2016 Volume 58 Part 1

Volume 58 | Issue 1 - 2016

Psychological Test and Assessment Modeling

ISSN 2190-0507 (Internet)

formerly Psychologische Beiträge / Psychology Science Quarterly

K. D. Kubinger: Editorial: Review and forecast on research in Psychological Test and Assessment Modeling

D. B. Wright: Treating all rapid responses as errors (TARRE) improves estimates of ability (slightly)

B. B. Reeve, J. A. Teresi: Overview to the two-part series: Measurement equivalence of the Patient-Reported Outcomes Measurement Information System[®] (PROMIS[®]) short-forms

J. A. Teresi, R. N. Jones: Methodological issues in examining measurement equivalence in Patient Reported Outcomes Measures: Methods overview to the two-part series, "Measurement equivalence of the Patient Reported Outcomes Measurement Information System[®] (PROMIS[®]) short forms"

M. Kleinman, J. A. Teresi: Differential item functioning magnitude and impact measures from item response theory models

R. E. Jensen, C. M. Moinpour, T. H. M. Keegan, R. D. Cress, X.-Cheng Wu, L. E. Paddock, A. M. Stroup, A. L. Potosky: The Measuring Your Health study: Leveraging community-based cancer registry recruitment to establish a large, diverse cohort of cancer survivors for analyses of measurement equivalence and validity of the Patient Reported Outcomes Measurement Information System[®] (PROMIS[®]) short form items

B. B. Reeve, L. C. Pinheiro, R. E. Jensen, J. A. Teresi, A. L. Potosky, M. K. McFatrich, M. Ramirez, W.-H. Chen: Psychometric evaluation of the PROMIS[®] Fatigue measure in an ethnically and racially diverse populationbased sample of cancer patients

J. A. Teresi, K. Ocepek-Welikson, M. Kleinman, M. Ramirez, G. Kim: Psychometric properties and performance of the Patient Reported Outcomes Measurement Information System[®] (PROMIS[®]) Depression short forms in ethnically diverse groups

J. A. Teresi, K. Ocepek-Welikson, M. Kleinman, M. Ramirez, G. Kim: Measurement equivalence of the Patient Reported Outcomes Measurement Information System[®] (PROMIS[®]) Anxiety short forms in ethnically diverse groups

SPECIAL TOPIC: MEASUREMENT EQUIVALENCE OF THE PATIENT REPORTED OUTCOMES MEASUREMENT INFORMATION SYSTEM® (PROMIS®) SHORT FORMS – PART I GUEST EDITORS: BRYCE B. REEVE & JEANNE A. TERESI

Editorial:

Review and forecast on research in Psychological Test and Assessment Modeling Klaus D. Kubinger (editor in chief) $\stackrel{\leftarrow}{\doteq}$ PDF of the full article

Treating all rapid responses as errors (TARRE) improves estimates of ability (slightly) Daniel B. Wright <u>Abstract</u> | <u>DF of the full article</u>

Overview to the two-part series: Measurement equivalence of the Patient Reported Outcomes Measurement Information System[®] (PROMIS[®]) short forms Bryce B. Reeve & Jeanne A. Teresi Abstract | $\stackrel{!}{ riangle}$ PDF of the full article

Methodological issues in examining measurement equivalence in patient reported outcomes measures: Methods overview to the two-part series, "Measurement equivalence of the Patient Reported Outcomes Measurement Information System^{*} (PROMIS^{*}) short forms" Jeanne A. Teresi & Richard N. Jones Abstract | idot PDF of the full article

Differential item functioning magnitude and impact measures from item response theory models *Marjorie Kleinman & Jeanne A. Teresi* <u>Abstract</u> | <u>in PDF of the full article</u>

The Measuring Your Health study: Leveraging community-based cancer registry recruitment to establish a large, diverse cohort of cancer survivors for analyses of measurement equivalence and validity of the Patient Reported Outcomes Measurement Information System[®] (PROMIS[®]) short form items *Roxanne E. Jensen, Carol M. Moinpour, Theresa H.M. Keegan, Rosemary D. Cress, Xiao-Cheng Wu, Lisa E. Paddock, Antoinette M. Stroup & Arnold L. Potosky* Abstract | ightarrow PDF of the full article

Psychometric Evaluation of the PROMIS[®] Fatigue measure in an ethnically and racially diverse population-based sample of cancer patients Bryce B. Reeve, Laura C. Pinheiro, Roxanne E. Jensen, Jeanne A. Teresi, Arnold L. Potosky, Molly K. McFatrich, Mildred Ramirez & Wen-Hung Chen Abstract | interprotect processing PDF of the full article

Psychometric properties and performance of the Patient Reported Outcomes Measurement Information System[®] (PROMIS[®]) depression short forms in ethnically diverse groups Jeanne A. Teresi, Katja Ocepek-Welikson, Marjorie Kleinman, Mildred Ramirez & Giyeon Kim Abstract | <u>PDF of the full article</u>

Measurement equivalence of the Patient Reported Outcomes Measurement Information System[®] (PROMIS[®]) Anxiety short forms in ethnically diverse groups *Jeanne A. Teresi, Katja Ocepek-Welikson, Marjorie Kleinman, Mildred Ramirez & Giyeon Kim* <u>Abstract</u> | <u>DF of the full article</u>



Volume 58 | Issue 2 - 2016

ISSN 2190-0493 (Print) SN 2190-0507 (Internet)

Psychological Test and Assessment Modeling

formerly Psychologische Beiträge / Psychology Science Quarterly

K. Haberkorn, S. Pohl & C. H. Carstensen: Incorporating different response formats of competence tests in an IRT model

D. Andrich: In memoriam Benjamin Wright (1926 - 2015)

R. Fieo, K. Ocepek-Welikson, M. Kleinman, J. P. Eimicke, P. K. Crane, D. Cella & J. A. Teresi: Measurement equivalence of the Patient Reported Outcomes Measurement Information System[®] (PROMIS[®]) Applied Cognition – General Concerns, short forms in ethnically diverse groups

J. A. Teresi, K. Ocepek-Welikson, K. F. Cook, M. Kleinman, M. Ramirez, M. Carrington Reid & A. Siu: Measurement equivalence of the Patient Reported Outcomes Measurement Information System[®] (PROMIS[®]) Pain Interference short form items: Application to ethnically diverse cancer and palliative care populations

R. E. Jensen, B. L. King-Kallimanis, E. Sexton, B. B. Reeve, C. M. Moinpour, A. L. Potosky, T. Lobo & J. A. Teresi: Measurement properties of PROMIS Sleep Disturbance short forms in a large, ethnically diverse cancer cohort

R. N. Jones, D. Tommet, M. Ramirez, R. Jensen & J. A. Teresi: Differential item functioning in Patient Reported Outcomes Measurement Information System® (PROMIS®) Physical Functioning short forms: Analyses across ethnically diverse groups

E. A. Hahn, M. A. Kallen, R. E. Jensen, A. L. Potosky, C. M. Moinpour, M. Ramirez, D. Cella & J. A. Teresi: Measuring social function in diverse cancer populations: Evaluation of measurement equivalence of the Patient Reported Outcomes Measurement Information System® (PROMIS®) Ability to Participate in Social Roles and Activities short form

J. A. Teresi & B. B. Reeve: Epilogue to the two-part series: Measurement equivalence of the Patient Reported Outcomes Measurement Information System® (PROMIS®) short forms

PABST SCIENCE PUBLISHERS . Lengerich

SPECIAL TOPIC: MEASUREMENT EQUIVALENCE OF THE PATIENT REPORTED OUTCOMES MEASUREMENT INFORMATION SYSTEM® (PROMIS®) SHORT FORMS – PART II GUEST EDITORS: BRYCE B. REEVE & JEANNE A. TERESI

Incorporating different response formats of competence tests in an IRT model Kerstin Haberkorn, Steffi Pohl & Claus H. Carstensen Abstract | DF of the full article

In memoriam Benjamin Wright (1926 – 2015) David Andrich <u>→ PDF of the full article</u>

Measurement equivalence of the Patient Reported Outcomes Measurement Information System^{*} (PROMIS^{*}) Applied Cognition – General Concerns, short forms in ethnically diverse groups *Robert Fieo, Katja Ocepek-Welikson, Marjorie Kleinman, Joseph P. Eimicke, Paul K. Crane, David Cella & Jeanne A. Teresi* <u>Abstract</u> | <u>UPDF of the full article</u>

Measurement equivalence of the Patient Reported Outcomes Measurement Information System[®] (PROMIS[®]) Pain Interference short form items: Application to ethnically diverse cancer and palliative care populations Jeanne A. Teresi, Katja Ocepek-Welikson, Karon F. Cook, Marjorie Kleinman, Mildred Ramirez, M. Carrington Reid & Albert Siu

Abstract | 🗄 PDF of the full article

Measurement properties of PROMIS Sleep Disturbance short forms in a large, ethnically diverse cancer cohort *Roxanne E. Jensen, Bellinda L. King-Kallimanis, Eithne Sexton, Bryce B. Reeve, Carol M. Moinpour, Arnold L. Potosky, Tania Lobo & Jeanne A. Teresi* Abstract | i PDF of the full article

Differential item functioning in Patient Reported Outcomes Measurement Information System^{*} (PROMIS^{*}) Physical Functioning short forms: Analyses across ethnically diverse groups *Richard N. Jones, Doug Tommet, Mildred Ramirez, Roxanne Jensen & Jeanne A. Teresi* <u>Abstract</u> | <u>LPDF of the full article</u>

Measuring social function in diverse cancer populations: Evaluation of measurement equivalence of the Patient Reported Outcomes Measurement Information System[®] (PROMIS[®]) Ability to Participate in Social Roles and Activities Short Form

Elizabeth A. Hahn, Michael A. Kallen, Roxanne E. Jensen, Arnold L. Potosky, Carol M. Moinpour, Mildred Ramirez, David Cella & Jeanne A. Teresi

Abstract | 🗄 PDF of the full article

Epilogue to the two-part series: Measurement equivalence of the Patient Reported Outcomes Measurement Information System[®] (PROMIS[®]) short forms Jeanne A. Teresi & Bryce B. Reeve Abstract | $\stackrel{!}{\sqcup}$ PDF of the full article

ABSTRACTS

Overview to the two-part series: Measurement equivalence of the Patient Reported Outcomes Measurement Information System[®] (PROMIS[®]) short forms *Bryce B. Reeve & Jeanne A. Teresi*

Abstract

Measurement equivalence across differing socio-demographic groups is essential for valid assessment. This is one of two issues of Psychological Test and Assessment Modeling that contains articles describing methods and substantive findings related to establishing measurement equivalence in self-reported health, mental health and social functioning measures.

The articles in this two part series describe analyses of items assessing eight domains: fatigue, depression, anxiety, sleep, pain, physical function, cognitive concerns and social function. Additionally, two overview articles describe the methods and sample characteristics of the data set used in these analyses. An additional article describes the important topic of assessing magnitude and impact of differential item functioning. These articles provide the first strong evidence supporting the measurement equivalence of the Patient Reported Outcomes Measurement Information System[®] (PROMIS[®]) short form measures in ethnically, socio-demographically diverse groups, and is a beginning step in meeting the international call for further study of their performance in such groups.

Key words: PROMIS, short form measures, patient-reported outcomes, measurement equivalence, differential item functioning

Bryce Reeve, PhD Department of Health Policy and Management Gillings School of Global Public Health University of North Carolina at Chapel Hill 1101-D McGavran-Greenberg Hall 135 Dauer Drive CB 7411, Chapel Hill NC 27599-7411, USA bbreeve@email.unc.edu

Methodological issues in examining measurement equivalence in patient reported outcomes measures: Methods overview to the two-part series, "Measurement equivalence of the Patient Reported Outcomes Measurement Information System[®] (PROMIS[®]) short forms" *Jeanne A. Teresi & Richard N. Jones*

Abstract

The purpose of this article is to introduce the methods used and challenges confronted by the authors of this two-part series of articles describing the results of analyses of measurement equivalence of the short form scales from the Patient Reported Outcomes Measurement Information System^{*} (PROMIS^{*}). Qualitative and quantitative approaches used to examine differential item functioning (DIF) are reviewed briefly. Qualitative methods focused on generation of DIF hypotheses. The basic quantitative approaches used all rely on a latent variable model, and examine parameters either derived directly from item response theory (IRT) or from structural equation models (SEM). A key methods focus of these articles is to describe state-of-the art approaches to examination of measurement equivalence in eight domains: physical health, pain, fatigue, sleep, depression, anxiety, cognition, and social function. These articles represent the first time that DIF has been examined systematically in the PROMIS short form measures, particularly among ethnically diverse groups. This is also the first set of analyses to examine the performance of PROMIS short forms in patients with cancer.

Latent variable model state-of-the-art methods for examining measurement equivalence are introduced briefly in this paper to orient readers to the approaches adopted in this set of papers. Several methodological challenges underlying (DIF-free) anchor item selection and model assumption violations are presented as a backdrop for the articles in this two-part series on measurement equivalence of PROMIS measures.

Key Words: methods, PROMIS, measurement equivalence, differential item functioning, ethnic diversity

Jeanne A. Teresi, Ed.D, Ph.D. Columbia University Stroud Center at New York State Psychiatric Institute 1051 Riverside Drive, Box 42, Room 2714, New York New York, 10032-3702, USA <u>Teresimeas@aol.com; jat61@columbia.edu</u>

Differential item functioning magnitude and impact measures from item response theory models *Marjorie Kleinman & Jeanne A. Teresi*

Abstract

Measures of magnitude and impact of differential item functioning (DIF) at the item and scale level, respectively are presented and reviewed in this paper. Most measures are based on item response theory models. Magnitude refers to item level effect sizes, whereas impact refers to differences between groups at the scale score level. Reviewed are magnitude measures based on group differences in the expected item scores and impact measures based on differences in the expected scale scores. The similarities among these indices are demonstrated. Various software packages are described that provide magnitude and impact measures, and new software presented that computes all of the available statistics conveniently in one program with explanations of their relationships to one another.

Key words: differential item functioning, magnitude, impact, item response theory

Marjorie Kleinman, M.S. New York State Psychiatric Institute 1051 Riverside Drive, Unit 72 New York, NY, 10032, USA <u>kleinmam@nyspi.columbia.edu</u>

The Measuring Your Health study: Leveraging community-based cancer registry recruitment to establish a large, diverse cohort of cancer survivors for analyses of measurement equivalence and validity of the Patient Reported Outcomes Measurement Information System[®] (PROMIS[®]) short form items

Roxanne E. Jensen, Carol M. Moinpour, Theresa H.M. Keegan, Rosemary D. Cress, Xiao-Cheng Wu, Lisa E. Paddock, Antoinette M. Stroup & Arnold L. Potosky

Abstract

The Measuring Your Health (MY-Health) study was designed to fill evidence gaps by validating eight Patient Reported Outcomes Measurement Information System^{*}(PROMIS^{*}) domains (Anxiety, Depression, Fatigue, Pain Interference, Physical Function, Sleep Disturbance, Applied Cognitive Function, and Ability to Participate in Social Roles and Activities) across multiple race-ethnic and age groups in a diverse cohort of cancer patients. This paper provides detailed information on MY-Health study design, implementation, and participant cohort; it identifies key challenges and benefits of recruiting a diverse community-based cancer cohort. Between 2010 and 2012, we identified eligible patients for the MY-Health study in partnership with four Surveillance, Epidemiology, and End Results (SEER) program cancer registries located in California, Louisiana, and New Jersey. The overall response rate for the MY-Health cohort (n = 5,506) was 34 %, with a median response time of 9.5 months after initial cancer diagnosis. The cohort represented meaningful diversity of age (22 % under 49 years of age) and race/ethnicity (41 % non-Hispanic White) across seven cancers. Challenges included lower response rates by race/ethnic minorities, young, and advanced-stage cancer patients, use of non-final registry information for eligibility identification, and lower use of translated surveys than expected. The MY-Health cohort represents one of the largest efforts to measure the full range of patient-reported symptoms experienced after initial cancer treatment. It provides sufficient diversity in terms of sociodemographics, symptoms, and function to provide a meaningful validation of eight PROMIS measures.

Key words: PROMIS, MY-Health, measurement equivalence, validity, cancer

Roxanne E. Jensen, Ph.D. Cancer Prevention and Control Program Lombardi Comprehensive Cancer Center Georgetown University 3300 Whitehaven Street NW, Suite 4100 Washington, DC 20007, USA rj222@georgetown.edu

Psychometric evaluation of the PROMIS[®] Fatigue measure in an ethnically and racially diverse population-based sample of cancer patients

Bryce B. Reeve, Laura C. Pinheiro, Roxanne E. Jensen, Jeanne A. Teresi, Arnold L. Potosky, Molly K. McFatrich, Mildred Ramirez & Wen-Hung Chen

Abstract

Aims: Fatigue is the most prevalent and distressing symptom related to cancer and its treatment affecting functioning and quality of life. In 2010, the National Cancer Institute's Clinical Trials Planning Meeting on cancer-related fatigue adopted the PROMIS[®] Fatigue measure as the standard to use in clinical trials. This study evaluates the psychometric properties of the PROMIS Fatigue measure in an ethnically/racially diverse population-based sample of adult cancer patients.

Methods: Patients were recruited from four US cancer registries with oversampling of minorities. Participants completed a paper survey 6 - 13 months post-diagnosis. The 14 fatigue items (5-point Likert-type scale; English-, Spanish-, and Chinese-versions) were selected from the PROMIS Fatigue short forms and larger item bank. Item response theory and factor analyses were used to evaluate item- and scale-level performance. Differential item functioning (DIF) was evaluated using the Wald test and ordinal logistic regression (OLR) methods. OLR-identified items with DIF were evaluated further for their effect on the scale scores (threshold r2 > .13).

Results: The sample included 5,507 patients (2,278 non-Hispanic Whites, 1,122 non-Hispanic Blacks, 1,053 Hispanics, and 917 Asians/ Pacific Islanders); 338 Hispanics were given the Spanish-language version of the survey and 134 Asians the Chinese version. One PROMIS item had poor discrimination as it was the only positively worded question in the fatigue measure. Among Hispanics, no DIF was found with the Wald test, while the OLR method identified five items with DIF comparing the English and Spanish versions; however, the effect of DIF on scores was negligible (r2 ranged from .006 - .015). For the English and Chinese translations, no single item was consistently identified by both DIF tests. Minimal or no impact was observed on the overall scale score comparisons among Whites, Blacks, Hispanics, and Asians using the English language scales. However, greater numbers of items with DIF appeared when comparing Asians/ Pacific Islanders with Whites, Blacks, and Hispanics. "How often were you too tired to think clearly" showed consistent DIF. Conclusions: Twelve of 14 PROMIS fatigue items performed well across the ethnically/racially diverse samples with minimal findings of DIF that would have any effect on comparing or combining scores across cancer populations. Supporting evidence of the validity and reliability of the PROMIS measures will enhance the adoption of the measures in oncology clinical research.

Keywords: differential item functioning, cancer, PROMIS, fatigue, patient-reported outcomes

Bryce Reeve, PhD Department of Health Policy and Management Gillings School of Global Public Health University of North Carolina at Chapel Hill 1101-D McGavran-Greenberg Hall 135 Dauer Drive CB 7411, Chapel Hill NC 27599-7411, USA bbreeve@email.unc.edu

Psychometric properties and performance of the Patient Reported Outcomes Measurement Information System[®] (PROMIS[®]) Depression short forms in ethnically diverse groups *Jeanne A. Teresi, Katja Ocepek-Welikson, Marjorie Kleinman, Mildred Ramirez & Giyeon Kim*

Abstract

Short form measures from the Patient Reported Outcomes Measurement Information System[®] (PROMIS[®]) are used widely. The present study was among the first to examine differential item functioning (DIF) in the PROMIS Depression short form scales in a sample of over 5000 racially/ethnically diverse patients with cancer. DIF analyses were conducted across different racial/ethnic, educational, age, gender and language groups.

Methods: DIF hypotheses, generated by content experts, informed the evaluation of the DIF analyses. The graded item response theory (IRT) model was used to evaluate the five-level ordinal items. The primary tests of DIF were Wald tests; sensitivity analyses were conducted using the IRT ordinal logistic regression procedure. Magnitude was evaluated using expected item score functions, and the non-compensatory differential item functioning (NCDIF) and T1 indexes, both based on group differences in the item curves. Aggregate impact was evaluated with expected scale score (test) response functions; individual impact was assessed through examination of differences in DIF adjusted and unadjusted depression estimates.

Results: Many items evidenced DIF; however, only a few had slightly elevated magnitude. No items evidenced salient DIF with respect to NCDIF and the scale-level impact was minimal for all group comparisons. The following short form items might be targeted for further study because they were also hypothesized to evidence DIF. One item showed slightly higher magnitude of DIF for age: nothing to look forward to; conditional on depression, this item was more likely to be endorsed in the depressed direction by individuals in older groups as contrasted with the cohort aged 21 to 49. This item was also hypothesized to show age DIF. Only one item (failure) showed DIF of slightly higher magnitude (just above threshold) for Whites vs. Asians/Pacific Islanders in the direction of higher likelihood of endorsement for Asians/Pacific Islanders. This item was also hypothesized to show DIF for minority groups. The impact of DIF was negligible. Conditional on depression, the items, worthless and hopeless were more likely to be endorsed in the depressed direction by respondents with less than high school education vs. those with a graduate degree; the magnitude of DIF was slightly above the T1 threshold, but not that of NCDIF. These items were also hypothesized to show DIF in the direction of more feelings of worthlessness by groups with lower education. While the magnitude and aggregate impact of DIF was small, in a few instances, individual impact was observed.

Information provided was relatively high, particularly in the middle upper (depressed) tail of the distribution. Reliability estimates were high (> 0.90) across all studied groups, regardless of estimation method.

Conclusions: This was the first study to evaluate measurement equivalence of the PROMIS Depression short forms across large samples of ethnically diverse groups. There were few items with DIF, and none of high magnitude, thus supporting the use of PROMIS Depression short form measures across such groups. These results could be informative for those using the short forms in minority populations or clinicians evaluating individuals with the depression short forms.

Key Words: depression, PROMIS[®], differential item functioning, item response theory, ethnic diversity

Jeanne A. Teresi, Ed.D, Ph.D. Columbia University Stroud Center at New York State Psychiatric Institute 1051 Riverside Drive, Box 42, Room 2714, New York New York, 10032-3702, USA <u>Teresimeas@aol.com; jat61@columbia.edu</u>

Measurement equivalence of the Patient Reported Outcomes Measurement Information System[®] (PROMIS[®]) Anxiety short forms in ethnically diverse groups

Jeanne A. Teresi, Katja Ocepek-Welikson, Marjorie Kleinman, Mildred Ramirez & Giyeon Kim

Abstract

This is the first study of the measurement equivalence of the Patient Reported Outcomes Measurement Information System^{*} (PROMIS^{*}) Anxiety short forms in a large ethnically diverse sample. The psychometric properties and differential item functioning (DIF) were examined across different racial/ethnic, educational, age, gender and language groups. Methods: These data are from individuals selected from cancer registries in the United States. For the analyses of race/ethnicity the reference group was non-Hispanic Whites (n = 2,263), the studied groups were non-Hispanic Blacks (n = 1,117), Hispanics (n = 1,043) and Asians/Pacific Islanders (n = 907). Within the Hispanic subsample, there were 335 interviews conducted in Spanish and 703 in English. The 11 anxiety items were from the PROMIS emotional disturbance item bank.

DIF hypotheses were generated by content experts who rated whether or not they expected DIF to be present, and the direction of the DIF with respect to several comparison groups. The primary method used for DIF detection was the Wald test for examination of group differences in item response theory (IRT) item parameters accompanied by magnitude measures. Expected item scores were examined as measures of magnitude. The method used for quantification of the difference in the average expected item scores was the non-compensatory DIF (NCDIF) index. DIF impact was examined using expected scale score functions. Additionally, precision and reliabilities were examined using several methods.

Results: Although not hypothesized to show DIF for Asians/Pacific Islanders, every item evidenced DIF by at least one method. Two items showed DIF of higher magnitude for Asians/Pacific Islanders vs. Whites: "Many situations made me worry" and "I felt anxious". However, the magnitude of DIF was small and the NCDIF statistics were not above threshold. The impact of DIF was negligible. For education, six items were identified with consistent DIF across methods: fearful, anxious, worried, hard to focus, uneasy and tense. However, the NCDIF was not above threshold and the impact of DIF on the scale was trivial. No items showed high magnitude DIF for gender. Two items showed slightly higher magnitude for age (although not above the cutoff): worried and fearful. The scale level impact was trivial. Only one item showed DIF with the Wald test after the Bonferroni correction for the language comparisons: "I felt fearful". Two additional items were flagged in sensitivity analyses after Bonferroni correction, anxious and many situations made me worry. The latter item also showed DIF of higher magnitude, with an NCDIF value (0.144) above threshold. Individual impact was relatively small.

Conclusions: Although many items from the PROMIS short form anxiety measures were flagged with DIF, item level magnitude was low and scale level DIF impact was minimal; however, three items: anxious, worried and many situations made me worry might be singled out for further study. It is concluded that the PROMIS Anxiety short form evidenced good psychometric properties, was relatively invariant across the groups studied, and performed well among ethnically diverse subgroups of Blacks, Hispanic, White non-Hispanic and Asians/Pacific Islanders. In general more research with the Asians/Pacific Islanders group is needed. Further study of subgroups within these broad categories is recommended.

Key words: anxiety, PROMIS, item response theory, differential item functioning, ethnic diversity

Jeanne A. Teresi, Ed.D, Ph.D. Columbia University Stroud Center at New York State Psychiatric Institute 1051 Riverside Drive, Box 42, Room 2714, New York New York, 10032-3702, USA Teresimeas@aol.com ; jat61@columbia.edu

Measurement equivalence of the Patient Reported Outcomes Measurement Information System^{*} (PROMIS^{*}) Applied Cognition – General Concerns, short forms in ethnically diverse groups

Robert Fieo, Katja Ocepek-Welikson, Marjorie Kleinman, Joseph P. Eimicke, Paul K. Crane, David Cella & Jeanne A. Teresi

Abstract

Aims: The goals of these analyses were to examine the psychometric properties and measurement equivalence of a self-reported cognition measure, the Patient Reported Outcome Measurement Information System^{*} (PROMIS^{*}) Applied Cognition – General Concerns short form. These items are also found in the PROMIS Cognitive Function (version 2) item bank. This scale consists of eight items related to subjective cognitive concerns. Differential item functioning (DIF) analyses of gender, education, race, age, and (Spanish) language were performed using an ethnically diverse sample (n = 5,477) of individuals with cancer. This is the first analysis examining DIF in this item set across ethnic and racial groups. Methods: DIF hypotheses were derived by asking content experts to indicate whether they posited DIF for each item and to specify the direction. The principal DIF analytic model was item response theory (IRT) using the graded response model for polytomous data, with accompanying Wald tests and measures of magnitude. Sensitivity analyses were conducted using ordinal logistic regression (OLR) with a latent conditioning variable. IRT-based reliability, precision and information indices were estimated.

Results: DIF was identified consistently only for the item, brain not working as well as usual. After correction for multiple comparisons, this item showed significant DIF for both the primary and sensitivity analyses. Black respondents and Hispanics in comparison to White non-Hispanic respondents evidenced a lower conditional probability of endorsing the item, brain not working as well as usual. The same pattern was observed for the education grouping variable: as compared to those with a graduate degree, conditioning on overall level of subjective cognitive concerns, those with less than high school education also had a lower probability of endorsing this item. DIF was observed for age for two items after correction for multiple comparisons for both the IRT and OLR-based models: "I have had to work really hard to pay attention or I would make a mistake" and "I have had trouble shifting back and forth between different activities that require thinking". For both items, conditional on cognitive complaints, older respondents had a higher likelihood than younger respondents of endorsing the item in the cognitive complaints direction. The magnitude and impact of DIF was minimal.

The scale showed high precision along much of the subjective cognitive concerns continuum; the overall IRT-based reliability estimate for the total sample was 0.88 and the estimates for subgroups ranged from 0.87 to 0.92. Conclusion: Little DIF of high magnitude or impact was observed in the PROMIS Applied Cogni-tion – General Concerns short form item set. One item, "It has seemed like my brain was not working as well as usual" might be singled out for further study. However, in general the short form item set was highly reliable, informative, and invariant across differing race/ethnic, educational, age, gender, and language groups.

Key words: PROMIS[®], cognitive concerns, item response theory, differential item functioning, race, ethnicity

Robert Fieo, Assistant Professor University of Florida College of Medicine Department of Geriatric Research 2004 Mowry Road Gainesville, FL 32611, USA <u>fieo@ufl.edu</u>

Measurement equivalence of the Patient Reported Outcomes Measurement Information System[®] (PROMIS[®]) Pain Interference short form items: Application to ethnically diverse cancer and palliative care populations

Jeanne A. Teresi, Katja Ocepek-Welikson, Karon F. Cook, Marjorie Kleinman, Mildred Ramirez, M. Carrington Reid & Albert Siu

Abstract

Reducing the response burden of standardized pain measures is desirable, particularly for individuals who are frail or live with chronic illness, e.g., those suffering from cancer and those in palliative care. The Patient Reported Outcome Measurement Information System[®] (PROMIS[®]) project addressed this issue with the provision of computerized adaptive tests (CAT) and short form measures that can be used clinically and in research. Although there has been substantial evaluation of PROMIS item banks, little is known about the performance of PROMIS short forms, particularly in ethnically diverse groups. Reviewed in this article are findings related to the differential item functioning (DIF) and reliability of the PROMIS pain interference short forms across diverse socio-demographic groups. Methods: DIF hypotheses were generated for the PROMIS short form pain interference items. Initial analyses tested item response theory (IRT) model assumptions of unidimensionality and local independence. Dimensionality was evaluated using factor analytic methods; local dependence (LD) was tested using IRT-based LD indices. Wald tests were used to examine group differences in IRT parameters, and to test DIF hypotheses. A second DIF-detection method used in sensitivity analyses was based on ordinal logistic regression with a latent IRT-derived conditioning variable. Magnitude and impact of DIF were investigated, and reliability and item and scale information statistics were estimated. Results: The reliability of the short form item set was excellent. However, there were a few items with high local dependency, which affected the estimation of the final discrimination parameters. As a result, the item, "How much did pain interfere with enjoyment of social activities?" was excluded in the DIF analyses for all subgroup comparisons. No items were hypothesized to show DIF for race and ethnicity; however, five items showed DIF after adjustment for multiple comparisons in both primary and sensitivity analyses: ability to concentrate, enjoyment of recreational activities, tasks away from home, participation in social activities, and socializing with others. The magnitude of DIF was small and the impact negligible. Three items were consistently identified with DIF for education: enjoyment of life, ability to concentrate, and enjoyment of recreational activities. No item showed DIF above the magnitude threshold and the impact of DIF on the overall measure was minimal. No item showed gender DIF after correction for multiple comparisons in the primary analyses. Four items showed consistent age DIF: enjoyment of life, ability to concentrate, day to day activities, and enjoyment of recreational activities, none with primary magnitude values above threshold. Conditional on the pain state, Spanish speakers were hypothesized to report less pain interference on one item, enjoyment of life. The DIF findings confirmed the hypothesis; however, the magnitude was small. Using an arbitrary cutoff point of theta (θ) \geq 1.0 to classify respondents with acute pain interfer-ence, the highest number of changes were for the education groups analyses. There were 231 respondents (4 % of the total sample) who changed from the designation of no acute pain interference to acute interference after the DIF adjustment. There was no change in the designations for race/ethnic subgroups, and a small number of changes for respondents aged 65 to 84. Conclusions: Although significant DIF was observed after correction for multiple comparisons, all DIF was of low magnitude and impact. However, some individual-level impact was observed for low education groups. Reliability estimates were high. Thus, the PROMIS short form pain items examined in this ethnically diverse sample performed relatively well; although one item was prob-lematic and removed from the analyses. It is concluded that the majority of the PROMIS pain interference short form items can be recommended for use among ethnically diverse groups, including those in palliative care and with cancer and chronic illness.

Key words: Differential item functioning, PROMIS[®], pain, measurement equivalence, palliative care, ethnicity, cancer

Jeanne A. Teresi, Ed.D, Ph.D. Columbia University Stroud Center at New York State Psychiatric Institute 1051 Riverside Drive, Box 42, Room 2714, New York

<u>top</u>

Measurement properties of PROMIS Sleep Disturbance short forms in a large, ethnically diverse cancer cohort *Roxanne E. Jensen, Bellinda L. King-Kallimanis, Eithne Sexton, Bryce B. Reeve, Carol M. Moinpour, Arnold L. Potosky, Tania Lobo & Jeanne A. Teresi*

Abstract

AIMS: To evaluate model fit, differential item function (DIF), and construct validity of select short forms from the PROMIS[®] Sleep Disturbance item bank.

METHODS: We recruited cancer survivors who were between 6 - 13 months post diagnosis (n = 4,956), as part of the Measuring Your Health (MY-Health) study. We measured sleep disturbance using 10 items commonly found in PROMIS Sleep Disturbance short forms (Sleep 4a, Sleep 6a, Sleep 8b), and which are frequently administered in computerized adaptive testing.

We evaluated domain reliability using Cronbach's coefficient alpha and factorial validity by fitting a PROMIS Sleep Disturbance unidimensional measurement model using confirmatory factor analy-sis (CFA). At the item-level, we examined DIF with respect to race/ethnicity (non-Hispanic White [NHW], non-Hispanic Black [NHB], Hispanic, and Asian/Pacific Islander), age, and sex. We used a multi-group CFA and multiple indicators, multiple methods (MIMIC) analyses. We then assessed construct validity (convergent, discriminate, and known groups) for sleep short forms, and a new "best fit" 6-item sleep disturbance short form.

RESULTS: We identified a satisfactory unidimensional sleep disturbance 6-item measure ($\chi^2(6)$ 37.6, p < 0.001, RMSEA = 0.031). To achieve this, we removed four items from the model with item content overlap and added residual covariances between positively worded items in order to address a method effect. We identified one instance of DIF: NHW participants were less likely to agree with the statement "I had difficulty falling asleep" compared to NHBs, Hispanics, or Asians/Pacific Islanders, who all reported the same level of sleep disturbance. After controlling for DIF, we extended this into a MIMIC model, identifying no additional DIF by age or sex. Across all race/ethnicity groups, the adjusted overall means suggest that older adults reported significantly lower sleep disturbance, and NHW, NHB, and Hispanic women reported significantly higher sleep disturbance than male survivors of the same race/ethnicity. CONCLUSIONS: We could not fit a unidimensional measurement model for either the full 10-items, or for any combination of sleep disturbance items used in PROMIS Sleep Disturbance short forms. However, after we removed the data well, with very little evidence of substantial DIF. This suggests this new measure (Sleep 6b) can be used in different groups across the adult lifespan, and in males and females in a heterogeneous cancer population. Our findings suggest further validation work is necessary to understand the impact of reverse-scored items, response set effects, and content overlap in this item bank.

Key words: sleep disturbance, PROMIS, differential item functioning, measurement invariance, methods effects

Roxanne E. Jensen, Ph.D. Cancer Prevention and Control Program Lombardi Comprehensive Cancer Center Georgetown University 3300 Whitehaven Street NW, Suite 4100 Washington, DC 20007, USA rj222@georgetown.edu top

Differential item functioning in Patient Reported Outcomes Measurement Information System[®] (PROMIS[®]) Physical Functioning short forms: Analyses across ethnically diverse groups

Richard N. Jones, Doug Tommet, Mildred Ramirez, Roxanne Jensen & Jeanne A. Teresi

Abstract

We analyzed physical functioning short form items derived from the PROMIS^{*} item bank (PF16) using data from more than 5,000 recently diagnosed, ethnically diverse cancer patients. Our goal was to determine if the short form items demonstrated evidence of differential item functioning (DIF) according to sociodemographic characteristics in this clinical sample. We evaluated responses for evidence of unidimensionality, local independence (given a single common factor), differential item functioning, and DIF impact. DIF was evaluated attributable to sex, age (middle aged vs. younger and older), race/ethnicity (White vs. Black or African-American, Asian/Pacific Islander, Hispanic) and level of education. We used a multiple group confirmatory factor analysis with covariates approach, a multiple indicators multiple causes (MIMIC) model. We confirmed essential unidimensionality but some evidence for multidimensionality is present, particularly for basic activities of daily living items, and many instances of local dependence. The presence of local dependence calls for further review of the meaning and measurement of the physical functioning domain among cancer patients.

Nearly every item demonstrated statistically significant DIF. In all group comparisons the impact of DIF was negligible. However, the Hispanic subgroup comparison revealed an impact estimate just below an arbitrary threshold for small impact. Within the limitations of local dependency violations, we conclude that items from a static short form derived from the PROMIS physical functioning item bank displayed trivial and ignorable DIF attributable to sex, race, ethnicity, age, and education among cancer patients.

Key words: PROMIS, physical function, differential item functioning, cancer, measurement equivalence, ethnicity

Richard N. Jones, Sc.D. Department of Psychiatry and Human Behavior Department of Neurology Warren Alpert Medical School Brown University, Butler Hospital 345 Blackstone Boulevard, Box G-BH, Providence Rhode Island 02906, USA richard jones@brown.edu

Measuring social function in diverse cancer populations: Evaluation of measurement equivalence of the Patient Reported Outcomes Measurement Information System[®] (PROMIS[®]) Ability to Participate in Social Roles and Activities short form

Elizabeth A. Hahn, Michael A. Kallen, Roxanne E. Jensen, Arnold L. Potosky, Carol M. Moinpour, Mildred Ramirez, David Cella & Jeanne A. Teresi

Abstract

Conceptual and psychometric measurement equivalence of self-report questionnaires are basic requirements for valid cross-cultural and demographic subgroup comparisons. The purpose of this study was to evaluate the psychometric measurement equivalence of a 10-item PROMIS[®] Social Function short form in a diverse population-based sample of cancer patients obtained through the Measuring Your Health (MY-Health) study (n = 5,301). Participants were cancer survivors within six to 13 months of a diagnosis of one of seven cancer types, and spoke English, Spanish, or Mandarin Chinese. They completed a survey on sociodemographic and clinical characteristics, and health status. Psychometric measurement equivalence was evaluated with an item response theory approach to differential item functioning (DIF) detection and impact. Although an expert panel proposed that many of the 10 items might exhibit measurement bias, or DIF, based on gender, age, race/ethnicity, and/or education, no DIF was detected using the study's standard DIF criterion, and only one item in one sample comparison was flagged for DIF using a sensitivity DIF criterion. This item's flagged DIF had only a trivial impact on estimation of scores. Social function measures are especially important in cancer because the disease and its treatment can affect the quality of marital relationships, parental responsibilities, work abilities, and social activities. Having culturally relevant, linguistically equivalent and psychometrically sound patient-reported measures in multiple languages helps to overcome some common barriers to including underrepresented groups in research and to conducting cross-cultural research.

Key words: patient-reported outcomes, social function, psychometrics, differential item function-ing, cancer

Elizabeth A. Hahn, PhD Department of Medical Social Sciences Northwestern University Feinberg School of Medicine 633 N. St. Clair St., Suite 1900 Chicago, IL 60611, USA <u>e-hahn@northwestern.edu</u>

Epilogue to the two-part series: Measurement equivalence of the Patient Reported Outcomes Measurement Information System[®] (PROMIS[®]) short forms

Jeanne A. Teresi & Bryce B. Reeve

Abstract

The articles in this two-part series of Psychological Test and Assessment Modeling describe the psychometric performance and measurement equivalence of the Patient Reported Outcomes Measurement Information System[®] (PROMIS[®]) short form measures in ethnically, socio-demographically diverse groups of cancer patients. Measures in eight health-related quality of life domains were evaluated: fatigue, depression, anxiety, cognition, pain, sleep, and physical and social function. State-of-the-art latent variable methods, most based on item response theory, and described in two methods overview articles in this series were used to examine differential item functioning (DIF). Findings were generally supportive of the performance of the PROMIS measures. Although use of powerful methods and large samples resulted in the identification of many items with DIF, practi-cally none were identified with high magnitude. The aggregate level impact of DIF was small, and minimal individual impact was detected. Some methodological challenges were encountered in-volving positively and negatively worded items, but most were resolved through modest item removal. Sensitivity analyses showed minimal impact of model assumption violation on the results presented.

A cautionary note is the observance of a few instances of individual-level impact of DIF in the analyses of depression, anxiety, and pain, and one instance of aggregate level impact just below threshold in the analyses of physical function. Although this sample of over 5,000 individuals was diverse, ethnically, a limitation was the lack of ability to examine language groups other than Spanish and English and specific ethnic subgroups within Hispanic, Asian/Pacific Islander, and Black subsamples.

Extensive qualitative and quantitative analyses were performed in the development of PROMIS item banks. These sets of analyses, performed by several teams of psychometricians, statisticians, and qualitative experts, were the first to examine measurement equivalence of PROMIS short forms among ethnically diverse groups, and were also the first examination of PROMIS short forms among adults with cancer. Results presented in these articles provide strong evidence supporting the measurement equivalence of PROMIS short forms.

Key words: Patient Reported Outcomes Measurement Information System[®], PROMIS[®], short forms, measurement equivalence, differential item functioning, item response theory, reliability, validity, ethnic diversity, cancer

Jeanne A. Teresi, Ed.D., Ph.D. Columbia University Stroud Center at New York State Psychiatric Institute 1051 Riverside Drive, Box 42, Room 2714 New York 10032-3702, USA Teresimeas@aol.com