2 patients each*)

01 34
02 45
03 33
04 21
05 54
06 22

Reliability of Ratings of Doctor

Source	df	SS	MS
Respondents (BMS)	5	15.67	3.13
Within (WMS)	6	2.00	0.33
Total	11	17.67	

$$\text{1-way} = \frac{3.13 - 0.33}{3.13} = \frac{2.80}{3.13} = \boxed{0.89}$$

Item-scale correlation matrix

	Depress	Anxiety	Anger
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.

Item-scale correlation matrix

	Depress	Anxiety	Anger
Item #1	0.50*	0.50	0.50
Item #2	0.50*	0.50	0.50
Item #3	0.50*	0.50	0.50
Item #4	0.50	0.50*	0.50
Item #5	0.50	0.50*	0.50
Item #6	0.50	0.50*	0.50
Item #7	0.50	0.50	0.50*
Item #8	0.50	0.50	0.50*
Item #9	0.50	0.50	0.50*

*Item-scale correlation, corrected for overlap.

Confirmatory Factor Analysis

	Depress	Anxiety	Anger
Item #1	0.80*	0.00	0.00
Item #2	0.80*	0.00	0.00
Item #3	0.80*	0.00	0.00
Item #4	0.00	0.80*	0.00
Item #5	0.00	0.80*	0.00
Item #6	0.00	0.80*	0.00
Item #7	0.00	0.00	0.80*
Item #8	0.00	0.00	0.80*
Item #9	0.00	0.00	0.80*

*Factor loading.

Confirmatory Factor Analysis Fit Indices

- Normed fit index: $\frac{\chi^2_{null} - \chi^2_{model}}{\chi^2_{null}}$

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

- Comparative fit index: $1 - \frac{\chi^2_{model} - df_{model}}{\chi^2_{null} - df_{null}}$

RMSEA = SQRT (λ^2 – df)/SQRT (df (N – 1))

Item Response Theory (IRT)

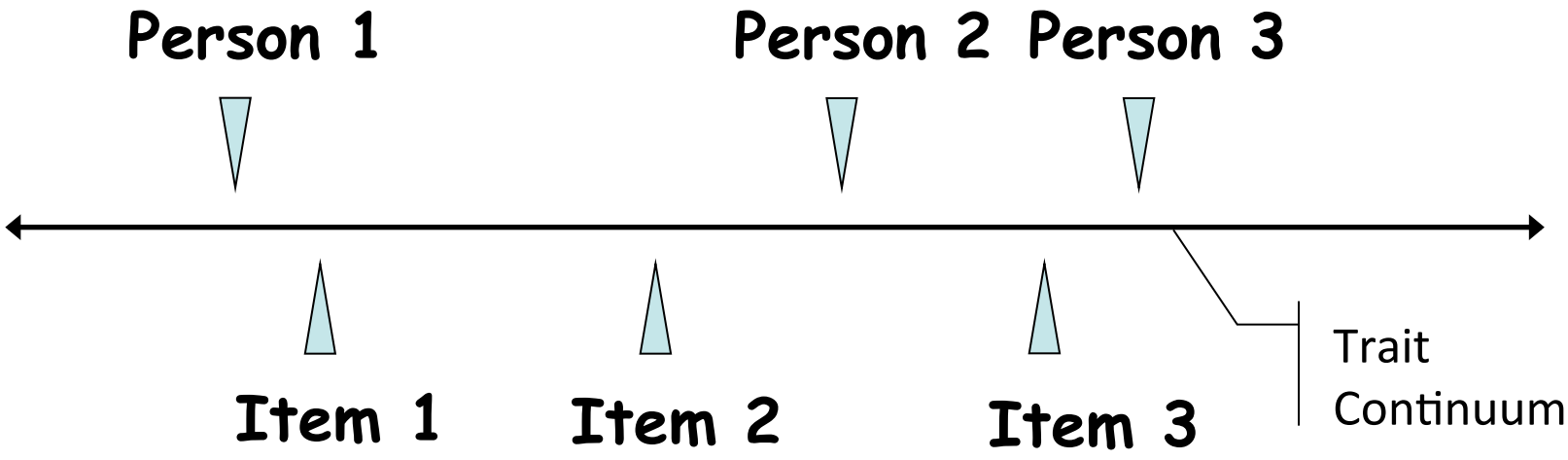
IRT models the relationship between a person's response Y_i to the question (i) and his or her level of the latent construct θ being measured by positing

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

b_{ik} estimates how difficult it is to get a score of k or more on item (i).

a_i is an estimate of the discriminatory power of the item.

Item Responses and Trait Levels



www.nihpromis.org

Computer Adaptive Testing (CAT)

Reliability Target for Use of Measures with Individuals

- Reliability ranges from 0-1
  - 0.90 or above is goal
  - SEM = SD $(1- \text{reliability})^{1/2}$
  - 95% CI = true score +/- 1.96 x SEM
    - if true z-score = 0, then CI: -.62 to +.62
      - Width of CI is 1.24 z-score units

- Reliability = 0.90 when SE = 3.2
  - T-scores (mean = 50, SD = 10)
  - Reliability = $1 - (SE/10)^2$

$$T = 50 + (z * 10)$$

In the past 7 days …

I was grouchy [1st question]

- Never [39]
- Rarely [48]
- Sometimes [56]
- Often [64]
- Always [72]

Estimated Anger = 56.1

SE = 5.7 (rel. = 0.68)

In the past 7 days …

I felt like I was ready to explode

[2nd question]

- Never
- Rarely
- Sometimes
- Often
- Always

Estimated Anger = 51.9

SE = 4.8 (rel. = 0.77)

In the past 7 days …

I felt angry [3rd question]

- Never
- Rarely
- Sometimes
- Often
- Always

Estimated Anger = 50.5

SE = 3.9 (rel. = 0.85)

In the past 7 days …

I felt angrier than I thought I should
[4th question]

- Never
- Rarely
- Sometimes
- Often
- Always

Estimated Anger = 48.8

SE = 3.6 (rel. = 0.87)

In the past 7 days …

I felt annoyed [5th question]

- Never
- Rarely
- Sometimes
- Often
- Always

Estimated Anger = 50.1

SE = 3.2 (rel. = 0.90)

In the past 7 days …

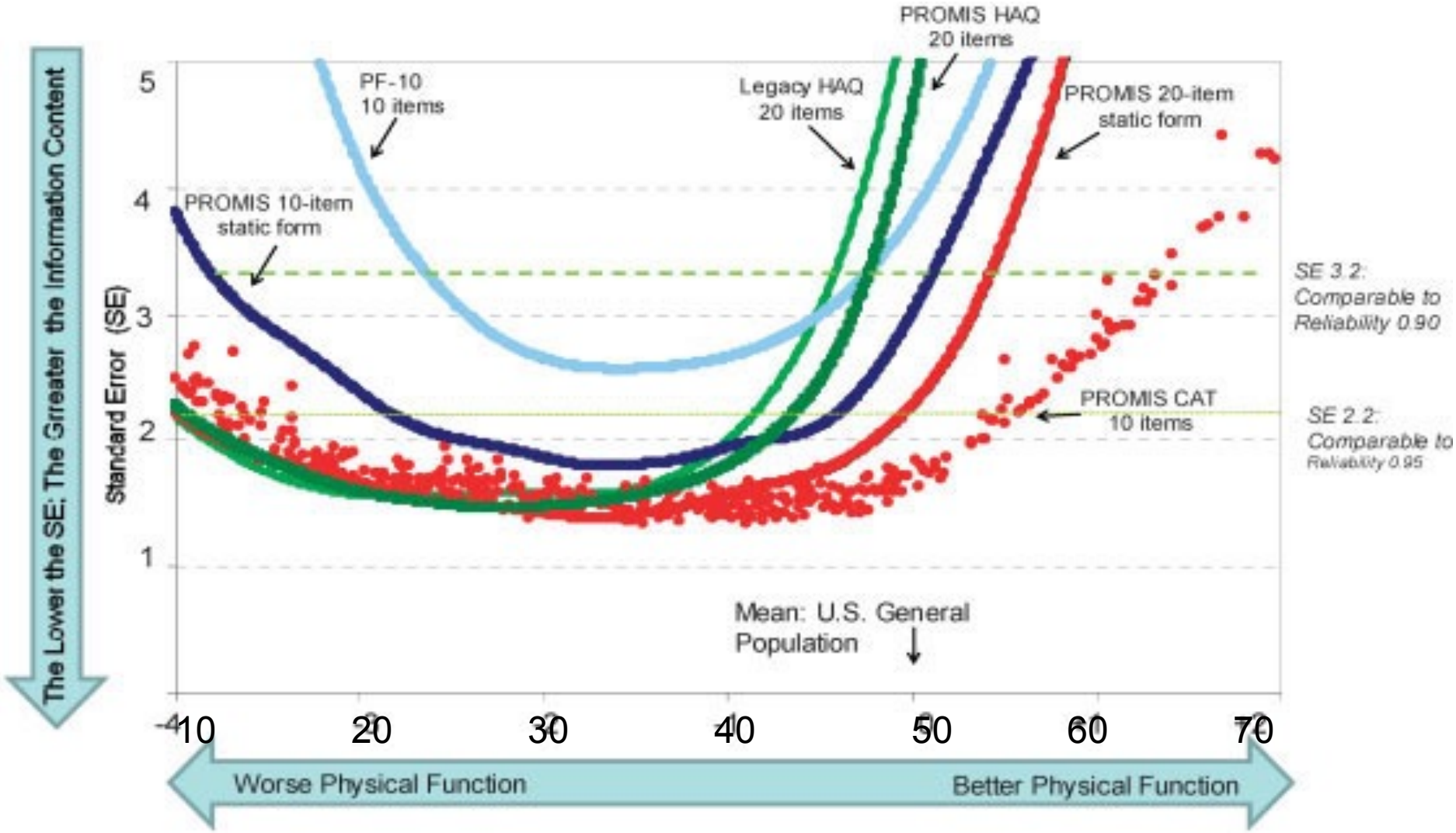
I made myself angry about something just by thinking about it. [6th question]

- Never
- Rarely
- Sometimes
- Often
- Always

Estimated Anger = 50.2

SE = 2.8 (rel = 0.92)

PROMIS Physical Functioning vs. “Legacy” Measures



Defining a Responder: Reliable Change Index (RCI)

$$\frac{X_2 - X_1}{(\sqrt{2})(SE)}$$

RCI >=1.96 is statistically significant individual change..

Thank you.

Powerpoint file is freely available at:
http://gim.med.ucla.edu/FacultyPages/Hays/

Contact information:
drhays@ucla.edu 310-794-2294
For a good time: http://twitter.com/RonDHays

Appendices
ANOVA Computations

- Candidate/Respondents SS

  $(7^2+9^2+6^2+3^2+9^2+4^2)/2 - 38^2/12 = \underline{15.67}$

- Rater/Item SS

  $(19^2+19^2)/6 - 38^2/12 = \underline{0.00}$

- Total SS

  $(3^2+ 4^2+4^2+5^2+3^2+3^2+2^2+1^2+5^2+4^2+2^2+2^2) - 38^2/10 = \underline{17.67}$

- Res. x Item SS= Tot. SS – (Res. SS+Item SS)

```
options ls=130 ps=52 nocenter;
options nofmterr;

data one;
 input id 1-2 rater 4 rating 5;
CARDS;
01 13
01 24
02 14
02 25
03 13
03 23
04 12
04 21
05 15
05 24
06 12
06 22
;
run;
**************;
```

```
proc freq;
tables rater rating;
run;
******************.
                   ;
proc means;
var rater rating;
run;
****************************************.
                                        ;
proc anova;
class id rater;
model rating=id rater id*rater;
run;
****************************************.
                                        ;
```

```
data one;
input id 1-2 rater 4 rating 5;
CARDS;
01 13
01 24
02 14
02 25
03 13
03 23
04 12
04 21
05 15
05 24
06 12
06 22
;
run;
*****************************************************************;
%GRIP(indata=one,targetv=id,repeatv=rater,dv=rating,
      type=1,t1=test of GRIP macro,t2=);
```

GRIP macro is available at: http://gim.med.ucla.edu/FacultyPages/Hays/util.htm

```
data one;
 input id 1-2 rater1 4 rater2 5;
 control=1;
CARDS;
01 34
02 45
03 33
04 21
05 54
06 22
;
run;
**************.
                ;
DATA DUMMY;
INPUT id 1-2 rater1 4 rater2 5;
CARDS;
01 11
02 22
03 33
04 44
05 55
RUN;
```

```
DATA NEW;
SET ONE DUMMY;
PROC FREQ;
TABLES CONTROL*RATER1*RATER2
/NOCOL NOROW NOPERCENT AGREE;
*****************************************.
                                         ;
data one;
 set one;
 *****************************************.
                                         ;
proc means;
var rater1 rater2;
run;
*****************************************.
                                         ;
proc corr alpha;
var rater1 rater2;
run;
```