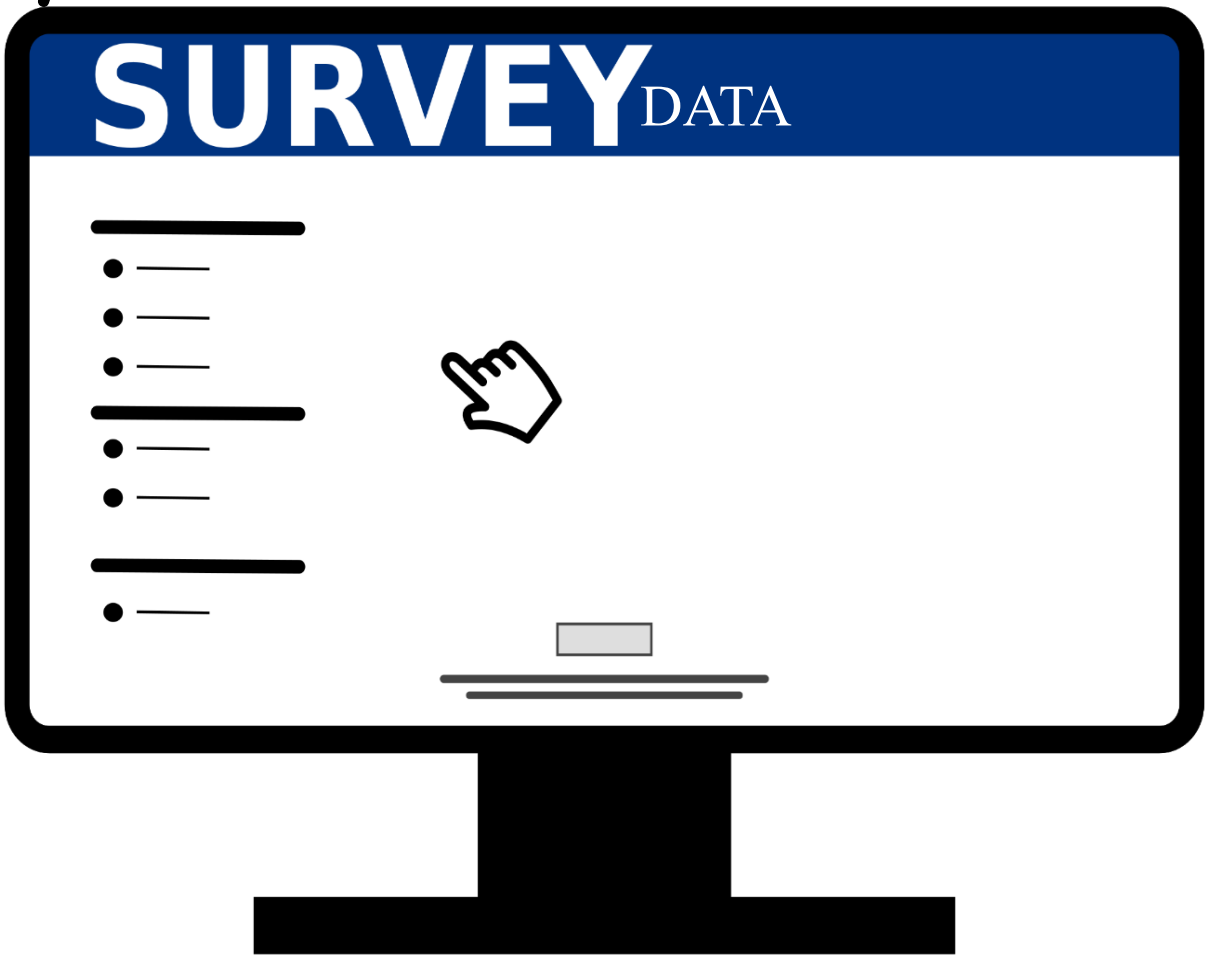
Psychometric Evaluation of



Questionnaire Design and Testing Workshop

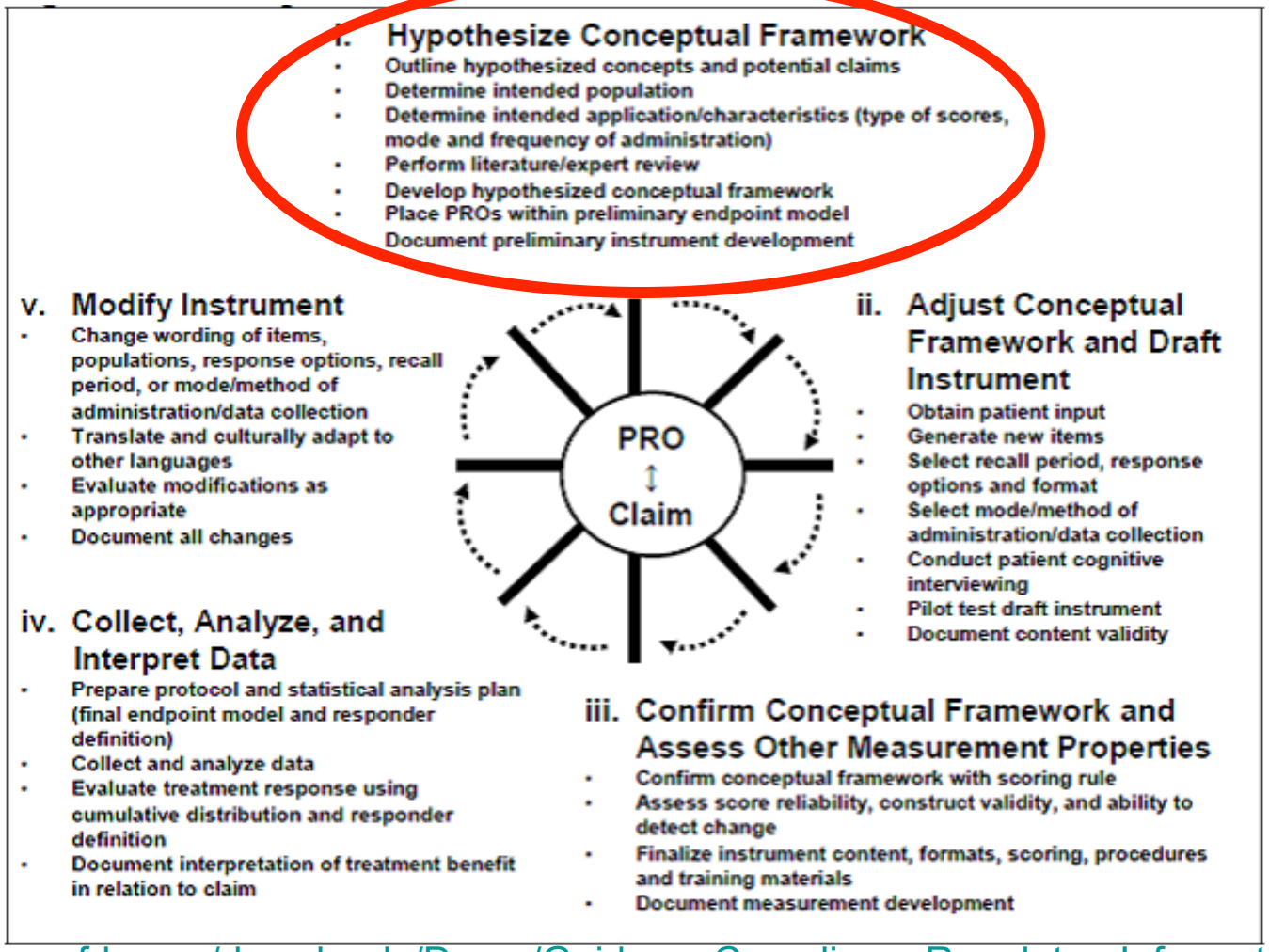
December 7 2015, 10:00-11:30 am

10940 Wilshire Suite 710

Patient-Reported Outcomes (PROs)

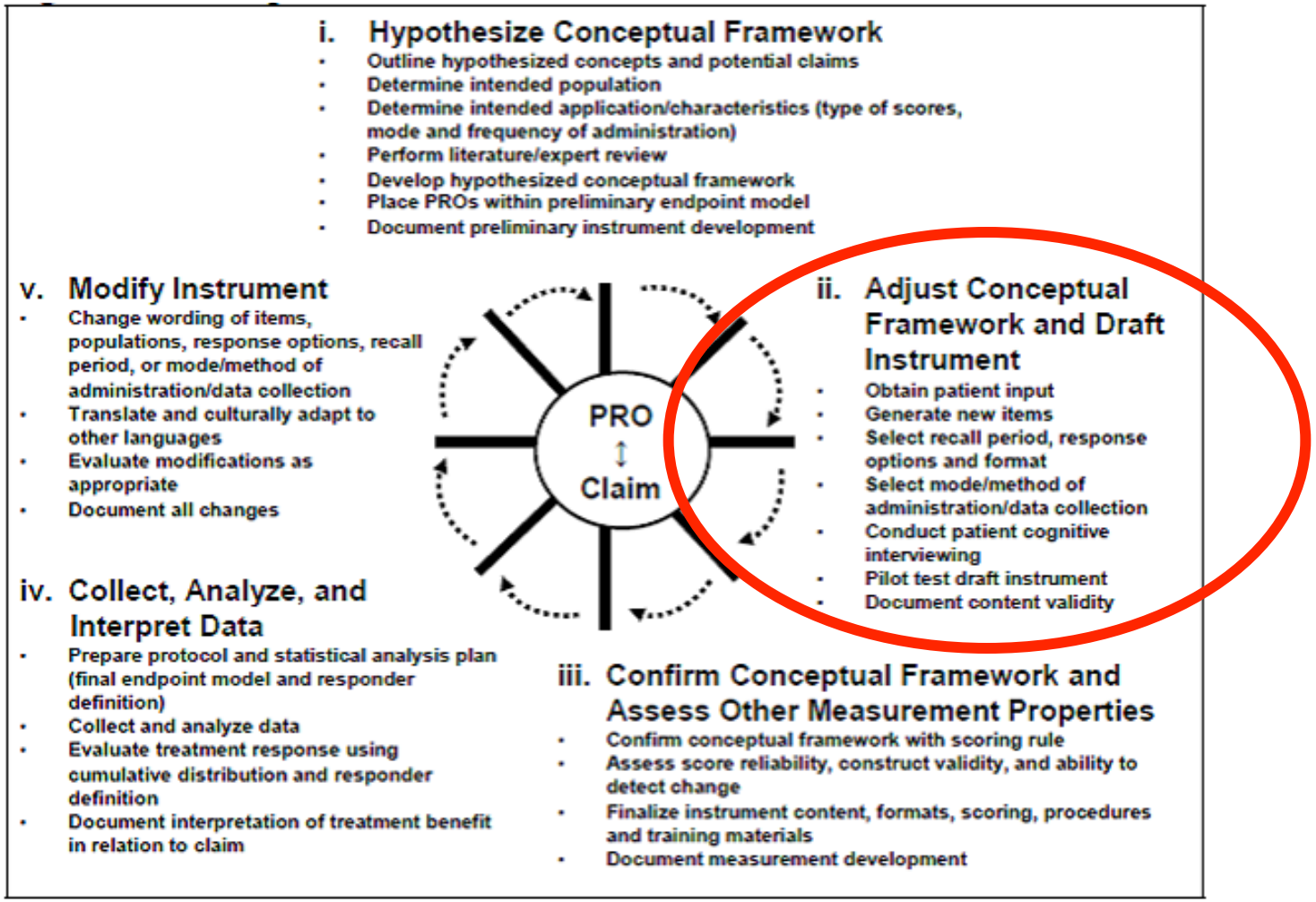
- "Any report of the status of a patient's health condition that comes directly from the patient, without interpretation of the patient's response by a clinician or anyone else"
  - Patient reports about their health
    - What they can do (functioning)
    - How they feel (well-being)

PRO Development Process



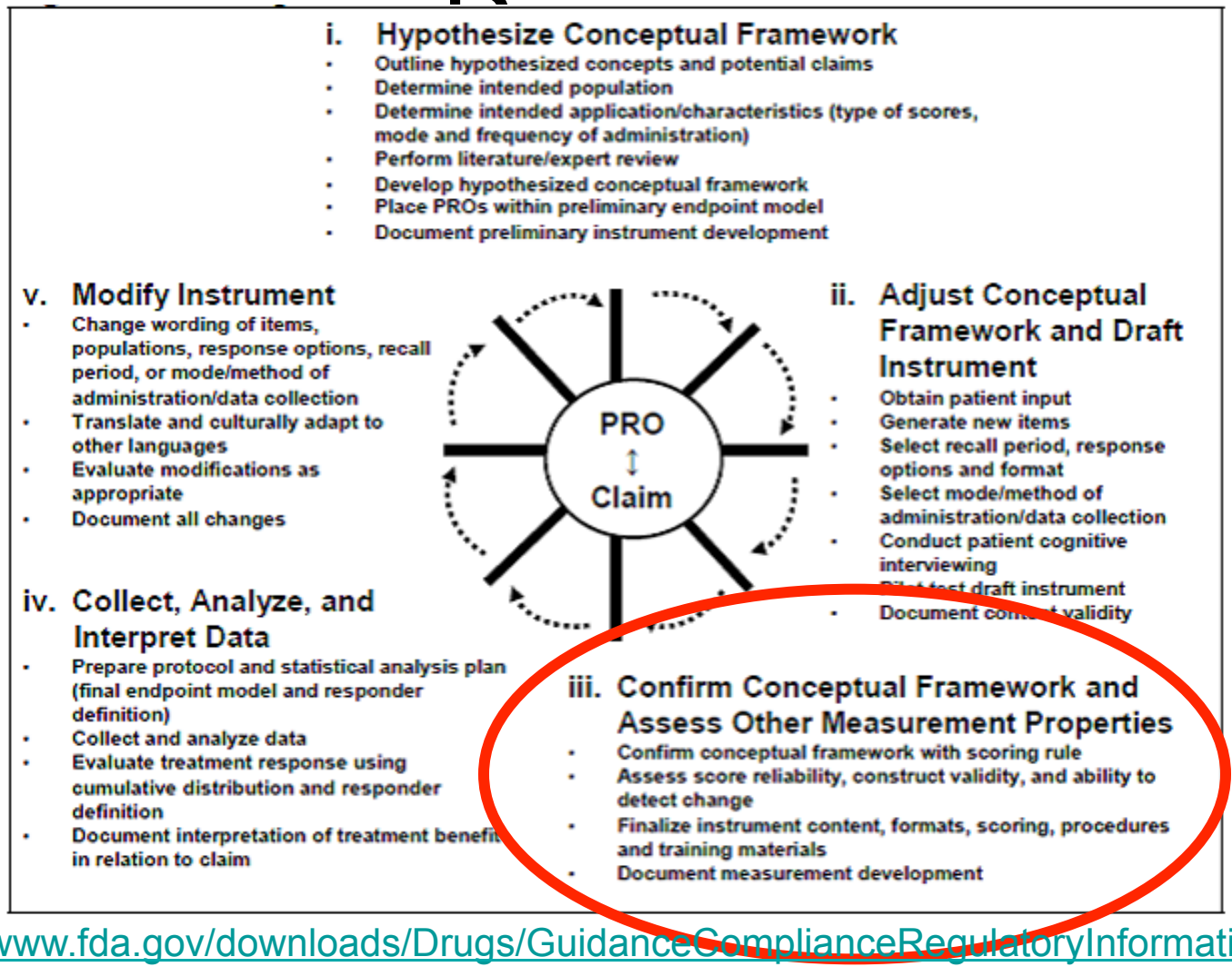
http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM205269.pdf

PRO Development Process



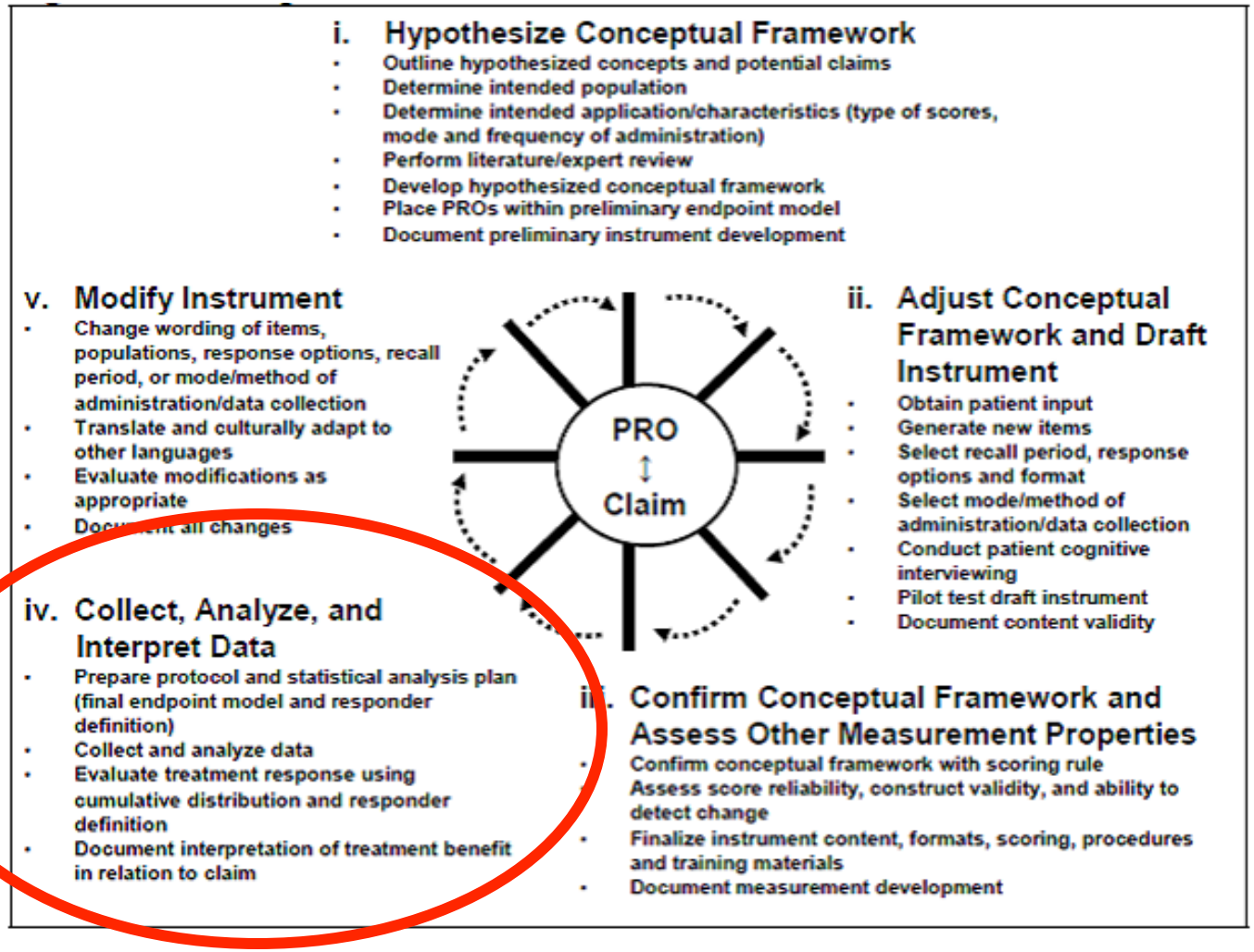
http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM205269.pdf

PRO Iterative Development



http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM205269.pdf

PRO Iterative Development Process



http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM205269.pdf

Physical Functioning

- Ability to conduct a variety of activities ranging from self-care to running
- Predictor of
  - Hospitalizations, institutionalization, and mortality
- Six physical functioning items included in 2010 Consumer Assessment of Healthcare Providers and Systems (CAHPS®) Medicare Survey

Because of a health or physical problem are you unable to do or have any difficulty doing the following activities?

- Walking?
- Getting in our out of chairs?
- Bathing?
- Dressing?
- Using the toilet?
- Eating?

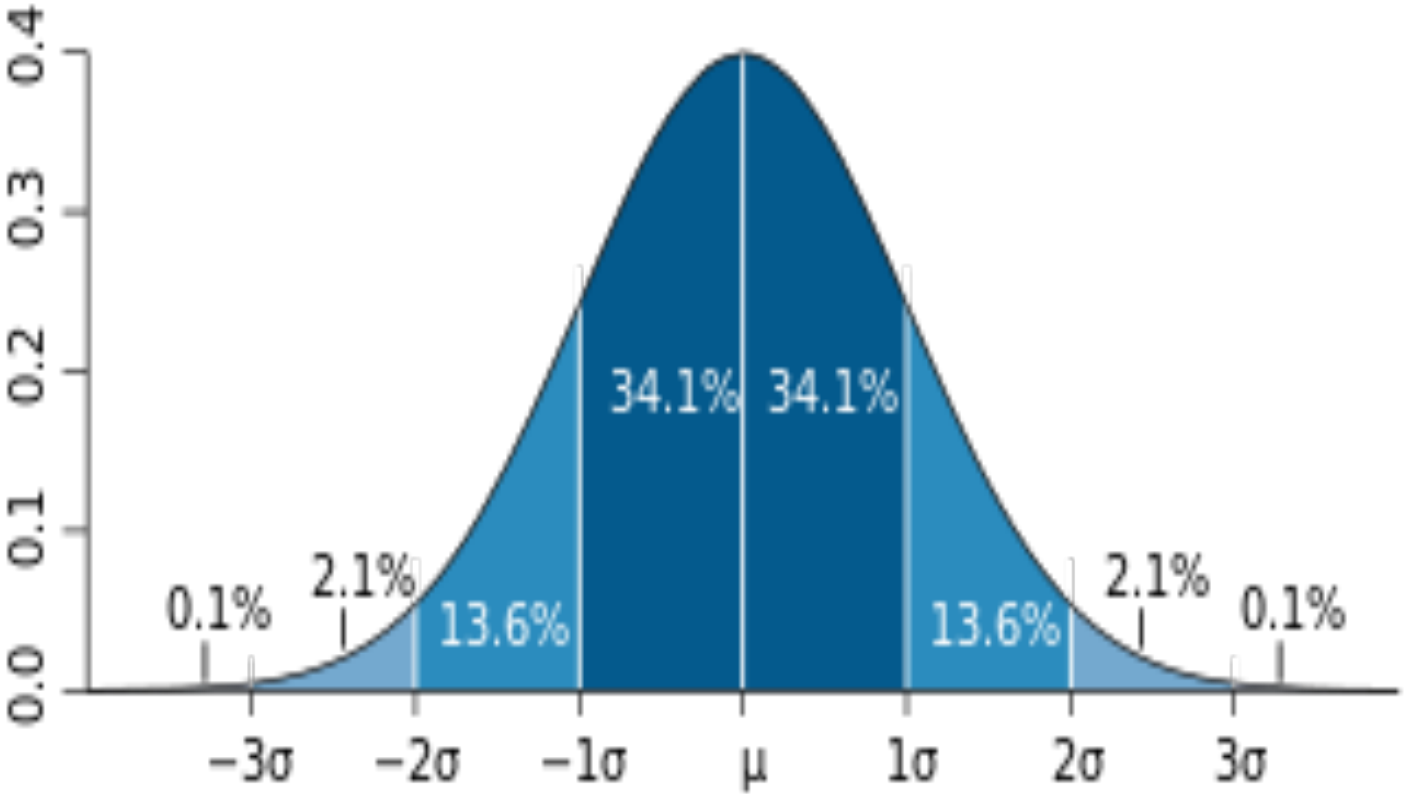
  - I am unable to do this activity (0)
  - Yes, I have difficulty (1)
  - No, I do not have difficulty (2)

Medicare beneficiary sample (n = 366,701)

- 58% female
- 57% high school education or less
- 14% 18-64; 48% 65-74, 29% 75-84, 9% 85+

Normal Curve

Bell-shaped “normal” curve (68.2%, 95.4%, and 99.6%)

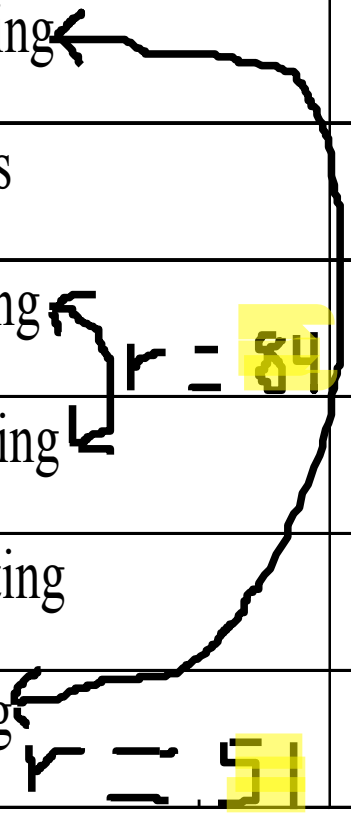


% of Medicare beneficiaries (n = 366,701) selecting each response option

Item	Unable to do	Have difficulty	No difficulty
Walking	4	27	69
Chairs	3	19	78
Bathing	4	11	85
Dressing	3	9	88
Toileting	3	6	91
Eating	3	3	94

% of Medicare beneficiaries (n = 366,701) selecting each response option

Item	Unable to do	Have difficulty	No difficulty
Walking	4	27	69
Chairs	3	19	78
Bathing	4	11	85
Dressing	3	9	88
Toileting	3	6	91
Eating	3	3	94

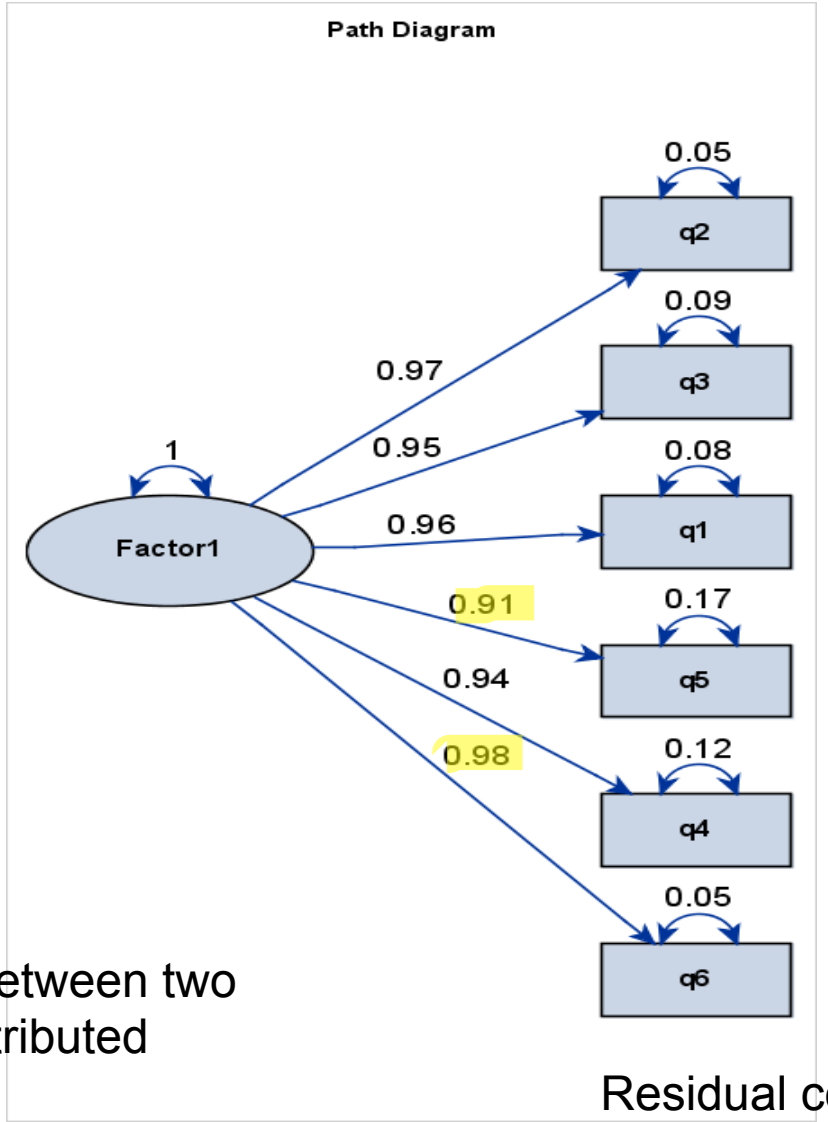


Item-Scale Correlations

Item	Item-Scale Correlations
Walking (0, 1, 2)	0.71
Chairs (0, 1, 2)	0.80
Bathing (0, 1, 2)	0.83
Dressing (0, 1, 2)	0.86
Toileting (0, 1, 2)	0.84
Eating (0, 1, 2)	0.75

Possible 6-item scale range: 0-12 (2% floor, 65% ceiling)

Confirmatory Factor Analysis (Polychoric* Correlations)



Dressing

Eating

Bathing

Walking

Chairs

Toileting

* Estimated correlation between two underlying normally distributed continuous variables

Residual correlations <= 0.04

Reliability

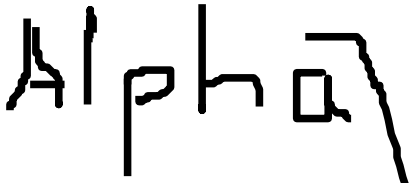
Degree to which the same score is obtained when the *target* or thing being measured (person, plant or whatever) has not changed.

- ✓Internal consistency (items)
  - ✓Need 2 or more items
- ✓Test-retest (administrations) correlations
  - ✓Need 2 or more time points

Reliability

Model	Reliability	Intraclass Correlation
Two-way random	$\frac{N(MS_{BMS} - MS_{EMS})}{NMS_{BMS} + MS_{JMS} - MS_{EMS}}$	$\frac{MS_{BMS} - MS_{EMS}}{MS_{BMS} + (k-1)MS_{EMS} + k(MS_{JMS} - MS_{EMS})/N}$
Two-way mixed	$\frac{MS_{BMS} - MS_{EMS}}{MS_{BMS}}$	$\frac{MS_{BMS} - MS_{EMS}}{MS_{BMS} + (k-1)MS_{EMS}}$
One-way	$\frac{MS_{BMS} - MS_{WMS}}{MS_{BMS}}$	$\frac{MS_{BMS} - MS_{WMS}}{MS_{BMS} + (k-1)MS_{WMS}}$

BMS = Between Ratee Mean Square N = n of ratees
WMS = Within Mean Square k = n of items or raters
JMS = Item or Rater Mean Square
EMS = Ratee x Item (Rater) Mean Square



Reliability Formulas

Model	Reliability	Intraclass Correlation
Two-way random	$\frac{N(MS_{BMS} - MS_{EMS})}{NMS_{BMS} + MS_{JMS} - MS_{EMS}}$	$\frac{MS_{BMS} - MS_{EMS}}{MS_{BMS} + (k-1)MS_{EMS} + k(MS_{JMS} - MS_{EMS})/N}$
Two-way mixed	$\frac{MS_{BMS} - MS_{EMS}}{MS_{BMS}}$	$\frac{MS_{BMS} - MS_{EMS}}{MS_{BMS} + (k-1)MS_{EMS}}$
One-way	$\frac{MS_{BMS} - MS_{WMS}}{MS_{BMS}}$	$\frac{MS_{BMS} - MS_{WMS}}{MS_{BMS} + (k-1)MS_{WMS}}$

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

Internal Consistency Reliability (Coefficient Alpha)

- Coefficient alpha = 0.92

  $(MS_{bms} - MS_{ems})/MS_{bms}$

- Ordinal alpha = 0.98

  http://support.sas.com/resources/papers/proceedings14/2042-2014.pdf

Item-scale correlation matrix ("Multi-trait Scaling")

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

*Item-scale correlation, corrected for overlap.

Item-scale correlation matrix

	Phyfun	Anger
Item #1	0.80*	0.80
Item #2	0.80*	0.80
Item #3	0.80*	0.80
Item #4	0.80*	0.80
Item #5	0.80*	0.80
Item #6	0.80*	0.80
Item #7	0.50	0.80*
Item #8	0.50	0.80*
Item #9	0.50	0.80*

*Item-scale correlation, corrected for overlap.

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

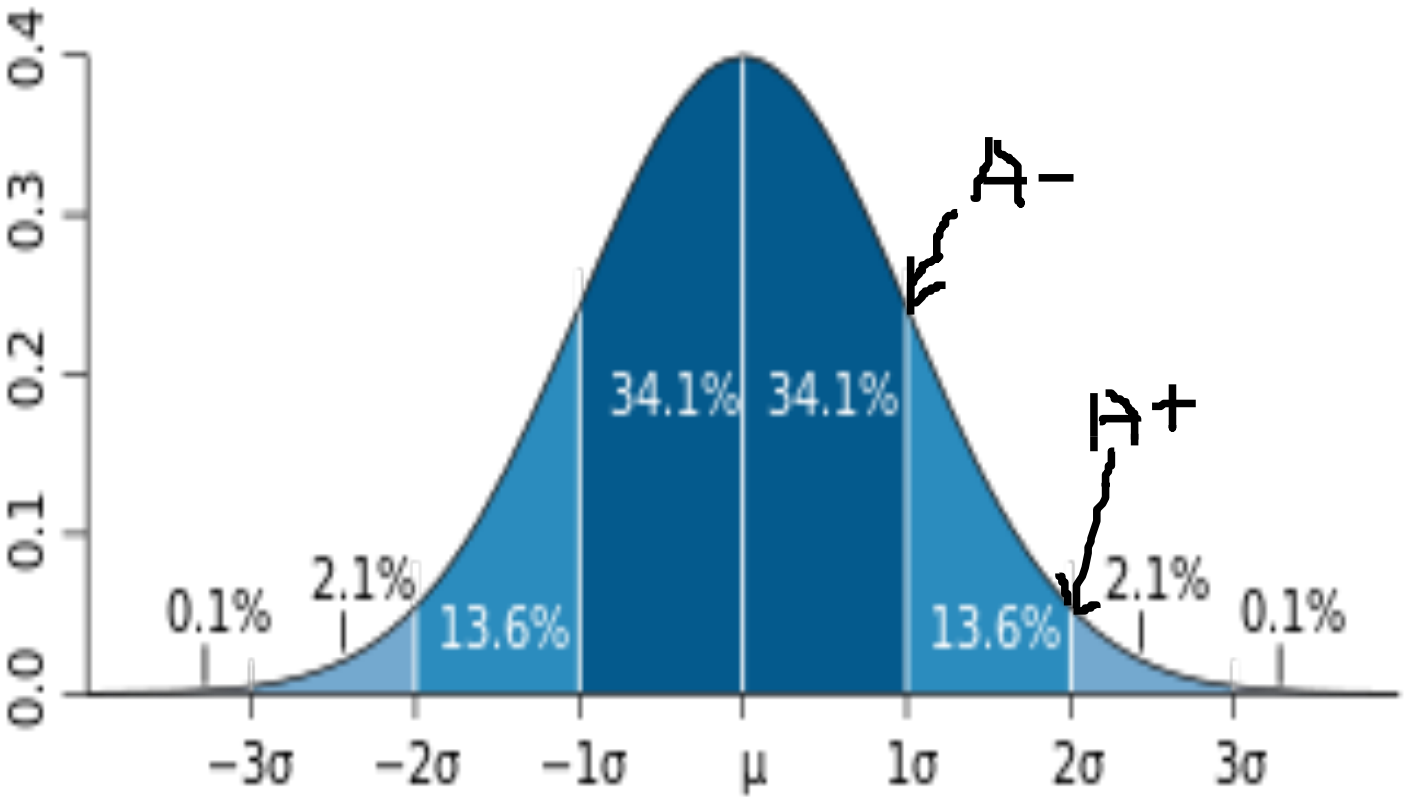
$$\Pr(Y_i \geq k) = \frac{1}{1 + \exp(-a_i\theta + b_{ik})}$$

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

a_i estimates the discriminatory power of the item.

Normal Curve (bell-shaped)

z = -1 to 1 (68.2%); z = -2 to 2 (95.4%); z = -3 to 3 (99.6%)



% of Medicare beneficiaries (n = 366,701) selecting each response option

Item	Unable to do 0	Have difficulty 1	No difficulty 2
Walking	4	27	69
Chairs	3	19	78
Bathing	4	11	85
Dressing	3	9	88
Toileting	3	6	91
Eating	3	3	94

% of Medicare beneficiaries (n = 366,701) selecting each response option

Item	Unable to do	Have difficulty	No difficulty
Walking	4	27	69
Chairs	3	19	78
Bathing	4	11	85
Dressing	3	9	88
Toileting	3	6	91
Eating	3	3	94

% of Medicare beneficiaries (n = 366,701) selecting each response option

Item	Unable to do	Have difficulty	No difficulty
Walking	4	27	69
Chairs	3	19	78
Bathing	4	11	85
Dressing	3	9	88
Toileting	3	6	91
Eating	3	3	94

Threshold #1 Parameter (Graded Response Model)

Physical Functioning	1st Threshold *Unable to do (lowest 2%)*
Walking	-1.86
Chairs	-1.91
Bathing	-1.72
Dressing	-1.78
Toileting	-1.87
Eating	-1.98

Threshold #2 Parameter (Graded Response Model)

Physical Functioning		2nd Threshold *Unable to do or have difficulty*
Walking		-0.55 (lowest 32%)
Chairs		-0.81
Bathing		-1.02
Dressing		-1.10
Toileting		-1.27
Eating		-1.53 (lowest 9%)

Item Parameters (Graded Response Model)

Physical Functioning	1st Threshold *Unable to do*	2nd Threshold *Have difficulty*	Slope (Discrimination)
Walking	-1.86	-0.55	4.63
Chairs	-1.91	-0.81	5.65
Bathing	-1.72	-1.02	6.34
Dressing	-1.78	-1.10	8.23
Toileting	-1.87	-1.27	7.23
Eating	-1.98	-1.53	4.87

The IRT Procedure

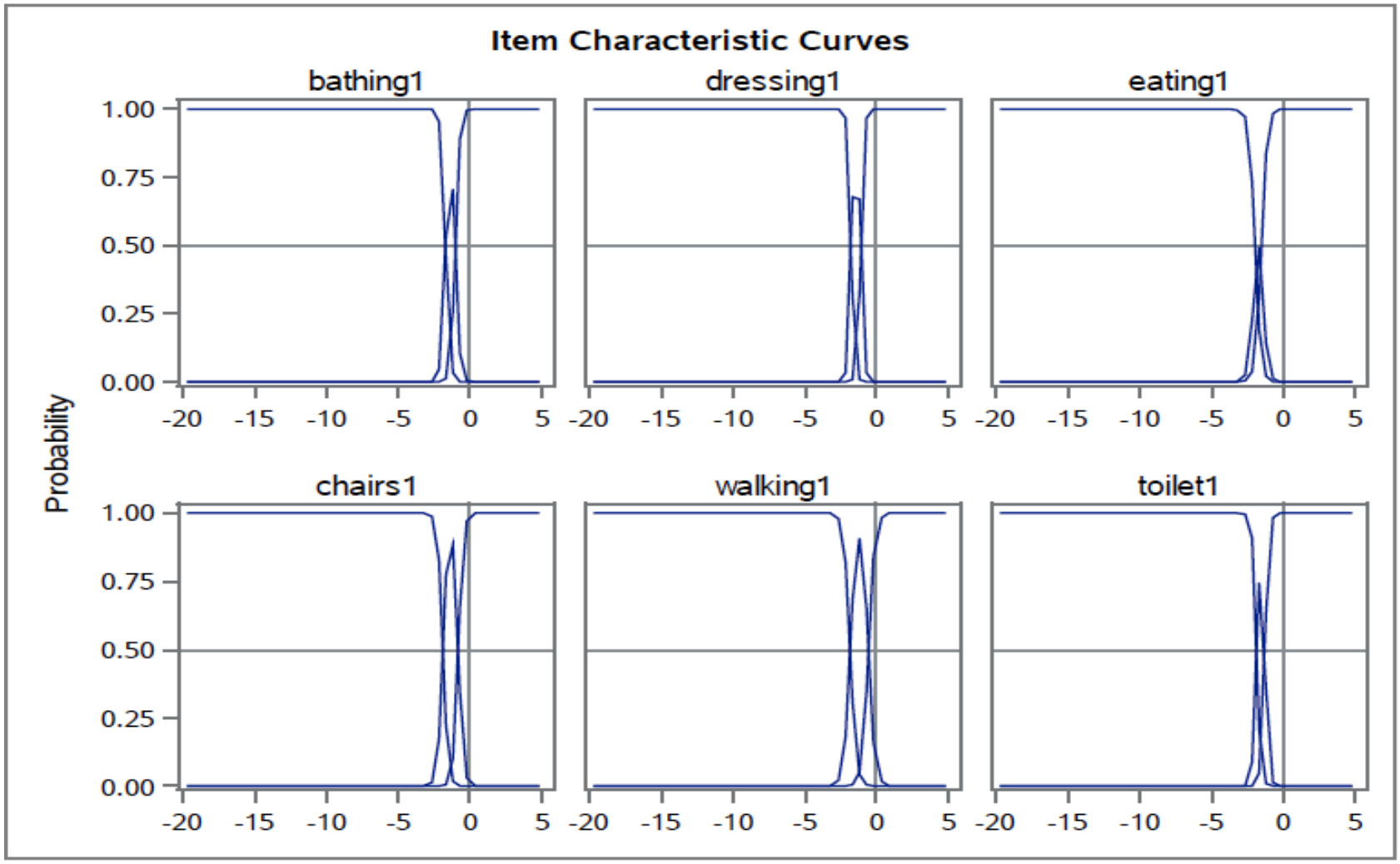


Figure 2. Person-Item Map

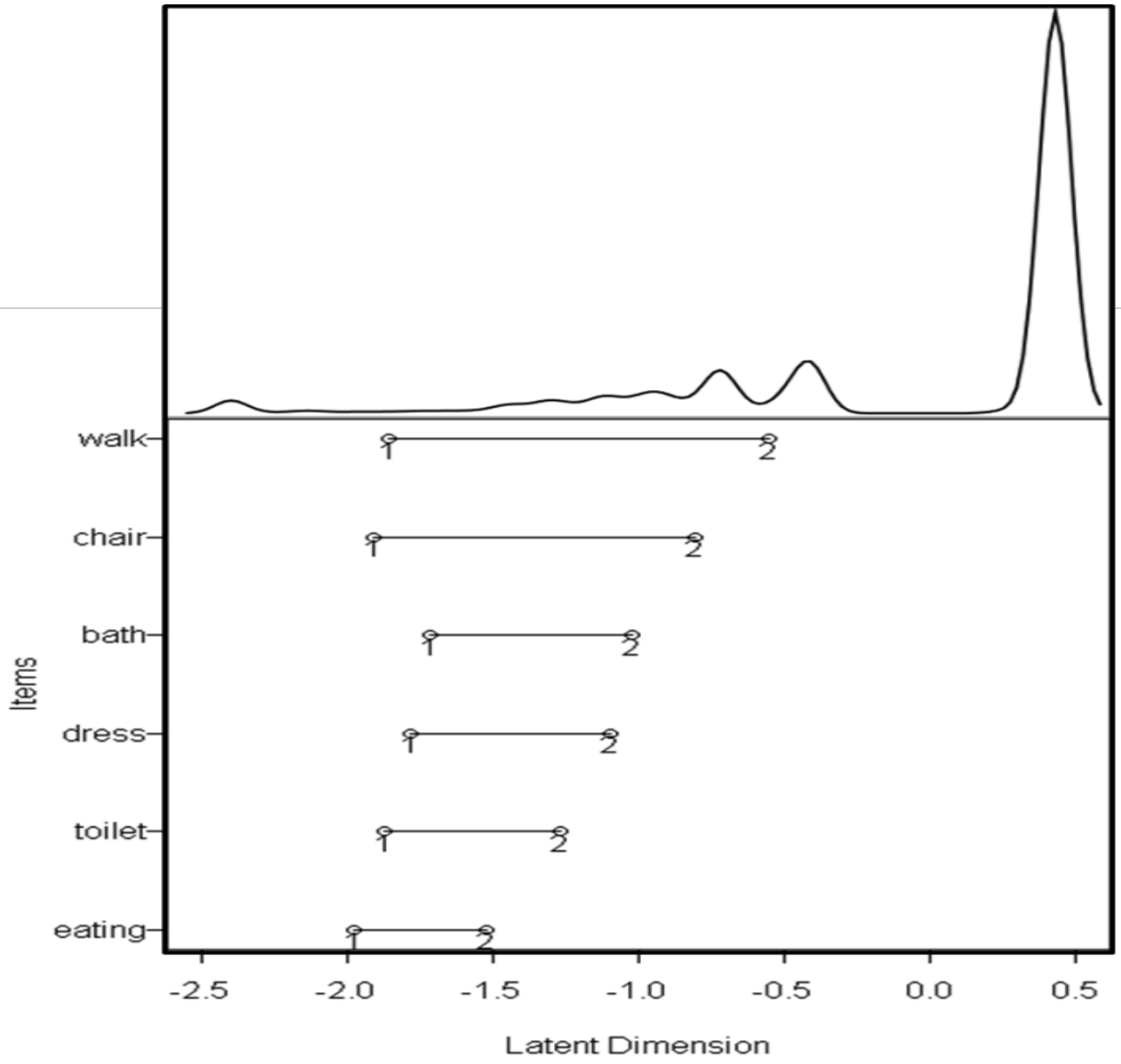
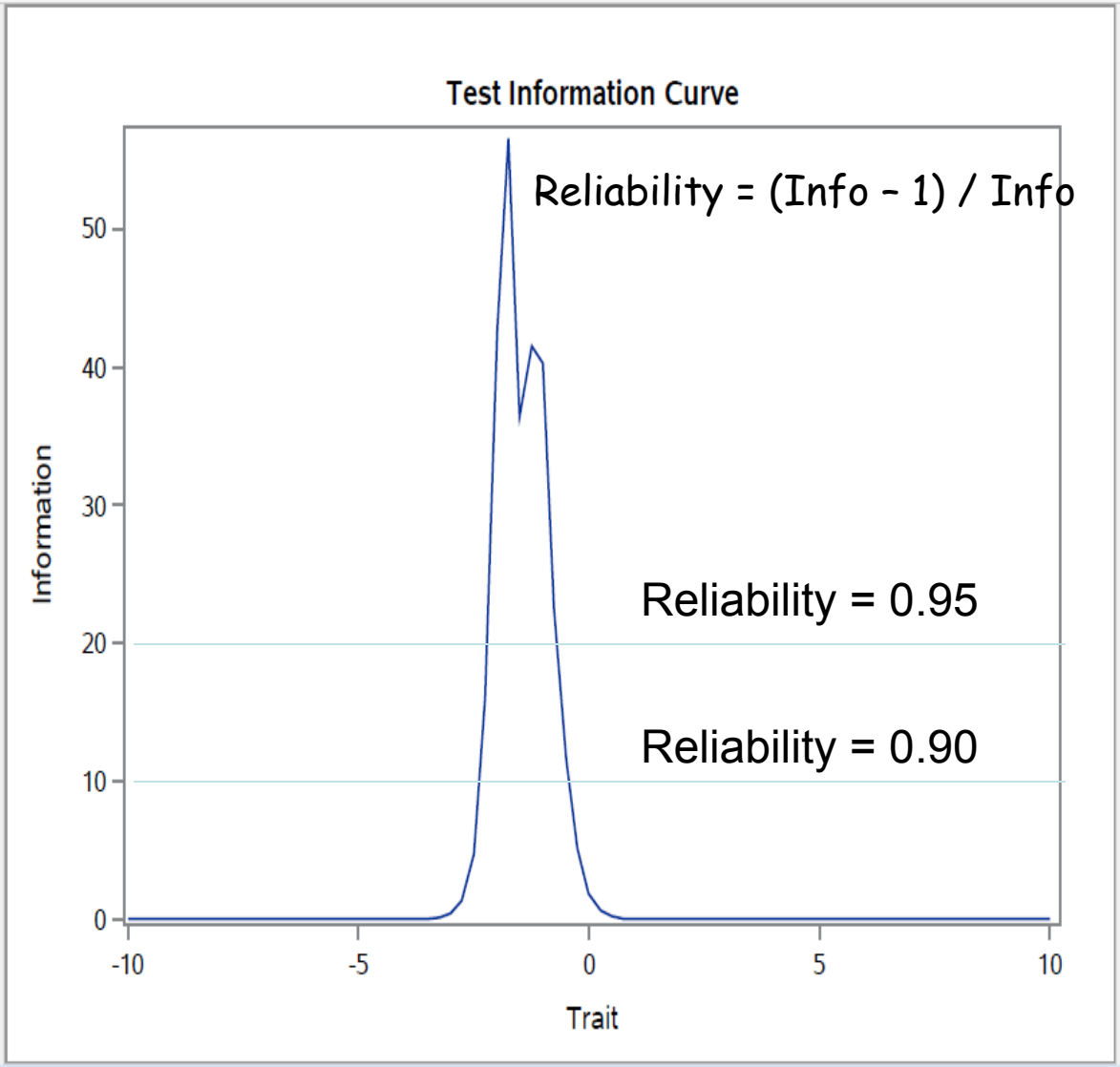


figure3.pdf - Adobe Acrobat Reader DC

File Edit View Window Help

Home Tools Document 1 / 1 113% Sign In



Validity

- Content validity: Does measure "appear" to reflect what it is intended to (expert judges or patient judgments)?
  - Do items operationalize concept?
  - Do items cover all aspects of concept?
  - Does scale name represent item content?
- Construct validity
  - Are the associations of the measure with other variables consistent with hypotheses?

Physical Function Scale Correlations

r = 0.39 (self-rated general health)

r = -0.23 (number of chronic conditions)

Cohen's rule of thumb for correlations that correspond to effect size rules of 0.20 SD, 0.50 SD and 0.80 SD are as follows:

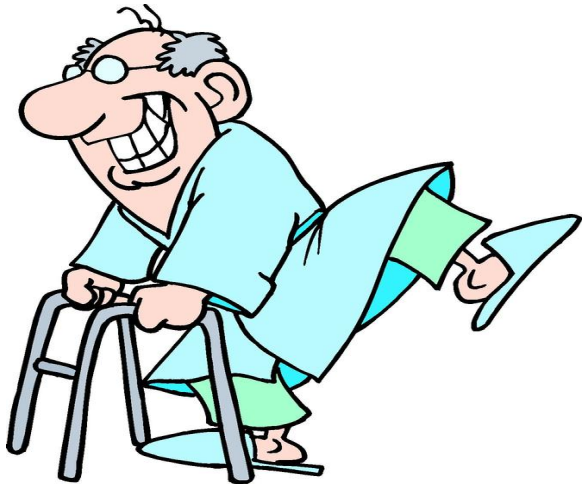
0.100 is small correlation
0.243 is medium correlation
0.371 is large correlation

(r's of 0.10, 0.30 and 0.50 are often cited as small, medium and large, respectively).

Summary

- The 6 physical functioning items target relatively easy activities

- Items representing higher levels of physical functioning are needed for the majority of Medicare beneficiaries.
  - Lifting or carrying groceries
  - Doing chores like vacuuming or yard work
  - Running a short distance

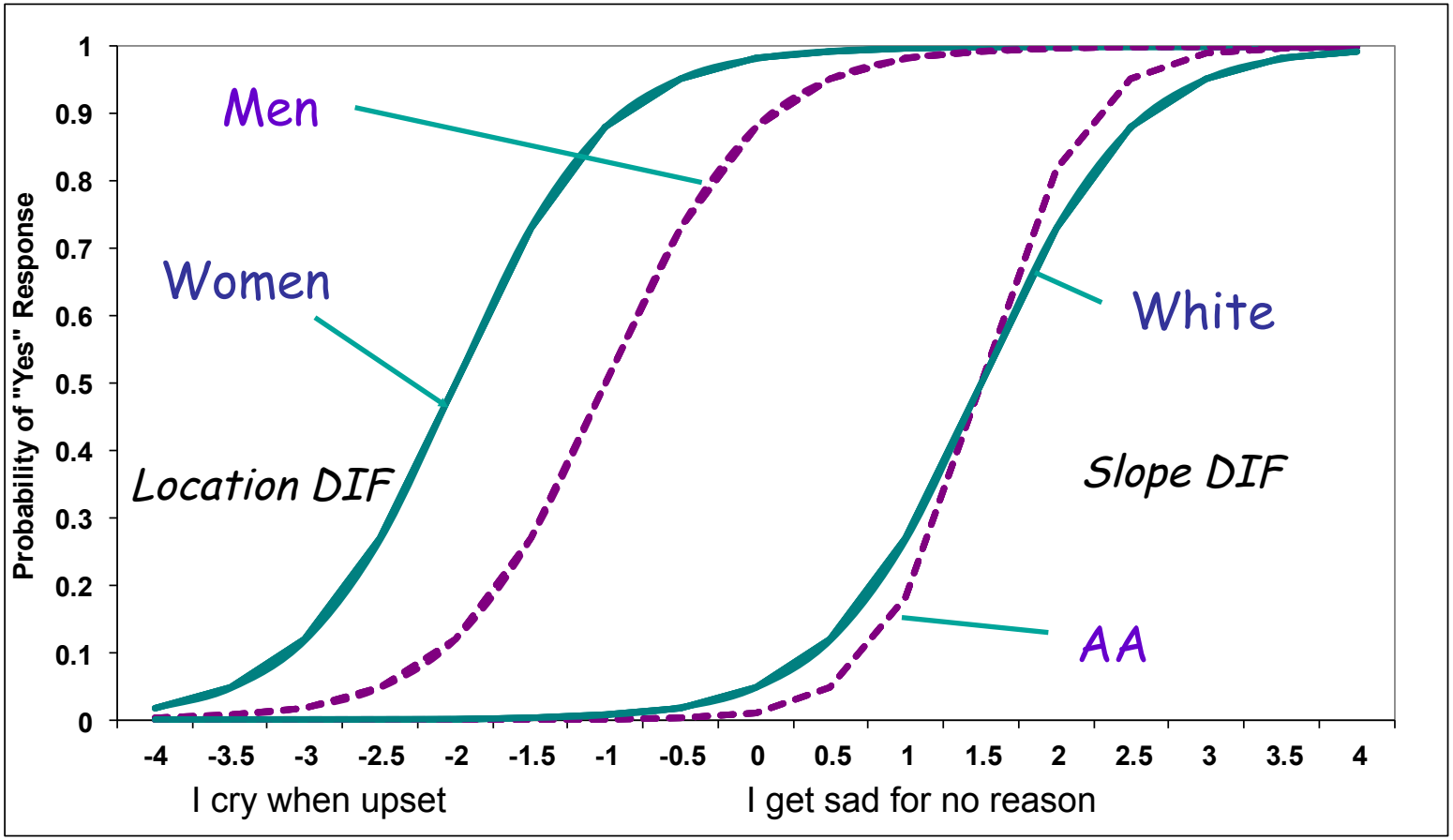


Differential Item Functioning (DIF)

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

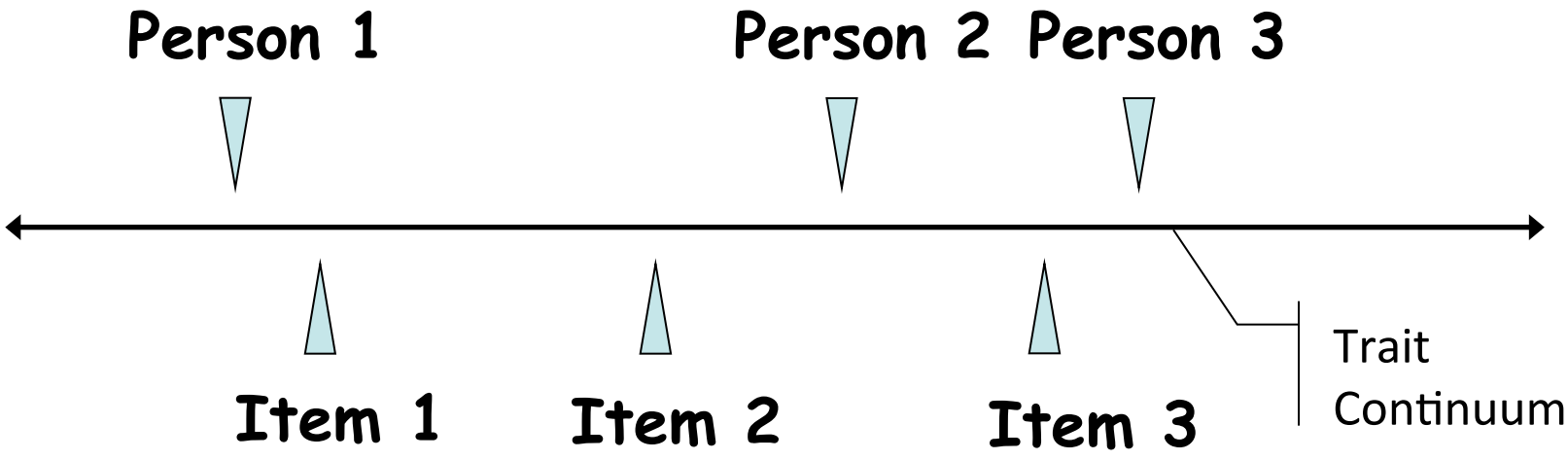
- Evaluation of DIF by subgroups

DIF (2-parameter model)



Higher Score = More Depressive Symptoms

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
- $SE = SD\ (1 - \text{reliability})^{1/2}$
- $\text{Reliability} = 1 - (SE/10)^2$
  - Reliability = 0.90 when SE = 3.2
  - 95% CI = true score +/- 1.96 x SE

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) (95% CI: 44.7-55.7)

Recommended Reading

- Cappelleri, J. C., Lundy, J.J., & Hays, R. D. (2014). Overview of classical test theory and item response theory for quantitative assessment of items in developing patient-reported outcome measures. Clinical Therapeutics, 36 (5), 648-662

Thank You!

drhays@ucla.edu (310-794-2294). http://gim.med.ucla.edu/FacultyPages/Hays/