CLASSICAL TEST THEORY- AND ITEM RESPONSE THEORY-BASED ASSESSMENT OF THE STRUCTURAL VALIDITY OF THE PROMIS - GLOBAL HEALTH (PROMIS-GH) QUESTIONNAIRE IN PATIENTS WITH MUSCULOSKELETAL DISORDERS

Giusti Emanuele Maria[1], Pellicciari Leonardo[2], Veronesi Giovanni[1], Terwee Caroline[3,4], Roorda Leo[5]

[1] *EPIMED Research Center, Department of Medicine and Surgery, University of Insubria, Varese, Italy*
[2] *IRCCS Istituto delle Scienze Neurologiche di Bologna, Bologna, Italy*
[3] *Department of Epidemiology and Data Science, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands.*
[4] *Amsterdam Public Health Research Institute, Amsterdam, The Netherlands.*
[5] *Amsterdam Rehabilitation Research Center | Reade, Amsterdam, The Netherlands.*

Introduction

Obtaining valid measures of health-related quality of life is essential for epidemiologic and intervention studies focused on patients with musculoskeletal disorders, as these conditions affect the physical, psychological and social health of sufferers [1,2]. Within the Patient-Reported Outcomes Measurement Information System (PROMIS®) initiative [3], the PROMIS Global Health (PROMIS-GH) has been recommended for the assessment of health-related quality of life in core sets for the evaluation of overall adult health and of the outcomes of low back pain patients [4,5]. However, in samples of patients with musculoskeletal disorders, existing research has mostly focused on its responsiveness (e.g. [6]) and convergent and discriminant validity [7], without addressing its structural validity. This is problematic, as structural validity is essential for calculating summary scores, comparing groups and evaluating changes over time [8,9]. Classical Test Theory (CTT) and Item Response Theory (IRT) can be integrated for studying the structural validity of the PROMIS-GH, using CTT to identify the best-fitting factor structure and IRT to assess the link between item responses and the latent trait.

Aims

To study the factor structure of the PROMIS-GH and the unidimensionality, local independence, monotonicity, fit to Graded Response Model (GRM), IRT parameters, information and measurement invariance across demographic and clinical parameters, as well as across clinical/general population, of its subscales in patients with musculoskeletal disorders.

Methods

We employed PROMIS-GH data collected from four separate cohorts of patients, one including patients with chronic musculoskeletal pain, one including patients with rheumatoid arthritis, one including patients with hip or knee osteoarthritis, and one including patients who underwent physiotherapy in the year before being enrolled. We firstly performed a Confirmatory Factor Analysis (CFA) to assess whether the two-factor structure was confirmed (criteria: CFI and TLI >.95, SRMR <.08 and RMSEA <.06). Second, we studied the fit of the two PROMIS-GH subscales, i.e. the Global Mental Health (GMH) and Global Physical Health (GPH), to a unidimensional model through a CFA and an exploratory bifactor analysis (criteria: Explained Common Variance [ECV] >.70 and Omega hierarchical [ω_H]> .80). Third, we investigated local independence of the subscales by examining

residual correlations from the unidimensional CFAs (criterion: no residual correlation ≥.20) and monotonicity through a Mokken scale analysis (criterion: scalability coefficient [H] >.50). Fourth, we investigated the fit to a GRM (criteria: non-significant Orlando and Thissen's S-χ^2 test, exploration of empirical plots), discrimination (criterion: item slopes [α]>1.00), targeting (minimum threshold [β] <-2.5) and item and total score information (i.e, precision across the latent trait; criterion: visual inspection of information curves). Finally, we assessed measurement invariance through Differential Item Functioning (DIF) analyses based on a logistic regression framework (criterion: McFadden pseudo-R^2 <.02 and presence of non-negligible differences between the Test Characteristic Curves of the compared groups) for age, sex, disease duration, educational level, cohort, administration modality, and, also using data from a sample from the general population (n=4370), across clinical/general population.

Results

The PROMIS-GH was completed by 4506 subjects (mean±SD age=55.6±13.8; 30% male). In both the two-factor and in the unidimensional CFAs on each subscale, the CFI, TLI and SRMR indicated adequate fit, while the RMSEA (>.06) indicated suboptimal fit. The exploratory bifactor analysis performed on the GMH and GPH subscales indicated that both the indices of the GPH were above the respective cut-offs (ECV= .91, ω_H=.86), while for the GMH subscale, the ω_H was slightly below (ECV=.84, ω_H=.79). Based on these results, we considered both subscales as unidimensional. No correlation between residuals of the unidimensional CFAs was >.20, indicating local independence.

Table 1 reports descriptive statistics, results of the Mokken scale analysis, fit to the GRM model and IRT parameters. Monotonicity was confirmed for both subscales. Based on S-χ^2 values, all items misfitted the GRM model. However, visual inspection of the empirical plots revealed that misfit was negligible for all items except for item Global10 ("Being bothered by emotional problems"). Items' discrimination and subscales' targeting were adequate. The item information curves showed that information was high across the latent trait for all items except for item Global10. Nonetheless, item Global10 exhibited the highest information at low values of the mental health latent trait. Regarding DIF, only a uniform DIF in Global10 across cohorts was detected (R^2=.04), which was considered negligible based on inspection of the Test Characteristic Curves of the cohorts.

Table 1. Descriptive statistics, scalability coefficients, fit to a Graded Response Model and Item Response Theory parameters of the PROMIS – Global Health (PROMIS-GH) items

Item	Mean ± SD	Mokken scale analysis	Graded Response Model fit		Item Response Theory parameters				
		H	S-χ^2	p-value	α	β_1	β_2	β_3	β_4
Global Mental Health									
Global02	2.92 ± 1.07	.73	62.3	<0.001	3.28	-1.58	-0.40	0.64	1.57
Global04	3.09 ± 1.13	.77	105.3	<0.001	4.09	-1.60	-0.50	0.42	1.22
Global05	2.97 ± 1.13	.74	85.6	<0.001	3.84	-1.45	-0.40	0.50	1.39

Global10	3.61 ± 1.11	.68	183.1	<0.001	2.06	-2.62	-1.19	-0.12	0.83
Global Physical Health									
Global03	2.47 ± 0.99	.68	52.0	<0.001	2.54	-1.28	0.20	1.27	2.24
Global06	3.61 ± 1.19	.70	88.7	<0.001	2.72	-2.13	-1.02	-0.13	0.61
Global07	3.22 ± 1.02	.67	78.9	<0.001	2.68	-2.74	-0.65	0.25	1.47
Global08	3.14 ± 1.06	.66	42.9	0.01	2.48	-2.00	-0.71	0.42	1.51

Notes. Abbreviations: SD: standard deviation.

Conclusions

Overall, the integrated use of CTT and IRT allowed to support the structural validity of the PROMIS-GH in patients with musculoskeletal disorders. Item Global10 showed misfit to the GRM model and low information across most of the latent trait, implying an inaccurate and imprecise contribution to the estimation of mental health in patients with medium to high mental health. Nonetheless, since it provided the highest information at low levels of mental health, we recommend replacing it with another item of similar difficulty rather than removing it. There were minor deviations from unidimensionality in both subscales, which might be attributed to the PROMIS-GH combining reflective and formative items, a complexity that neither framework fully accommodates.

References

[1] Beaudart C, Biver E, Bruyère O, et al. Assessment of Quality of Life in Musculo-Skeletal Health. Aging Clin Exp Res 2018;30:413–8.

[2] Roux CH, Guillemin F, Boini S, et al. Impact of musculoskeletal disorders on quality of life: an inception cohort study. Annals of the Rheumatic Diseases 2005;64:606–11.

[3] Cella D, Yount S, Rothrock N, et al. The Patient-Reported Outcomes Measurement Information System (PROMIS): progress of an NIH Roadmap cooperative group during its first two years. Med Care 2007;45:S3–11.

[4] Gangannagaripalli J, Albagli A, Myers SN, et al. A Standard Set of Value-Based Patient-Centered Outcomes and Measures of Overall Health in Adults. Patient 2022;15:341–51.

[5] Chiarotto A, Terwee CB, Kamper SJ, et al. Evidence on the measurement properties of health-related quality of life instruments is largely missing in patients with low back pain: a systematic review. Journal of Clinical Epidemiology 2018;102:23–37.

[6] Shim J, Hamilton D. Comparative responsiveness of the PROMIS-10 Global Health and EQ-5D questionnaires in patients undergoing total knee arthroplasty. The Bone & Joint Journal 2019;101:832–7.

[7] Lapin B, Davin S, Stilphen M, et al. Validation of PROMIS CATs and PROMIS global health in an interdisciplinary pain program for patients with chronic low back pain. Spine 2020;45:E227–35.

[8] Mokkink LB, Terwee CB, Patrick DL, et al. The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status

measurement instruments: an international Delphi study. Qual Life Res 2010;19:539–49.
[9] Maassen E, D'Urso ED, van Assen MALM, et al. The dire disregard of measurement invariance testing in psychological science. Psychol Methods 2023.