Are the 6 Physical Functioning Items in the CAHPS Medicare Survey Informative?

#### Ron D. Hays, Ph.D. UCLA Department of Medicine March 17, 2016 12 noon – 1pm

Live Streaming Webinar With Audio: Uclahealth.webex.com Meeting Number: 802702422 Meeting Password: uclagim

## Physical Functioning

- Able to do a range of activities from basic (e.g., self-care) to advanced (e.g., running)
- Six physical functioning items included in the 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?



## Medicare beneficiary sample (n = 366,701)

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



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

Item (Some difficulty)	Unable to do	Have difficulty	No difficulty
Walking (1/3)	4	27	69
Chairs (1/5)	3	19	78
Bathing (1/7)	4	11	85
Dressing (1/9)	3	9	88
Toileting (1/10)	3	6	91
Eating (1/16)	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 Sr = 34	4	11	85
Dressing <b>L</b>	3	9	88
Toileting	3	6	91
Eating 51	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

# Alpha Reliability Formulas



## Internal Consistency Reliability (Coefficient Alpha)

- Coefficient alpha = 0.92(MS<sub>bms</sub> - MS<sub>ems</sub>)/MS<sub>bms</sub>
- Ordinal alpha = 0.98

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

-<u>http://gim.med.ucla.edu/FacultyPages/Hays/utils/</u>

## Confirmatory Factor Analysis (Polychoric\* Correlations)



\*Estimated correlation between two underlying normally distributed continuous variables

**Residual correlations** <= 0.04



R. M. Kaplan and D. P. Saccuzzo, Psychological Testing: Principles, Applications, and Issues (2<sup>nd</sup> Edition). Brooks/Cole Publishing Company1989 (page 152).



FIGURE 6-3 Item characteristic curve for a test item that discriminates well at low levels of performance but not at higher levels.

### Item Characteristic Curves



#### MINNESOTA LIVING WITH HEART FAILURE® QUESTIONNAIRE

The following questions ask how much your heart failure (heart condition) affected your life during the past month (4 weeks). After each question, circle the 0, 1, 2, 3, 4 or 5 to show how much your life was affected. If a question does not apply to you, circle the 0 after that question.

Did your heart failure prevent you from living as you wanted during		Very				Very
the past month (4 weeks) by -	No	Little				Much
<ol> <li>causing swelling in your ankles or legs?</li> <li>making you sit or lie down to rest during</li> </ol>	0	1	2	3	4	5
the day?	0	1	2	3	4	5
3. making your walking about or climbing stairs difficult?	0	1	2	3	4	5
4. making your working around the house or yard difficult?	0	1	2	3	4	5
5. making your going places away from home difficult?	0	1	2	3	4	5
<ol><li>making your sleeping well at night difficult?</li></ol>	0	1	2	3	4	5
<ol><li>making your relating to or doing things with your friends or family difficult?</li></ol>	0	1	2	3	4	5
8. making your working to earn a living difficult?	0	1	2	3	4	5
9. making your recreational pastimes, sports						
or hobbies difficult?	0	1	2	3	4	5
10. making your sexual activities difficult?	0	1	2	3	4	5

## Item Characteristic Curve for Emotional Health Scale

The IRT Procedure



## Simple-summated Scoring of Physical Functioning Scale

- I am unable to do this activity (0)
- Yes, I have difficulty (1)
- No, I do not have difficulty (2)
- Possible 6-item scale range: 0-12
  Mean = 11 (2% floor, 65% ceiling)





## Reliability = (Info - 1) / Info

The IRT Procedure



#### Item Response Theory (IRT)

IRT graded response model estimates relationship between a person's response  $Y_i$  to the question (i) and his or her level on the latent construct ( $\theta$ ):

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

 $\mathbf{b}_{ik}$  estimates how difficult it is to have a score of k or more on item (i).

a<sub>i</sub> estimates item discrimination.

## People and Items on Same *z-score* metric





Are you able to get in and out of bed? Are you able to stand without losing your balance for 1 minute? Are you able to walk from one room to another? Are you able to walk a block on flat ground? Are you able to run or jog for two miles? Are you able to run five miles?



- BIGSTEPS and WINSTEPS
- PARSCALE and MULTILOG

IRTPRO and FLEXMIRT

SAS and STATA



#### Computer Adaptive Testing (CAT)







www.nihpromis.org

### Reliability Target for Use of Measures with Individuals

- z-score (mean = 0, SD = 1)
- Reliability ranges from 0-1
  - 0.90 or above is goal
  - SE = SD (1- reliability)<sup>1/2</sup>
  - Reliability = 1 SE<sup>2</sup>
  - Reliability = 0.90 when <u>SE = 0.32</u>
- 95% CI = true score +/- 1.96 x SE

$$(CI = -0.63 \rightarrow 0.63$$
 z-score when reliability = 0.90)

#### DIF (2-parameter model)



## Person Fit

- Large negative  $Z_L$  values indicate misfit.
- One person in PROMIS project had  $Z_L = -3.13$
- This person reported that they could do 13 physical functioning activities (including running 5 miles) <u>without any</u> <u>difficulty</u>, but
  - This person reported <u>a little difficulty</u> being out of bed for most of the day.

#### IRT Distortions

- "Parameter values are identical in separate subgroups or across different measurement conditions."
- It is the often misunderstood feature of parameter invariance that is frequently cited in introductory or advanced texts" (Rupp & Zumbo, 2006).

### Item Parameters (Graded Response Model)

Physical Functioning	1 <sup>st</sup> Threshold <i>Unable to do</i>	2 <sup>nd</sup> Threshold <i>Have difficulty</i>	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

## Interval-Level?

- "Modern day psychometric analyses such as Rasch analysis convert ordinal data to an interval scale so that response scores meet the criteria for measurement"
- Correlation (product-moment and ICC) between simple-summated scoring and IRT estimated score for physical functioning = 0.91

## Ben Wright or Been Wrong?

- "Application of the Rasch model to the data set estimates a measure that can be considered valid."
- The "Rasch model is the only valid approach to measurement"
  - Bergan, 2013, Rasch versus Birnbaum: New arguments in an old debate (p. 3)



#### **Questions?**



Hays, R. D., Mallett, J. S., Gaillot, S., & Elliott, M. N. (2015). Performance of the Medicare Consumer Assessment of Healthcare Providers and Systems (CAHPS®) Physical Functioning Items. <u>Medical Care, 54</u>, 205-209