BRIEF REPORT

Six Patient-Reported Outcome Measurement Information System Short Form Measures Have Negligible Age- or Diagnosis-Related Differential Item Functioning in Individuals With Disabilities

Karon F. Cook, PhD, Alyssa M. Bamer, MPH, Dagmar Amtmann, PhD, Ivan R. Molton, PhD, Mark P. Jensen, PhD

ABSTRACT. Cook KF, Bamer AM, Amtmann D, Molton IR, Jensen MP. Six Patient-Reported Outcome Measurement Information System short form measures have negligible age- or diagnosis-related differential item functioning in individuals with disabilities. Arch Phys Med Rehabil 2012;93:1289-91.

Objective: To evaluate the measurement invariance of 6 self-report measures selected for an ongoing longitudinal study of individuals with spinal cord injury, muscular dystrophy, postpolio syndrome, and multiple sclerosis.

Design: Participants completed and returned by mail surveys that included the targeted self-report measures. Ordinal logistic regressions methods were applied to evaluate items for differential item functioning (DIF) by diagnosis and age range.

Setting: Community.

Participants: Participants (N=2479) who had 1 of the 4 target diagnoses.

Interventions: None.

Main Outcome Measures: Six short-form measures from the Patient-Reported Outcome Measurement Information System (PROMIS) were administered to participants to measure fatigue, pain interference, satisfaction with social roles, sleep disturbance, sleep-related impairment, and depression.

Results: One item of 1 measure (fatigue) exhibited DIF by diagnosis based on a published standard for meaningful DIF. However, scores corrected for this DIF were highly correlated with uncorrected scores ($r>.999$). No DIF by age range was found for any of the measures.

Conclusions: Study findings support the use of the selected PROMIS short forms for comparing symptoms and quality of life indicators across different diagnoses and age ranges.

Key Words: Aging; Disabled persons; Outcome assessment (health care); Psychometrics; Rehabilitation.

CONSISTENT WITH THE general aging of the U.S. population, persons with disabling conditions are living longer. Age-related changes contribute to physical and psychological difficulties over and above any effects of the disability itself. Essential to understanding the interaction between aging and disability is the availability of brief, but reliable and valid measures of the impact of symptoms and quality of life indicators. Of particular importance in studies of the longitudinal impact of aging in disability populations is the confidence that outcome measures perform consistently across diagnostic and age groups, that is, that the scores the measures generate are measurement invariant across diagnosis and age. In an item response theory framework, measurement invariance is evaluated at the item level using methods that assess differential item functioning (DIF).

Recently, we undertook a longitudinal study in samples of persons with spinal cord injury (SCI), muscular dystrophy (MD), postpolio syndrome (PPS), and multiple sclerosis (MS). A major goal of the study is to evaluate the natural course of aging with a disability, especially with respect to secondary health conditions. The purpose of the current analysis was to use baseline data (time 1) to evaluate whether the measures selected for the longitudinal study exhibited DIF with respect to diagnosis or aging.

METHODS

Participants

Individuals with MS, MD, PPS, or SCI were recruited through direct mail invitation from advocacy organizations (eg, the National Multiple Sclerosis Society), advertisements in organization newsletters and websites, and by inviting participants in the university's registry of persons with disabilities and other condition-specific registries (eg, SCI Model Systems). Participants from prior studies who had agreed to be contacted about future studies also were recruited. Eligibility requirements included self-reported diagnosis of MS, MD, PPS, or SCI, ability to read and write English, and written consent. The Institutional Review Board at the University of Washington, Seattle reviewed and approved study procedures. Only responses in the baseline study (time 1) were included for the current analyses.

List of Abbreviations

DIF	differential item functioning
MD	muscular dystrophy
MS	multiple sclerosis
PPS	postpolio syndrome
PROMIS	Patient-Reported Outcome Measurement Information System
SF	short form
SCI	spinal cord injury

From the Department of Medical Social Sciences, Northwestern University, Chicago, IL (Cook); and Department of Rehabilitation Medicine, University of Washington, Seattle, WA (Bamer, Amtmann, Molton, Jensen).

Supported by the Department of Education, National Institute on Disability and Rehabilitation Research (grant nos. H133B080024 and H133B080025). However, these contents do not necessarily represent the policy of the Department of Education, and one should not assume endorsement by the Federal Government.

No commercial party having a direct financial interest in the results of the research supporting this article has or will confer a benefit on the authors or on any organization with which the authors are associated.

Correspondence to Karon F. Cook, PhD, Dept of Medical Social Sciences, Northwestern University Feinberg School of Medicine, Faculty, Dept of Psychology, 625 N Michigan Ave, Ste 2700, Chicago, IL 60611, e-mail: *karon.cook@northwestern.edu*. Reprints are not available from the author.

In-press corrected proof published online on Mar 2, 2012, at www.archives-pmr.org.

0003-9993/12/9307-00918$36.00/0
doi:10.1016/j.apmr.2011.11.022

Table 1: Clinical and Demographic Characteristics of Study Participants (N=2479)

Characteristics	Values
Age (y)	55.5±12.8
Duration of disease (y)	15.5±10.4
Women	68.1
Education level	
≤High school/general equivalency degree	14.1
Technical/some college	33.0
Bachelor's degree	29.8
Graduate degree	23.1
Married or living with partner	64.8
Race/ethnicity*	
White	94.6
Hispanic/Chicano	1.6
Other	6.0
Paid employment	29.7
Disease type	
MS	48.4
MD	13.7
PPS	18.0
SCI	19.9

NOTE. Values are mean ± SD or percent.
*More than 1 response allowed.

Measures

Short forms (SFs) developed by the Patient-Reported Outcome Measurement Information System (PROMIS) were used to measure symptoms and quality of life indicators. PROMIS is a National Institutes of Health-sponsored project to develop item banks and standard measures that assess key symptoms and outcomes applicable to a range of chronic conditions.[1] Surveys were developed that included PROMIS SFs to measure: pain interference,[2] sleep disturbance and sleep-related impairment,[3] satisfaction with social roles,[4] depression,[5] and fatigue.[6] (Note: Copies of SFs are available from the PROMIS Assessment Center at http://www.nihpromis.org.) Surveys were mailed to participants who completed and returned them in a postage-paid envelope. The PROMIS SFs used in the current study have a 7-day time frame and 5 response categories. Higher scores indicate more of the trait for which the measure is named (eg, more depression, more satisfaction with social roles). PROMIS scores are calibrated to a T score metric where 50 is the mean and 10 is the SD of a census-weighted sample of the general United States population.[1]

Methods

DIF analyses. The software package, lordif,[7,a] was used to evaluate DIF. Lordif assesses DIF using an ordinal logistic regression framework. In these analyses, the dependent variable was the item response (1–5), and the predictor was the level of the trait being measured. For example, consider the PROMIS Fatigue SF. When the items of this SF were evaluated for DIF, a base model (model 1) was posited in which only the level of fatigue predicted people's responses. A second model (model 2) was posited in which both level of fatigue and age group predicted responses. If model 2 predicts item responses statistically significantly better than model 1, then DIF has been identified. A third model (model 3) includes an interaction

Table 2: Sample Sizes and Scores by Age Group, Diagnosis, and Measure

Sample Sizes and Scores	Fatigue	Pain Interference	Satisfaction With Social Roles	Sleep Disturbance	Sleep-Related Impairment	Depression
By diagnosis						
MD						
n	337	171	335	337	165	174
$\bar{X}$ ± SD	56.1±8.2	54.3±8.8	45.5±9.3	51.8±9.1	52.0±9.6	53.0±9.0
MS						
n	1193	794	572	1194	908	899
$\bar{X}$ ± SD	58.2±8.4	56.4±8.4	45.1±8.7	52.4±10.1	53.3±9.5	51.9±9.3
PPS						
n	441	218	432	445	218	225
$\bar{X}$ ± SD	58.7±7.2	58.0±7.9	43.8±7.0	51.6±9.1	49.7±8.7	53.6±7.7
SCI						
n	478	239	483	490	248	243
$\bar{X}$ ± SD	52.4±7.7	56.8±8.3	45.5±8.1	52.6±9.7	49.8±9.7	53.9±8.8
By age group						
0–44y						
n	462	266	349	463	289	289
$\bar{X}$ ± SD	55.1±9.5	54.4±8.9	47.3±9.2	52.8±10.1	53.6±10.5	52.7±9.3
45–54y						
n	618	378	414	616	407	409
$\bar{X}$ ± SD	60.0±8.8	57.4±7.8	45.2±8.4	53.6±9.9	53.2±9.5	52.7±9.4
55–64y						
n	802	488	587	803	494	518
$\bar{X}$ ± SD	57.7±7.7	56.8±8.6	44.5±7.9	52.8±9.5	52.3±9.4	52.9±8.7
≥65y						
n	584	309	493	588	351	327
$\bar{X}$ ± SD	57.0±7.6	56.6±8.1	43.8±7.8	49.6±9.2	49.3±8.5	51.8±8.8

NOTE. Sample sizes vary depending on incomplete data and depending on which survey version was completed by each participant.

term and tests whether DIF has a consistent impact across trait levels (uniform DIF) or if the impact of DIF varies by trait level (nonuniform DIF).

A known issue with identifying DIF based on statistical criteria is that chi-square values are heavily influenced by sample size. Alternative standards have been suggested for identifying negligible DIF. For the current study we used the following criteria for meaningful DIF: (1) change in pseudo R^2 statistic $<.13$[8] and (2) $<5\%$ change in β coefficients comparing model 1 (no group effect) with model 2 (group effect).[9]

For DIF analyses by diagnosis, participants were grouped into the 4 diagnostic categories (MS, PPS, SCI, MD). For DIF analyses by age they were grouped into 4 age ranges: <44, 45–54, 55–64, and ≥65 years of age. These age groupings were chosen to maximize comparisons and minimize sparse cells (N<200). Because the study targeted aging with a disability, few younger persons were included. Analyses were conducted within domain (eg, fatigue or pain interference) but across disability groups to ensure sufficient sample size in each age comparison group.

We tested the omnibus hypothesis that no items of a given SF had meaningful DIF by diagnosis/age. If items with meaningful DIF were identified, scores were corrected for DIF by developing age- or diagnosis-specific item parameters and rescoring. Post hoc analyses evaluated the impact of DIF on individual item scores.

RESULTS

Table 1 presents clinical and demographic characteristics of the sample.

DIF by Diagnosis

In the omnibus tests of DIF by diagnosis, only 1 item of 1 SF (PROMIS Fatigue) was identified as having DIF based on 1 criterion ($<5\%$ change in β coefficients). After conditioning on the fatigue level, the probability of different responses to the item, "How often were you too tired to take a bath or shower" varied slightly by diagnosis (nonuniform DIF). Diagnosis-specific item parameters were calculated for this item and DIF-corrected scores were estimated. Scores before and after DIF correction were highly associated ($r>.999$), and differences between corrected and uncorrected means on PROMIS Fatigue SF were $<.02$ SDs.

DIF by Age Group

Based on study criteria, no items of any PROMIS SF were identified as having meaningful DIF by age group. Therefore, no post hoc analyses of impact on scores were conducted.

Table 2 presents sample sizes and scores by diagnosis and age group for each SF.

DISCUSSION

The PROMIS SFs used in a longitudinal study to evaluate the natural course of aging with a disability do not exhibit DIF with respect to age. Although 1 PROMIS Fatigue SF item met criteria for meaningful DIF, impact on respondents' scores was negligible.

Study Limitations

Our samples reflected the demographics of the Seattle geographic area and are not representative of the general population of persons with MS, SCI, PPS, or MD.

CONCLUSIONS

These findings support the use of the selected PROMIS SFs for comparing symptoms and quality of life indicators across different diagnoses and age ranges in the current study. Future studies should evaluate the measures in more representative samples of persons with these disabilities, especially with respect to education and race/ethnicity.

References

1. Cella D, Riley W, Stone A, et al. The Patient-Reported Outcomes Measurement Information System (PROMIS) developed and tested its first wave of adult self-reported health outcome item banks: 2005-2008. J Clin Epidemiol 2010;63:1179-94.
2. Amtmann D, Cook KF, Jensen MP, et al. Development of a PROMIS item bank to measure pain interference. Pain 2010;150: 173-82.
3. Flynn KE, Shelby RA, Mitchell SA, et al. Sleep-wake functioning along the cancer continuum: focus group results from the Patient-Reported Outcomes Measurement Information System (PROMIS(®)). Psychooncology 2010;19:1086-93.
4. Hahn EA, Devellis RF, Bode RK, et al. Measuring social health in the Patient-Reported Outcomes Measurement Information System (PROMIS): item bank development and testing. Qual Life Res 2010;19:1035-44.
5. Pilkonis PA, Choi SW, Reise SP, Stover AM, Riley WT, Cella D. Item banks for measuring emotional distress from the Patient-Reported Outcomes Measurement Information System (PROMIS ®): depression, anxiety, and anger. Assessment 2011;18:263-83.
6. Lai JS, Cella D, Choi S, et al. How item banks and their applications can influence measurement practice in rehabilitation medicine: a PROMIS fatigue item bank example. Arch Phys Med Rehabil 2011;92(10 Suppl):S20-7.
7. Choi SW, Gibbons LE, Crane PK. Lordif: an R package for detecting differential item functioning using iterative hybrid ordinal logistic regression/item response theory and Monte Carlo simulations. J Stat Softw 2011;39:1-30.
8. Zumbo BD. A handbook on the theory and methods of differential item functioning (DIF): logistic regression modeling as a unitary framework for binary and Likert-type (ordinal) item scores. Ottawa: Directorate of Human Resources Research and Evaluation, Department of National Defense; 1999.
9. Crane PK, Gibbons LE, Ocepek-Welikson K, et al. A comparison of three sets of criteria for determining the presence of differential item functioning using ordinal logistic regression. Qual Life Res 2007;16(Suppl 1):69-84.

Supplier

a. Lordif; freeware available at: http://cran.r-project.org/web/packages/lordif/lordif.pdf.