Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

HHS Public Access

Author manuscript

Spine (Phila Pa 1976). Author manuscript; available in PMC 2023 February 13.

Published in final edited form as:

Spine (Phila Pa 1976). 2015 March 15; 40(6): E366–E371. doi:10.1097/BRS.0000000000000764.

Item Response Theory analysis of the Modified Roland-Morris Disability Questionnaire in a population-based study

Thelma J. Mielenz, PT, PhD[1,2,3], **Timothy S. Carey, MD, MPH**[2,4], **Michael C. Edwards, PhD**[5]

[1]Department of Epidemiology, Mailman School of Public Health, New York City, New York, 10032

[2]Cecil G. Sheps Center for Health Services Research, University of North Carolina, Chapel Hill, 27599

[3]Department of Allied Health Sciences, Division of Physical Therapy, School of Medicine, University of North Carolina, Chapel Hill, 27599

[4]Department of Medicine, School of Medicine, University of North Carolina, Chapel Hill, 27599

[5]Department of Psychology, The Ohio State University, Columbus, Ohio 43210

Abstract

Study Design: This is a secondary analysis of a cross-sectional population-based survey.

Objective: Shorten the 23-item Roland scale using item response theory methods and describe where in the functional disability range each scale is the most precise.

Summary of Background Data: The Roland-Morris Disability Questionnaire is recommended for a functional disability outcome measure in low back pain (LBP)patients. One commonly used version is the modified 23-item Roland. It is unknown where in the functional disability range the modified 23-item Roland measures.

Methods: One candidate individual with LBP in randomly selected households was interviewed, identifying 694 adults with chronic LBP. To justify the use of a unidimensional 2-parameter logistic item response theory model, we performed both exploratory and confirmatory factor analysis.

Results: Exploratory factor analysis revealed one dominant eigenvalue. Confirmatory factor analysis results indicate that the one factor model fit well. Item response theory analysis revealed variability in the slopes, ranging from 1.07 to 3.10. The marginal reliability, an item response theory-based analog to coefficient alpha, was 0.88. The modified 23-item Roland produces reliable scores (i.e., with a standard error less than 0.3) from 1.4 standard deviations below the mean to roughly 0.2 above the mean.

Conclusions: The modified 23-item Roland measures one construct. The modified 23-item Roland appears to be an excellent tool for measuring just-below-average levels of functional disability. The modified 23-item Roland measures high levels of functional disability with

Address all manuscript correspondence (including reprint requests) to: Thelma J. Mielenz, PT, PhD, OCS, Assistant Professor, Department of Epidemiology, Mailman School of Public Health, Columbia University, 722 West 168th St. Rm 811, New York, NY 10032, Phone: 212-342-0169, Fax: 212-342-5168, tjm2141@columbia.edu.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

relatively poor reliability and may be more appropriate for a less disabled population with LBP. We demonstrate that the modified 23-item Roland can be shortened to 11 items with minimal loss of information. We show that there are different ways to go about selecting the set of 11 items which yield short forms with different strengths.

Precis

Surveying randomly selected households, we identified 694 adults with chronic low back pain. Item response theory methods determined that the modified 23-item Roland is the most appropriate for measuring just-below-average levels of functional disability and described alternatives for shortening this scale to 11 items with minimal loss of information.

Keywords

item response theory; Modified Roland-Morris Disability; Chronic low back pain

Introduction

Low back pain (LBP) patient-reported outcomes (PROs) are in large part assessed through measurement of functional status. The modified Roland (mRoland) and the Oswestry Disability Index (Oswestry) are recommended by experts and are the most common disease-specific functional disability measures used in LBP patients.[1–5] Over the past ten years, several articles report that the Oswestry may be more precise in patients who are more functionally disabled and the mRoland may be more precise in a population with less functional disability.[3, 5–9] This manuscript focuses only on the mRoland.

The mRoland comes from the Roland-Morris Disability Questionnaire or Roland.[2] The Roland was derived from the Sickness Impact Profile and has been translated into at least 12 languages.[6] The Roland can be completed in five minutes and consists of 24 functional activity limitations due to LBP items.[2]

Seven different versions of the Roland are reported, the two most commonly used being the original 24-item version and the modified 23-item (mRoland) version published by Patrick.[6, 10–11] The mRoland differs from the original by dropping five items and adding four items (i.e. sexual functioning, daily housework, expressing concern, and rubbing or holding the body areas that hurt).[6, 10] The 23 items of the mRoland are listed in Table 1. The 23 items are rated either agree (1) or disagree (0). The items are typically summed so that a higher score represents more disability.[2]

Psychometric analyses of the different versions of the Roland and the mRoland, primarily done in the classical test theory framework, suggest that scores from these modifications (as well as the original) display adequate levels of reliability and validity. Stroud et al. performed an item response theory (IRT) analysis on the original Roland using data from a population with chronic pain with the purpose of shortening the Roland and reported a reliable and valid 11-item version.[11] Two studies used IRT on the mRoland with the primary focus on differential item functioning across gender, education, age, and other.[12, 13] Differential item functioning is not a focus of this manuscript.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

The purpose of the current study is twofold. The main purpose is to address the question posed above as to where in the functional disability range the mRoland measures in community dwelling adults with chronic low back pain and to test the related hypothesis that the mRoland is more precise in a population with less functional disability. The second purpose is to describe how to create a shorter version of the mRoland using IRT methods.

Materials and Methods

This is a secondary analysis of a cross-sectional population-based survey of North Carolinians regarding the prevalence of and care use from LBP.[14] The survey follows up on a 1992 prevalence survey in NC, which found that 3.9% of the adult population had chronic back pain.[15] Carey and colleagues replicated the survey in 2006.

Households were sampled using a stratified probability sample in North Carolina.[14] Telephone numbers were selected by six sampling strata which included regions of the strata and percent of African Americans in the population.[14] The response rates were as follows: 66% for households, 86% for individuals, and 57% for overall. If the household had more than one adult with back pain, then one was randomly selected and interviewed, identifying 712 adults with chronic LBP. Surveying approximately 10,000 individuals, they found that the prevalence of chronic, impairing low back pain had doubled to 10.2% of the population.[14]

LBP was “defined as pain at the level of the waist or below with or without buttock and/or leg pain.” Chronic LBP was defined as 1) report of “pain and activity limitations nearly every day for the past 3 months or 2) more than 24 episodes of pain that limited activity for 1 day or more in the past year”. [14, 15] The mRoland was part of a 30 minute computer-assisted telephone interview of chronic back pain care seeking and functional status.

The methods and results of the 2006 sample selection are published in detail elsewhere including tables of demographics for the larger sample (n=9924) and the demographic and clinical characteristics of this sample with chronic LBP (n=723).[14] This manuscript uses an earlier data set than Freburger et al. with a slightly smaller n (712 versus 723).[14] The chronic LBP sample had a mean age of 53 years (range 21-96), was 62% female and 71% non-Hispanic white and the mean summary mRoland score was 14.9.[14]

IRT Analyses

IRT has become increasingly popular in the assessment of patient-reported outcomes (PROs). The 2-parameter logistic model (2PLM) is an appropriate model to use when the item responses are dichotomous as in the mRoland. The 2PLM describes the probability that an individual will endorse a particular item based on two properties of the item (discrimination and severity) and the underlying level of the construct, functional disability, which the individual possesses. For a more detailed description of the 2PLM see this chapter in *Test Scoring* by Thissen et al.[16] Multilog is used for the IRT analysis.[17] It is widely accepted that a sample size greater than 500 is sufficient for IRT analyses.[18]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

IRT predicts that reliability will vary depending on the score being given and the particular properties of the test being administered. These varying levels of reliability are summarized in two closely related ways: the test information function (TIF) and the standard error curve (SEC). TIF tells where a particular scale provides more (or less) information about respondents. Higher values are better (i.e., more information), but beyond that it is difficult to make meaningful statements about particular amounts of information being “good” or “bad”.

A closely related concept to information is that of a standard error. As information increases, the precision with which a score can be estimated also increases. This results in a decrease in the standard error (which can also be thought of as an increase in reliability). The TIF and the SEC are mathematically related such that the standard error at any particular value of the latent construct is equal to the reciprocal of the square root of the information at that level of the latent construct. When the latent trait being measured is assumed to follow a standard normal distribution (as is typically the case), IRT-based scores and their standard errors are both in a standard normal metric (i.e., a z-score metric).

In addition to the TIF and SEC, MULTILOG also produces a marginal reliability estimate.[19] Marginal reliability is an IRT analog to more commonly used reliability coefficients (e.g., coefficient alpha) that averages over the range of theta to create a one number summary of how reliable scores from a scale will be. This is most useful when reliability is fairly constant over scores. However, as can be seen in Figures 1 and 2, this is not the case in the scales examined here.

EFA and CFA to Test Unidimensionality and Local Dependence for IRT Analysis

An assumption of the 2PLM presented here is that only one latent construct is being measured by the items in question. To justify the use of a unidimensional 2PLM, we performed both exploratory and confirmatory factor analysis (EFA and CFA, respectively). We used the CEFA software package to conduct the EFAs and LISREL for the CFAs.[20–21] EFA was performed on tetrachoric correlations using ordinary least squares (OLS) estimation. When more than one factor was retained, oblique quartimax rotations were used. CFA was also performed on a matrix of tetrachoric correlations using diagonally weighted least squares estimation (DWLS). To evaluate the fit of the one-factor model we used the root mean square error of approximation (RMSEA), comparative fit index (CFI), the goodness of fit index (GFI), and the root mean residual (RMR).[22,23]

The number of participants included in the factor analyses decreased to 604, as listwise deletion was used to deal with missing data for these analyses. The EFA analysis revealed one dominant eigenvalue (13.00) and had only two additional eigenvalues greater than one (1.6 and 1.2). We next fit a one factor CFA model to the 23 items of the mRoland. The CFA results indicate that the one factor model fit well: RMSEA=0.058, CFI=0.99, GFI=0.987, RMR=0.075. The EFA and CFA results both support the assertion that the mRoland is sufficiently unidimensional to proceed with the IRT analyses.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Using IRT to Create Short Forms

An earlier IRT analysis of the original Roland by Stroud et al. (2004) was primarily focused on using IRT to shorten the scale.[11] The method through which the short form was created is not entirely clear, but the resulting 11-item short form (hereafter referred to as Stroud) is described in some detail. To better understand the implications of any efforts to shorten the 23-item mRoland, we created two 11-item versions based on different guiding rules and compared those to the mRoland and the Stroud. We created one 11-item form by choosing the 11 items with the highest slopes (HS11) and the other by trying to evenly space the 11 items across the severity range being measured by these items (BR11).

Results

IRT Analysis

The sample size for the IRT analysis was 670, as participants were not removed due to missing responses. The estimated item parameters are reported in Table 1. Again, in the 2PLM a slope and severity parameter are estimated for each item. Estimated slopes ranged from 1.07 to 3.43 and severity parameters ranged from −2.21 to 1.22.

The TIF and SEC for the mRoland are shown as the short dashed lines in Figures 1 and 2, respectively (we will address the additional lines in those figures below). In general, higher values are better for TIFs and lower values are better for SECs. When viewing TIFs or SECs it is useful to mentally super-impose a standard normal distribution on the X-axis. This reminds us that most scores (68%) are within one standard deviation from the mean and nearly all scores (95%) are within two standard deviations from the mean.

The TIF and SEC for the mRoland suggest that it measures with good reliability (i.e., standard errors <0.3) those individuals who have a level of functional disability between −1.4 and 0.2 (in a standard normal metric). Beyond this range the scores degrade in their reliability quite rapidly. Scores in the higher range of functional disability are measured with less reliability than those at the lower end, which in the SEC is visible in greater increase on the right side than on the left side. For the 23-item mRoland the marginal reliability was 0.88, which is considered good by most standards.

Using IRT to Create Short Forms

Figure 3 contains 2PLM trace lines (known as item characteristic curves) for 11 different items. The Y-axis in these plots is probability of endorsement and the X-axis represents the latent construct being measured (functional disability). A score of zero on the X-axis indicates an individual who has an average level of functional disability for the population from which this sample was drawn. Similarly, a score of 1.5 indicates an individual who is one and a half standard deviations above the average level of functional disability for the population from which this sample was drawn.

The trace lines in Figure 3 help illustrate both differences in slope parameters and differences in severity parameters across items. Four of the items have noticeably shallower curves than the others, indicating that these items have a lower slope. In general, differences

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

in slopes are reflected in the steepness with which the trace lines rise as one looks from left to right in the figure. The severity parameters indicate where the 50% mark on the Y-axis crosses a particular item. This controls the left-to-right shift observed among the items in Figure 3. The further the item is to the left, the lower the severity parameter is (indicating that an item is easy to endorse even for those with very little functional disability). The further an item is to the right, the higher the severity parameter is.

Figure 4 contains 11 trace lines corresponding to the 11 items with the highest slopes from the 23-item mRoland. A comparison with Figure 3 reveals that, although there are some common items between the two 11-item subsets, the breadth of coverage obtained in the "broad" 11-item subset is obtained at the expense of choosing items that are slightly weaker in their association to the construct of interest.

In Figures 1 and 2 we compare the three 11-item short forms (Stroud, HS11, & BR11) with the mRoland. Differences are more pronounced in the information plots, so we will focus there. The first striking feature of Figure 1 is the drop in information from the 23-item mRoland to any of the shorter versions. This is to be expected when shortening a scale, but it is interesting to note some the more subtle differences among the three short forms. The HS11 and Stroud are fairly similar, which supports the impression that the Stroud item selection was primarily based on slopes. Despite their similarity, if one were forced to choose, there is little to recommend the Stroud set of items over the HS11 set. The BR11 was selected to try and represent the range of severities observed in the mRoland items. This short form sacrifices some precision at lower levels of functional disability to achieve greater (but still not much) precision at higher levels of functional disability. There is generally strong agreement between the scores produced by the short form and the scores from the original 23-item mRoland. The HS11 scores correlate 0.95 with the mRoland scores and the BR11 scores correlate 0.96 with the mRoland scores. The HS11 and BR11 scores correlate 0.93.

Discussion

The modified Roland (mRoland) and the Oswestry Disability Index (Oswestry) are recommended by experts and are the most common disease-specific functional disability measures in LBP patients.[2, 5] Over the past 10 years, several articles report that the Oswestry may be more sensitive in patients who are more functionally disabled and the mRoland may be more sensitive in a population with less functional disability.[3,5–9]

Summary of findings

The results of this analysis lend support to these previous findings that the Roland is more appropriate for a less disabled population. In this vein, the mRoland can produce fairly precise scores (i.e., with a standard error less than 0.4) from two standard deviations below the mean to roughly one above the mean. It is apparent from the item parameters, the TIF, and the SEC that the mRoland is not as reliable at measuring higher levels (i.e., greater than 1 standard deviation above the population average) of functional disability from chronic LBP in this community dwelling population.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

When using IRT to aid in shortening a scale, this example demonstrates again the importance of considering the goal of the resulting scale. Here we demonstrated two possible alternative short forms of the mRoland and used TIFs and SECs to convey information about the loss of precision that accompanies the shorter scales. Since the mRoland only takes a few minutes to complete, shortening this scale from 23 to 11 items is probably not enough to justify any loss in precision in the clinical or research setting. Particularly in the clinical setting where preserving the ability to detect change over time is most important. These functional status measures are now being intergrated into the electronic medical records and patient portals.

Limitations

A limitation of this study and Stroud et al.'s is the lack of generalizability to a pure clinical LBP setting. Our study consisted of chronic LBP in a community-dwelling population although a large majority (~84%) sought care in the past year.[14] Stroud et al.'s population consisted of patients from a multidisciplinary pain program presenting with different sites of pain (36% LBP).[11]

Implications for practice

When in the IRT framework, it is thus possible to consider the desired use of a scale and select items to create a short form that maximize the scale's performance with respect to some criterion. Selecting high slope items will yield the most precise measurement, but may not cover the desired range. Selecting items based on location spread may be a more effective way to create a short form, depending on the desired function. For example, a neurosurgery clinic may target precision at a lower level of functional disability and use the Oswestry as an PRO whereas a primary care clinic may target a higher level of functional disability and utilize the mRoland.

Future Directions

The NIH Patient-Reported Outcomes Measurement Information System or PROMIS has a Physical Function item bank and short forms ranging from 4 items to 20 items.[24] As IRT and complementary methods become more widely used, researchers and clinicians may continue to want to use the well-validated mRoland or Owestry as legacy measures of low back pain function. This is supported in a recent NIH Task Force on Research Standards for Chronic Low Back Pain.[25] Another NIH effort called PROsetta Stone, allows clinicians and researchers to transform scores (via score transformation tables) from commonly used legacy measures such as the mRoland and the Owestery to the PROMIS metric. This will potentially perpetuate the continue use of the mROland and Oswestry, disases specific legacy measures (not necessarily a bad thing) compared to the more general physical function measures.[24] Future research should administer both the mRoland and the Oswestry and the PROMIS Physical Functions Short Forms or the PROMIS Physical Function item bank to a wide range of clinical LBP patients to further test this hypothesis of where in the functional disability range each measures covers the most precisely.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Conclusions

The mRoland appears to be an excellent tool for measuring just-below-average levels of functional disability (remembering that average is relative to the population studied here). However, if a researcher is interested in measuring a more severely disabled group, the mRoland does not appear to provide reliable assessment of high levels of functional disability due to LBP. This study advances our knowledge when selecting PROs to measure LBP.

Funding Source:

Grant support for this manuscript includes: a Patient-Centered Outcomes Research Institute (PCORI) Award (1IP2PI000797-01), NIAMS RO-1 AR051970, the Foundation for Physical Therapy *New Investigator Fellowship Training Initiative*, the American College of Rheumatology *Research and Education Foundation Health Professional New Investigator Award*, a National Arthritis Foundation *New Investigator Award*. It was supported in part by Grant 1 R49 CE002096-01 from the National Center for Injury Prevention and Control, Centers for Disease Control and Prevention to the Center for Injury Epidemiology and Prevention at Columbia University. Its contents are solely the responsibility of the authors and do not necessarily represent the official views of the Centers for Disease Control and Prevention.

The following funds were received in support of this work: a Patient-Centered Outcomes Research Institute (PCORI) Award (1IP2PI000797-01), NIAMS RO-1 AR051970, the Foundation for Physical Therapy New Investigator Fellowship Training Initiative, the American College of Rheumatology Research and Education Foundation IRT on Modified Roland-Morris Disability. Health Professional New Investigator Award, a National Arthritis Foundation NewInvestigator Award. It was supported in part by Grant 1 R49 CE002096-01 from theNational Center for Injury Prevention and Control, Centers for Disease Control andPrevention to the Center for Injury Epidemiology and Prevention at Columbia University.Its contents are solely the responsibility of the authors and do not necessarily representhe official views of the Centers for Disease Control and Prevention.

Relevant financial activities outside the submitted work: royalties, payment for lecture, grants/grants pending, board membership, consultancy, employment

Edwards: no relevant

Carey: royalties, payment for lectures, grants, board membership, consultancy, employment

Mielenz: grants

References

(1). Demoulin C, Ostelo R, Knottnerus J, Smeets R. What factors influence the measurement properites of the Roland-Morris disability questionnarie? Eur J Pain 2010 February;14(2):200–6. [PubMed: 19443246]

(2). Ostelo RW, de Vet HC. Clinically important outcomes in low back pain. Best Pract Res Clin Rheumatol 2005 August;19(4):593–607. [PubMed: 15949778]

(3). Bombardier C Outcome assessments in the evaluation of treatment of spinal disorders: summary and general recommendations. Spine 2000 December 15;25(24):3100–3. [PubMed: 11124724]

(4). Resnik L, Dobrykowski E. Outcomes measurement for patients with low back pain. Orthop Nurs 2005 January;24(1):14–24. [PubMed: 15722968]

(5). Grotle M, Brox JI, Vollestad NK. Concurrent comparison of responsiveness in pain and functional status measurements used for patients with low back pain. Spine 2004 November 1;29(21):E492–E501. [PubMed: 15507789]

(6). Roland M, Fairbank J. The Roland-Morris Disability Questionnaire and the Oswestry Disability Questionnaire. Spine 2000 December 15;25(24):3115–24. [PubMed: 11124727]

(7). Carey TS, Mielenz TJ. Measuring outcomes in back care. Spine 2007 May 15;32(11 Suppl):S9–14. [PubMed: 17495590]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

(8). Fairbank JC, Pynsent PB. The Oswestry Disability Index. Spine 2000 November 15;25(22):2940–52. [PubMed: 11074683]

(9). Leclaire R, Blier F, Fortin L, Proulx R. A cross-sectional study comparing the Oswestry and Roland-Morris Functional Disability scales in two populations of patients with low back pain of different levels of severity. Spine 1997 January 1;22(1):68–71. [PubMed: 9122784]

(10). Patrick DL, Deyo RA, Atlas SJ, Singer DE, Chapin A, Keller RB. Assessing health-related quality of life in patients with sciatica. Spine 1995 September 1;20(17):1899–908. [PubMed: 8560339]

(11). Stroud MW, McKnight PE, Jensen MP. Assessment of self-reported physical activity in patients with chronic pain: development of an abbreviated Roland-Morris disability scale. J Pain 2004 June;5(5):257–63. [PubMed: 15219257]

(12). Crane PK, Cetin K, Cook KF, Johnson K, Deyo R, Amtmann D. Differential item functioning impact in a modified version of the Roland-Morris Disability Questionnaire. Qual Life Res 2007 August;16(6):981–90. [PubMed: 17443419]

(13). Pietrobon R, Taylor M, Guller U, Higgins LD, Jacobs DO, Carey T. Predicting gender differences as latent variables: summed scores, and individual item responses: a methods case study. Health Qual Life Outcomes 2004;2:59. [PubMed: 15500700]

(14). Freburger JK, Holmes GM, Agans RP et al. The rising prevalence of chronic low back pain. Arch Intern Med 2009 February 9;169(3):251–8. [PubMed: 19204216]

(15). Carey TS, Evans A, Hadler N, Kalsbeek W, McLaughlin C, Fryer J. Care-seeking among individuals with chronic low back pain. Spine 1995 February 1;20(3):312–7. [PubMed: 7732467]

(16). Thissen D, Nelson L, Rosa K, McLeod L. Item response theory for items scored in two categories. In: Thissen D, Wainer H, editors. Test Scoring. Mahwah, NJ: Lawrence Erlbaum Associate, Inc.; 2001. p. 73–140.

(17). MULTILOG: Multiple, categorical item analysis and test scoring using item response theory [computer program]. Mooresville, IL: Scientific Software; 1991.

(18). Reise SP, Yu J. Parameter recovery in the graded response model using MULTILOG. Journal of Educational Measurement 1990; 27:133–144.

(19). Green BF, Bock RD, Humphreys LG, Linn RL, Reckase MD. Technical guidelines for assessing computerized adaptive tests. J Educational Measurement 1984;21:347–60.

(20). CEFA: Comprehensive exploratory factor analysis [computer program]. Version 2 2004.

(21). LISREL [computer program]. Version 8.7. Chicago, IL: Scientific Software International; 2004.

(22). Bentler PM. Comparative fit indexes in structural models. Psychological Bulletin 2007;107:238–46.

(23). Browne MW, Cudeck R. Alternative ways of assessing model fit. In: Bollen KA, Long JS, editors. Testing Structural Equation Models. Newbury Park, CA: Sage; 1993. p. 136–62.

(24). [NIH Patient Reported Outcomes Measurement System] Available at: http://www.nihpromis.org/. Accessed July 22, 2014.

(25). Deyo RA, Dworkin SF, Amtmann D, Andersson G, Borenstein D, Carragee E, Carrino J, Chou R, Cook K, DeLitto A, Goertz C, Khalsa P, Loeser J, Mackey S, Panagis J, Rainville J, Tosteson T, Turk D, Von Korff M, Weiner DK. Report of the NIH Task Force on Research Standards for Chronic Low Back Pain. Spine J. 2014 Jun 17.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

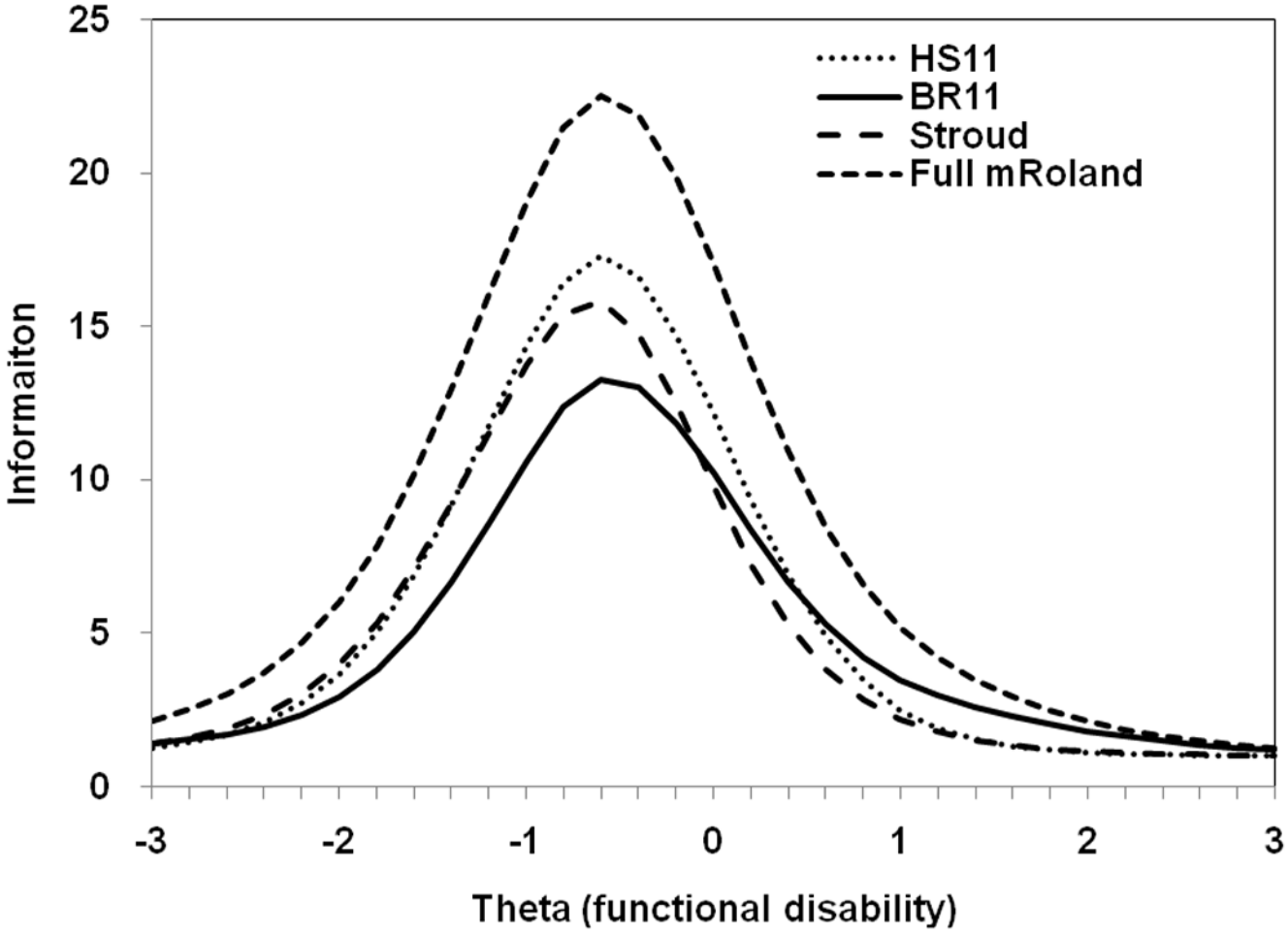

Figure 1.

Test information functions four modified variants: 1) HS11 - 11 items with the highest IRT slope estimates, 2) BR11 - 11 items with the widest spread of severity estimates, 3) Stroud et al (2004) - 11-item short form 4) full mRoland - 23 items in the modified Roland.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

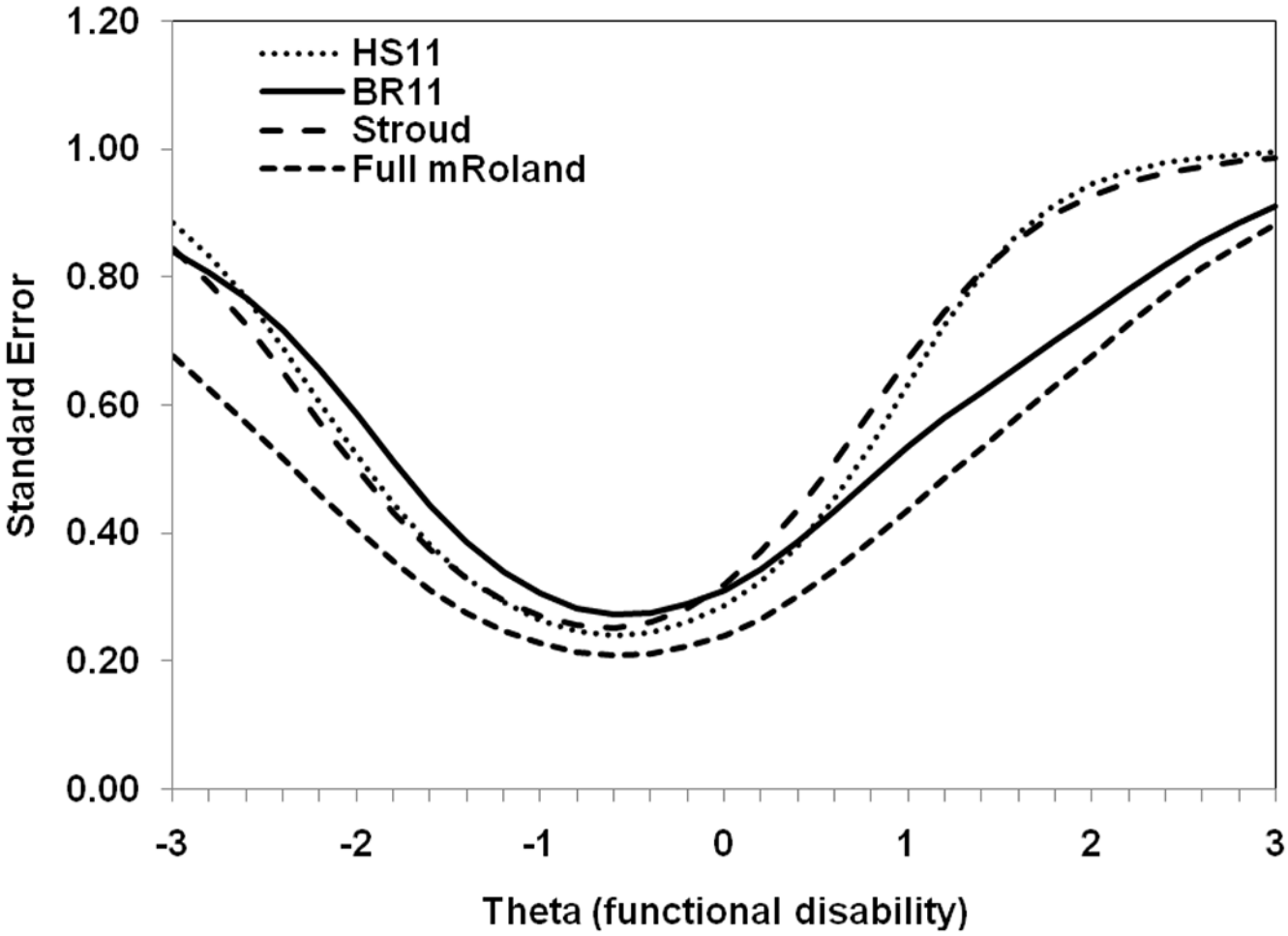

Figure 2.

Standard error curves four modified variants: 1) HS11 - 11 items with the highest IRT slope estimates, 2) BR11 - 11 items with the widest spread of severity estimates, 3) Stroud et al (2004) - 11-item short form 4) full mRoland - 23 items in the modified Roland

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

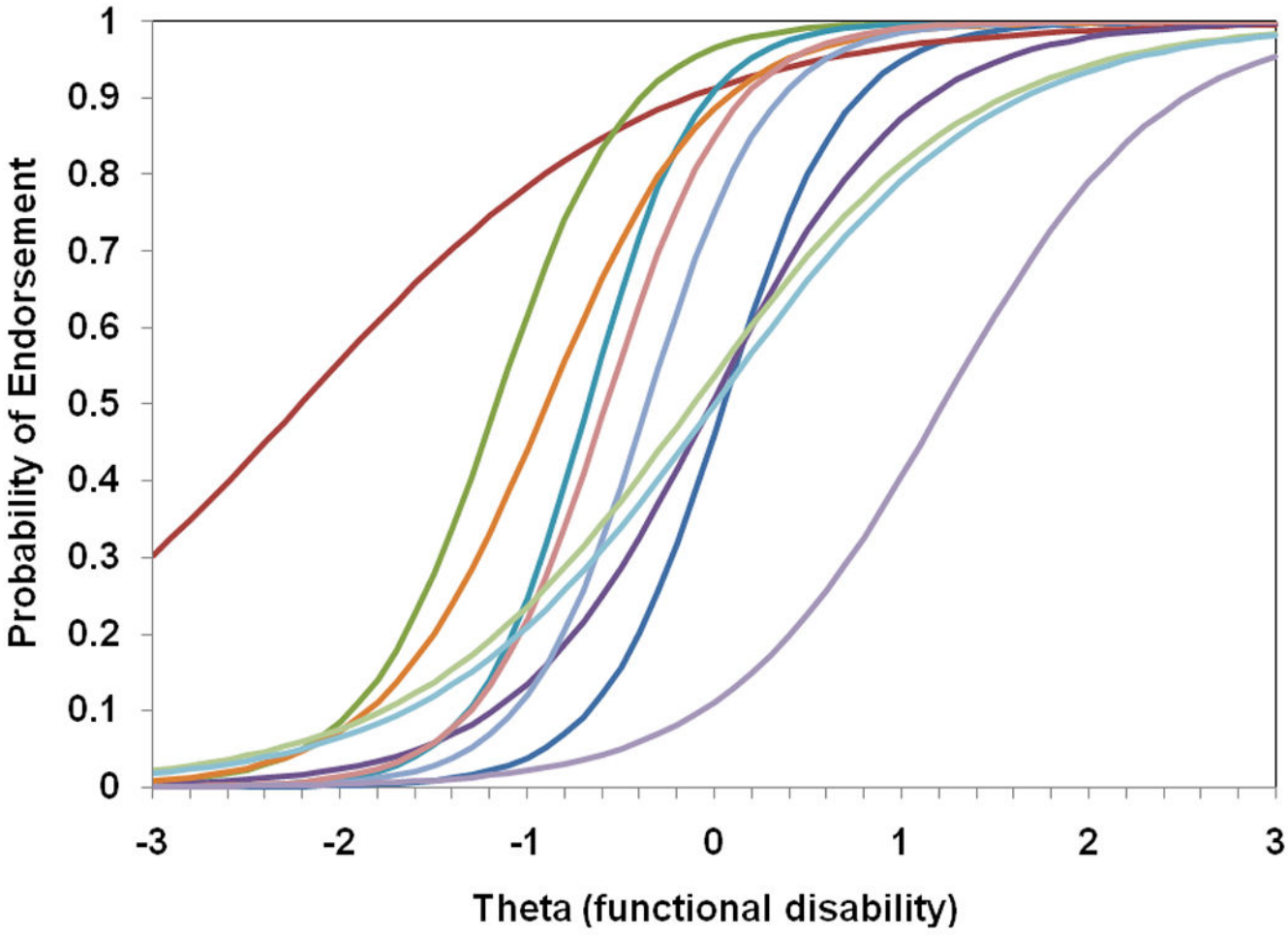

Figure 3.

2PLM trace lines for an 11-item short form of the modified Roland-Morris Disability Questionnaire which was selected to have a broad range of severity estimates (*b*-parameters).

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

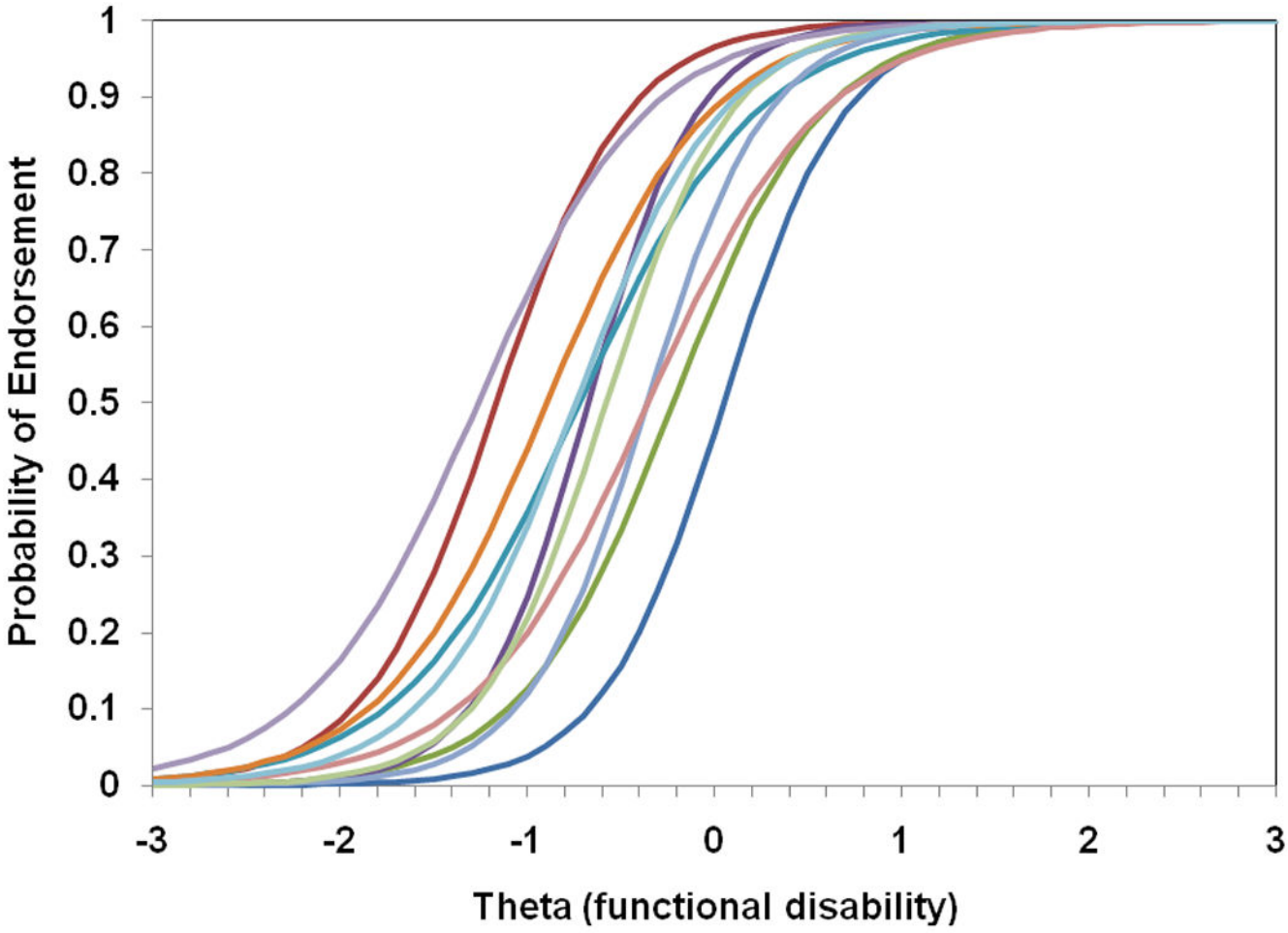

Figure 4.

2PLM trace lines for an 11-item short form of the modified Roland-Morris Disability Questionnaire which was selected by choosing the items with the 11 highest slope estimates (*a*-parameters).

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Table 1.

Items and item parameters from the modified Roland-Morris Disability Questionnaire

Stroud	HS11	BR11	Item Text	A	b
		x	change position frequently	1.07	−2.21
			sleep less well	1.12	−1.42
x	x		avoid heavy jobs around the house	2.22	−1.26
x	x	x	walk more slowly than usual	2.86	−1.17
			painful almost all of the time	1.33	−1.07
x			go up stairs more slowly than usual	1.83	−1.00
x	x	x	try not to bend or kneel down	2.30	−0.90
			rubbing or holding areas of my body that hurt	1.41	−0.84
	x		doing less of the daily work around the house	2.57	−0.74
x	x		only stand for short periods of time	2.11	−0.72
x	x	x	get dressed more slowly than usual	3.43	−0.68
x	x	x	only walk short distances	3.01	−0.58
x			use a handrail to get upstairs	1.20	−0.57
			find it difficult to turn over in bed	1.61	−0.39
x	x	x	find it difficult to get out of a chair	3.10	−0.36
x	x		trouble putting on socks or stockings	2.15	−0.35
			sexual activity is decreased	1.61	−0.35
x	x		hold on to something to get out of an easy chair	2.48	−0.22
		x	more irritable and bad tempered than usual	1.33	−0.11
		x	not doing any of the jobs around the house	1.89	−0.01
		x	express concern about health	1.34	0.00
	x	x	stay at home most of the time	3.07	0.05
		x	Stay in bed most of the time	1.70	1.22

Note. Stroud = 11-item short form reported by Stroud et al. (2004). HS11 = 11-item short form created by choosing the 11 items with the highest IRT slope values. BR11 = 11-item short form created by choosing 11 items with a wide range (relative to the available pool) of IRT severity values. In the first three columns, an “x” indicates that an item is included on that particular short form. IRT slopes (a-parameters) are provided in the column labeled “a” and IRT severity values (b-parameters) are provided in the column labeled “b”.