

Remaining Challenges and What to Do Next: Undiscovered Areas

Ron D. Hays

UCLA Division of General Internal Medicine &
Health Services Research

[\(drhays@ucla.edu\)](mailto:drhays@ucla.edu)

Item Response Theory in Health Measurement

University of Twente Symposium

Van der Valk Hotel Schiphol, Netherlands

September 17, 2013

This Presentation is Dedicated to Last Night's Hero (Santa Claus)



Five Future Areas of Opportunity

- Person fit
- Distributional Issues
- Presence/severity
- Group-level Models
- Multidimensional IRT/Hierarchical Models

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:
 - *Longer response time*
 - *More chronic conditions*
 - *Younger age*

Distributional Issues

- Normal distribution for latent trait assumed in estimating item parameters using marginal ML
- Degree to which non-normality of latent distribution is consequential for IRT modeling
 - Extent to which violating normality assumption distorts item slope and threshold parameters
- Ramsay Curve IRT (Carol Woods)
 - Detects and adjusts for non-normal latent variables
 - Need “enough” items and respondents

Memorial Symptom Assessment Scale

Presence/Severity Example

(Liu & Verkuilen, 2013, Applied Psych Measurement)

Did you have diarrhea in the past 2 weeks?

- No (0)
- Yes

How much did diarrhea bother or distress you?

- Not at all (1)
- A little bit (2)
- Somewhat (3)
- Quite a bit (4)
- Very much (5)

Presence/Severity Ordinal Scoring

0 = No diarrhea in the past 2 weeks.

1 = Diarrhea and bothered/distressed not at all.

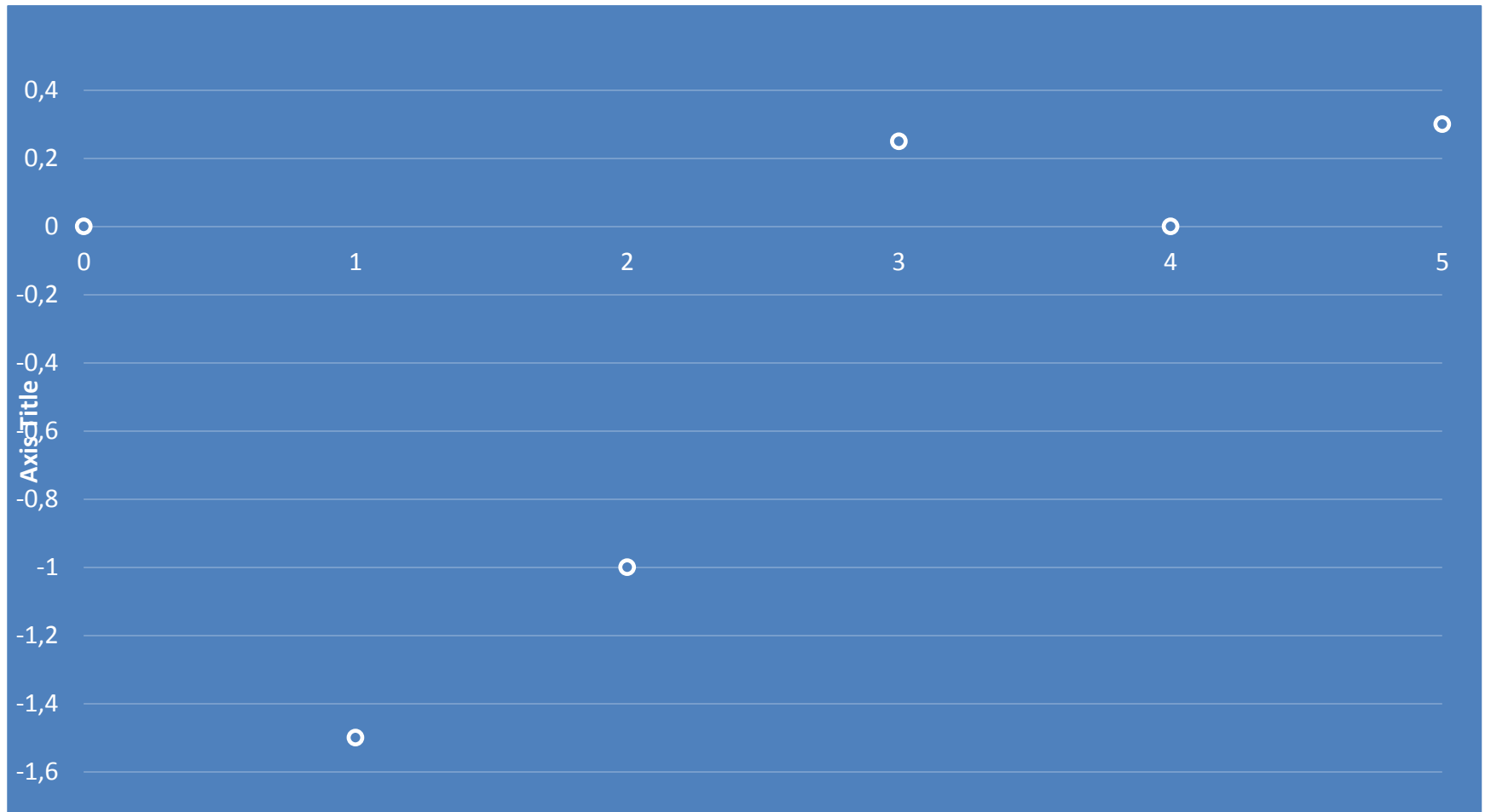
2 = Diarrhea and bothered/distressed a little bit.

3 = Diarrhea and bothered/distressed somewhat.

4 = Diarrhea and bothered/distressed quite a bit.

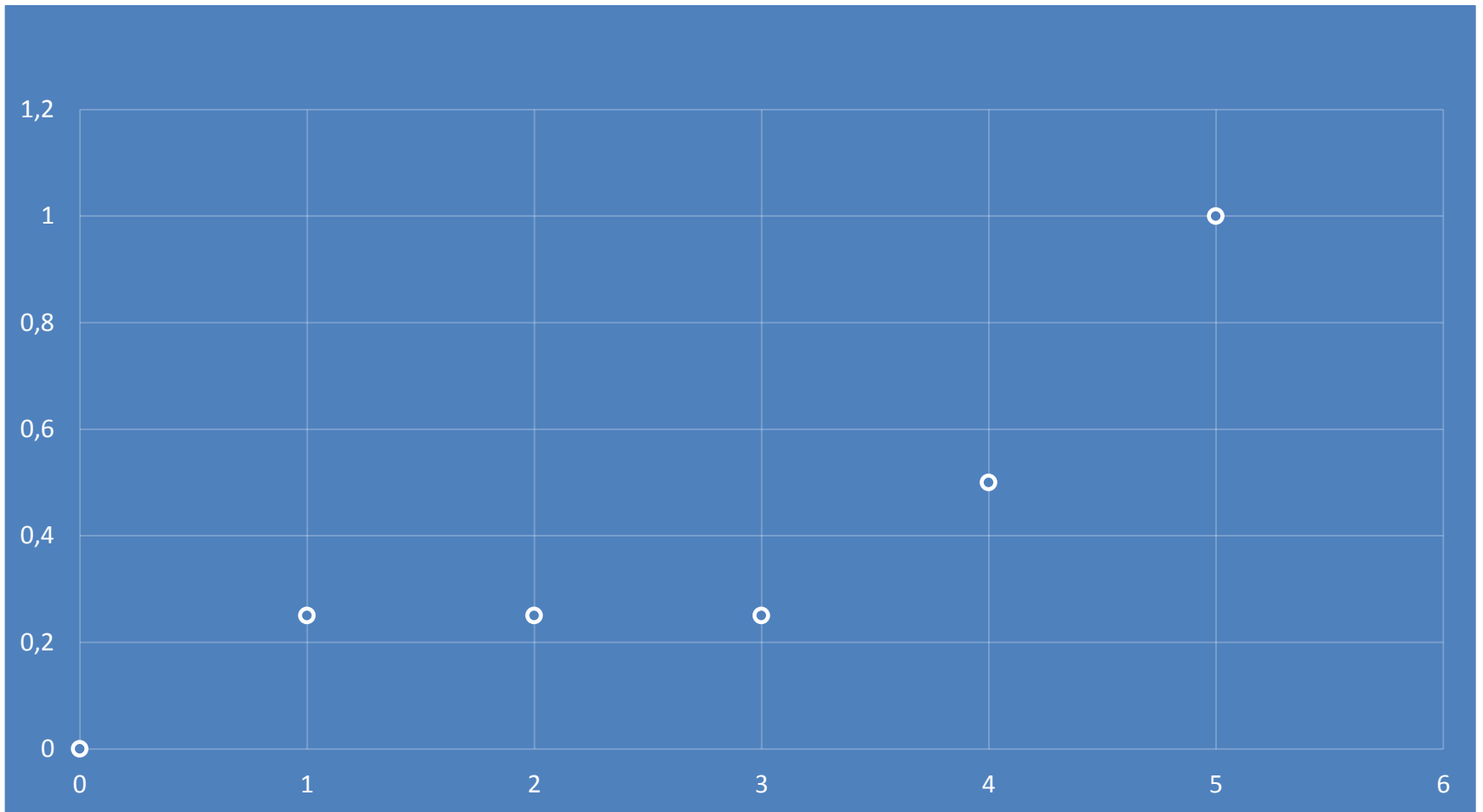
5 = Diarrhea and bothered/distressed very much.

δ - plot for Diarrhea item (Nominal Response Model)



x axis = 0 -5 categories; y axis = $a_k - a_0$ category severity

δ - plot for Pain Item



x axis = 0 -5 categories; y axis = $a_k - a_0$ category severity 10

Alternatives to Logistic/Normal-Ogive for Item Response Curves

- “Positive trait” item response models
 - Alcohol dependency (below average level of alcohol dependency is not meaningful)
- Latent trait scale begins at $\theta = 0$
 - Log-logistic (with $B = 1$ is Rasch’s original item response model)
 - Lognormal (similar to Steven’s psychophysical stimulus-response function)
 - Weibull (frequently used in biostatistics)
- Markov Chain Monte Carlo

CAHPS[®] Health Plan Survey

- 35,572 adults in 131 plans (n = 271 per plan)
- *Never/Sometimes/Usually/Always* items
 - Got help/advice needed when phoned
 - Got appointment for routine care as soon as wanted
 - Got appointment when ill or injured as soon wanted
 - Doctors listed carefully to you
 - Doctors explained things so you could understand
 - Doctors respect what you had to say
 - Doctors spent enough time with you
 - Office staff treated you with courtesy and respect
 - Office staff helpful

Group-Level Modeling

- Analogous to person-based IRT
- Estimate relationships between group trait level and proportion of group members (patients) that endorse each item

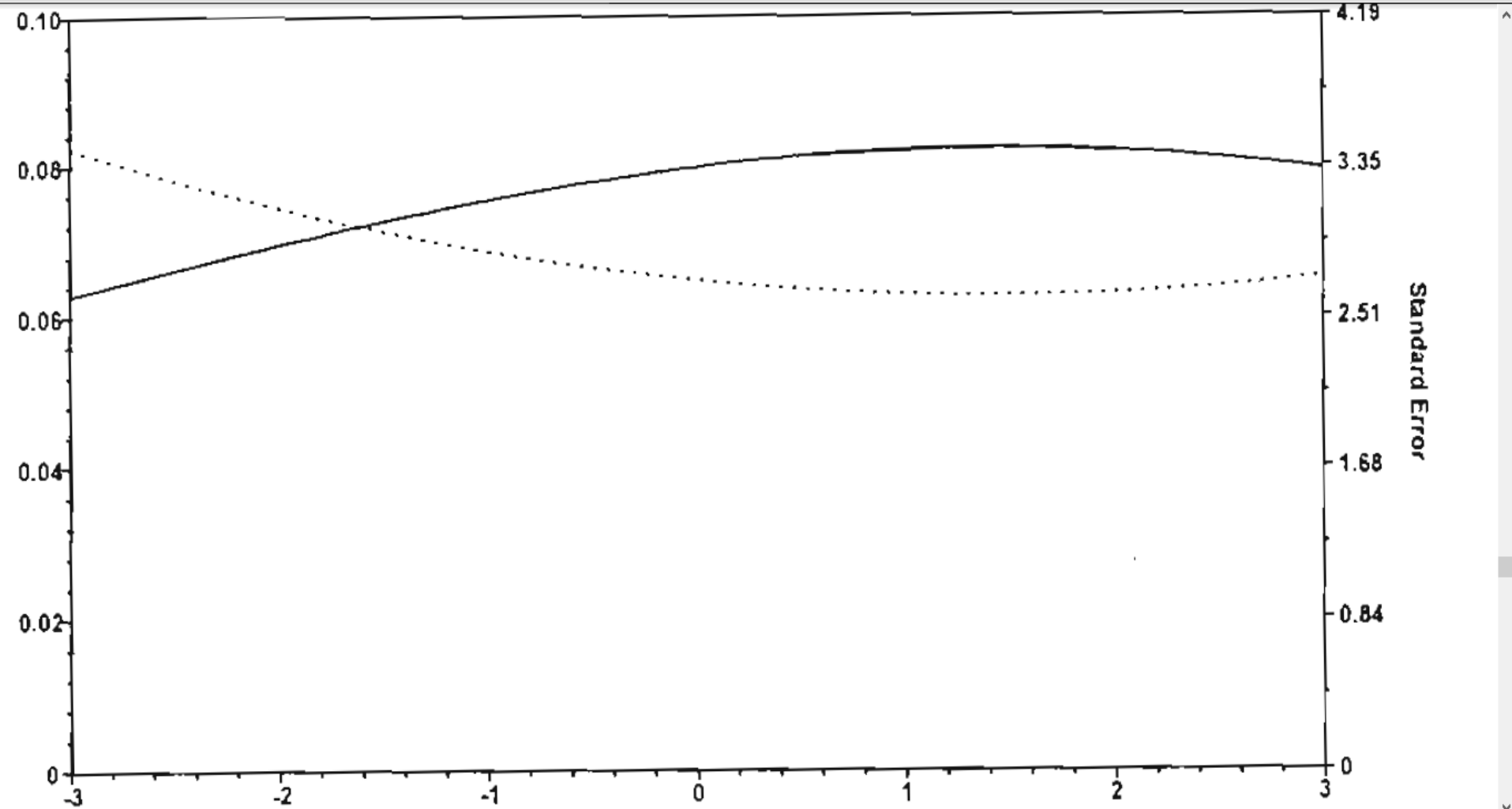


Reise, Meijer et al., 2006, Multi Beh Res

Items Do Not Discriminate Between Groups As Well as Among Persons

| Item | Person-Level Discrimination | Group (Plan) Level Discrimination |
|------|-----------------------------|-----------------------------------|
| C1 | 0.83 | 0.13 |
| C2 | 0.76 | 0.13 |
| C3 | 0.68 | 0.12 |
| C4 | 1.78 | 0.14 |
| C5 | 1.16 | 0.10 |
| C6 | 1.83 | 0.13 |
| C7 | 1.50 | 0.18 |
| C8 | 1.22 | 0.14 |
| C9 | 1.37 | 0.14 |

Health Plan-level Scale Information and SE for 9-Item Scale



Standard Error for Each Plan Depends Upon

- Plan's estimated θ
- Number of respondents



Multidimensional and Other Hierarchical Item Factor Models

<https://www.facebook.com/groups/370354816400530/members/>

- Multidimensional nominal models
 - Historically not easy to do, but IRTPRO makes it easier
- Cai, L. (2010). A two-tier full-information item factor analysis model with applications. *Psychometrika*, 75, 581-612.
 - Includes correlated traits multidimensional IRT, bifactor model, and testlet response theory models

Questions?

