

UCLA

UCLA Previously Published Works

Title

Validity of PROMIS physical function measured in diverse clinical samples.

Permalink

<https://escholarship.org/uc/item/9hd8v5nw>

Authors

Schalet, Benjamin D

Hays, Ron D

Jensen, Sally E

et al.

Publication Date

2016-05-01

DOI

10.1016/j.jclinepi.2015.08.039

Peer reviewed

Validity of PROMIS physical function measured in diverse clinical samples

Benjamin D. Schalet^{a,*}, Ron D. Hays^b, Sally E. Jensen^a, Jennifer L. Beaumont^a,
James F. Fries^c, David Cella^a

^aDepartment of Medical Social Sciences, Northwestern University Feinberg School of Medicine, Chicago, IL 60611, USA

^bDepartment of Medicine, Division of General Internal Medicine and Health Services Research, UCLA, Los Angeles, CA 90095-1736, USA

^cDepartment of Medicine, Stanford University School of Medicine, Palo Alto, CA 94304, USA

Accepted 21 August 2015; Published online 9 March 2016

Abstract

Objectives: To evaluate the validity of the Patient-Reported Outcomes Measurement Information System (PROMIS) Physical Function measures using longitudinal data collected in six chronic health conditions.

Study Design and Setting: Individuals with rheumatoid arthritis (RA), major depressive disorder (MDD), back pain, chronic obstructive pulmonary disease (COPD), chronic heart failure (CHF), and cancer completed the PROMIS Physical Function computerized adaptive test or fixed-length short form at baseline and at the end of clinically relevant follow-up intervals. Anchor items were also administered to assess change in physical function and general health. Linear mixed-effects models and standardized response means were estimated at baseline and follow-up.

Results: A total of 1,415 individuals participated (COPD $n = 121$; CHF $n = 57$; back pain $n = 218$; MDD $n = 196$; RA $n = 521$; cancer $n = 302$). The PROMIS Physical Function scores improved significantly for treatment of CHF and back pain patients but not for patients with MDD or COPD. Most of the patient subsamples that reported improvement or worsening on the anchors showed a corresponding positive or negative change in PROMIS Physical Function.

Conclusion: This study provides evidence that the PROMIS Physical Function measures are sensitive to change in intervention studies where physical function is expected to change and able to distinguish among different clinical samples. The results inform the estimation of meaningful change, enabling comparative effectiveness research. © 2016 Elsevier Inc. All rights reserved.

Keywords: PROMIS; Physical function; Patient-reported outcome; Chronic conditions; Item bank; Validity

1. Introduction

Physical function is a concept that refers to the ability to carry out activities ranging from self-care (activities of daily living) to more vigorous behaviors that require increasing degrees of mobility, strength, or endurance [1–6]. Physical function emphasizes abilities above and below the population mean and thus reflects a more comprehensive range of abilities than the construct of disability [7]. It may be affected by chronic health

conditions, including rheumatoid arthritis (RA), back pain, cancer, chronic obstructive pulmonary disease (COPD), and chronic heart failure (CHF) [8,9].

Several extensively evaluated measures of physical function and disability exist, such as the Health Assessment Questionnaire Disability Index (HAQ-DI) [10,11] and the short form-36 (SF-36) Physical Function scale [12]. To improve on these measures, the Patient-Reported Outcomes Measurement Information System (PROMIS[®]) developed a large pool of physical function items and calibrated it using item response theory to allow for administration using either computerized adaptive testing (CAT) or fixed-length SFs [7,13,14].

Prior published articles have described the development process and information about the precision of the PROMIS Physical Function item bank in cross-sectional administrations [13,14]. Examining data from a prospective observational

Conflict of interest: David Cella is an unpaid member of the board of directors and officer of the PROMIS Health Organization. The remaining authors have no conflicts of interests to disclose.

* Corresponding author. Tel.: +1-312-503-3640; fax: +1-312-503-4800.

E-mail address: b-schalet@northwestern.edu (B.D. Schalet).

What is new?**Key findings**

- Patient-Reported Outcomes Measurement Information System (PROMIS) Physical Function measures are responsive in intervention studies that target back pain and congestive heart failure.

What this adds to what was known?

- Clinical groups differ in hypothesized ways on the PROMIS Physical Function measures.

What is the implication and what should change now?

- The PROMIS Physical Function measures are suitable for use in clinical trials and comparative effectiveness studies.

study of RA, an additional report examines the responsiveness and minimally important difference for the PROMIS Physical Function 20-item SF [15]. The current article reports on an evaluation of the PROMIS Physical Function measure in six longitudinal studies of adults with different chronic health conditions: RA, back pain, cancer, COPD, CHF, and major depressive disorder (MDD).

We hypothesize improvements in physical function after treatment for CHF (heart surgery), back pain (spinal injection), and the resolution of COPD exacerbation [16]. No firm a priori predictions are possible for physical function change in treatment for MDD or for physical function changes in people with cancer or RA when followed in an observational cohort study in the time line of these studies (6–12 weeks for cancer; 12 months for RA). Nevertheless, we expect that subsamples of cancer and RA patients with improvements or deteriorations on general health anchors should show corresponding positive or negative changes in PROMIS Physical Function scores [17]. Finally, from a cross-sectional perspective, we also hypothesize that the MDD sample will have better physical function than samples targeting individuals with physical diseases and that patients experiencing a COPD exacerbation will have worse physical function than patients with stable COPD [18].

2. Method*2.1. Measures**2.1.1. PROMIS Physical Function*

The PROMIS wave 1 physical function item bank consists of 124 items that assess mobility (lower extremity), dexterity (upper extremity), axial or central (neck and back function), and complicated actions that cover multiple

domains (daily living activities) [13,19]. The items in the 10-item SF were selected to represent the range of physical function with high levels of precision. The 10-item SF correlates very highly ($r = 0.96$) with the full item bank [6]. Other forms of the instrument selected from the 124-item bank include a brief CAT, a 20-item SF [20], and another 10-item SF intended for use in a cancer population [21].

For the COPD and CHF samples, both CAT and the 10-item SF were administered. For MDD and back pain samples, only the CAT was administered. For RA, only the SFs were given (both 10- and 20-item versions). For cancer, a different 10-item SF was administered [21]; however, because all PROMIS SFs are scored using PROMIS item parameters, all resulting SF scores are on the same T-score metric. This article evaluates PROMIS Physical Function scores based on CAT administration for the samples where CAT was administered on the 10-item SFs in the remaining samples. The PROMIS CAT administration applied the following stopping rules: items were administered until the standard error was <3.0 and the number of items administered was ≥ 4 items and ≤ 12 items (no more than 12 CAT items were administered). CAT item selection followed maximum posterior weighted information criterion [22,23].

2.1.2. Change anchors

Each sample was administered either a retrospective global measure of change in health or global measures of health administered at baseline and follow-up. In addition, for all samples except MDD and back pain, we also had a physical function anchor item. We used either the retrospective measure of change or the difference between follow-up and baseline on these global measures as anchors to evaluate prospective change on the PROMIS Physical Function measure. We defined three change groups based on the anchors: better, about the same, or worse than baseline. For the specific items used as anchors, please see the overview article in this volume [24].

2.2. Samples and research design

The samples and research designs for each clinical condition are described in detail in Cook et al. (this volume) [24]. Briefly, the samples were drawn from the following clinical populations: (1) back pain, (2) cancer, (3) MDD, (4) COPD, subdivided into exacerbation and stable groups, (5) CHF, and (6) RA. The studies of MDD, back pain, and CHF followed patients as they enrolled in new treatments. Patients with COPD exacerbation were treated for their condition, which was expected to resolve over the course of the study. Both RA and cancer samples were heterogeneous with respect to intervention but were dominated by participants who were already receiving treatments by the time they enrolled in the present study. We examined the longitudinal data at baseline and follow-up, namely, 3 months after start of study (MDD, back pain, and COPD), 8–12 weeks after heart transplantation (CHF), 6–12 weeks

after enrollment (cancer), and 12 months after enrollment (RA). Although the COPD stable, cancer, and RA groups were not enrolled in new treatments, we apply the terms “baseline” and “follow-up” to all study groups for consistency. Missing follow-up data percentages for PROMIS instruments were as follows: 4% for COPD and MDD, 9% for RA, 10% for cancer, and 16% for CHF, and 21% for pain.

2.3. Statistical analyses

Physical function measures were administered to all six clinical groups at baseline and follow-up. Linear mixed models were estimated with random subject effects to account for repeated observations under the missing at random assumption [25–28]. Least squares means, standard errors, and 95% confidence intervals were estimated.

As noted previously, we divided each of the clinical samples into three subgroups, representing better health, about the same health, or worse health. For each of these subgroups, we calculated the change in PROMIS Physical Function score and the standardized response mean (SRM), defined as the mean change divided by the standard deviation of that change [29–32]. Because of missing data at follow-up, subgroup sample sizes for the anchor-based analysis do not sum to the sample sizes for the mixed models analysis.

Because the cancer and RA samples were large ($N = 302$; $N = 521$) and we expected some patients in these groups to improve while others would deteriorate, we also computed least square means for the subsamples of patients who reported globally that their physical function became better, worse, and remained about the same over the course of the study.

3. Results

Least squares means and 95% confidence intervals from the mixed models are provided in Table 1.

3.1. Cross-sectional group differences

Physical function in the MDD group was at least 4.7 points higher than the other chronic conditions. Physical function was lower for the COPD-exacerbation group at baseline than the stable COPD group by 2.0 points, but this difference was not statistically significant, $t(119) = 1.59$, $P = 0.114$. At follow-up the difference was 2.3 points in the expected direction, but not statistically significant: $t(114) = 1.85$, $P = 0.067$.

3.2. Responsiveness to change

Fig. 1 illustrates the least square mean changes from baseline to follow-up. Consistent with our hypothesis, CHF and back pain groups improved significantly (mean changes of 7.2 and 3.4 T-score points, respectively).

Unexpectedly, patients with COPD exacerbations did not change significantly on physical function between baseline and follow-up (3 months). As expected, the mean score of the remaining groups (cancer, COPD stable, MDD, and RA) did not change significantly. However, as Table 2 shows, the cancer and RA change subgroups showed significant mean changes over time in the expected direction. For cancer, these effects were larger (2.7 and -4.6 T-score points) than for RA (1.3 and -1.9 T-score points).

3.3. Prospective change on PROMIS Physical Function by global ratings of change anchors

Changes on the PROMIS Physical Function measure generally corresponded to changes on the global anchors (see Table 3). Individuals grouped in the better health subsamples on the general health anchor showed SRMs ranging from 0.21 (RA) to 1.05 (CHF). Patients in the worse health group showed the expected negative change in PROMIS scores for cancer (-0.22), COPD stable (-0.55), and RA (-0.19). Although the remaining worse health subgroups (back pain, COPD exacerbation, MDD, CHF) showed positive change (SRMs = 0.04 to 0.29), each of these were smaller than the corresponding better health subgroups (differences in worse-better SRMs ranged from 0.26 to 0.35, excluding CHF [$n = 2$]).

Using the general physical function anchor, the results were also mostly in line with expectations. The better physical function subgroups showed positive SRMs change on the PROMIS measure, ranging from 0.11 (COPD stable) to 1.20 (CHF). With the exception of COPD stable, the worse physical function subgroups also showed expected negative SRMs for PROMIS change: -0.77 (cancer), -0.38 (COPD exacerbation), and -0.46 (RA), excluding CHF [$n = 3$]. For COPD stable, however, the worse physical health subgroup showed an SRM of 0.02. The small SRMs for COPD stable using the general physical function anchor, however, stand in contrast to the effects using the general health anchor in the same clinical group, showing a worse health SRM of -0.55 and a better health SRM of 0.70.

4. Discussion

The usefulness of a measure for clinical research and practice depends on its ability to detect change over time and response to clinical intervention. It is equally important that measures demonstrate stability in scores when no real change is present. This study extends that work by evaluating the performance on the measure longitudinally in six diverse clinical condition samples.

The prospective observational studies reported here provide an opportunity to compare PROMIS physical function scores across six clinical samples (back pain, cancer, COPD, CHF, MDD, and RA). These comparisons are made

Table 1. PROMIS Physical Function scores in different clinical samples

	Back pain (n = 218)	Cancer (n = 302)	CHF (n = 57)	COPD: exacerbation (n = 45)	COPD: stable (n = 76)	MDD (n = 196)	Rheumatoid arthritis (n = 521)
Baseline	37.5 (36.6–38.4)	41.8 (40.8–42.8)	34.8 (33.1–36.5)	36.0 (33.9–38.0)	38.0 (36.5–39.4)	46.5 (45.1–48.0)	40.7 (39.9–41.5)
Follow-up	40.9 (39.9–41.9)	41.9 (40.8–42.9)	42.0 (40.2–43.8)	36.2 (34.1–38.3)	38.2 (36.8–39.7)	47.3 (45.9–48.8)	40.3 (39.5–41.1)
Change	3.4 (2.5–4.3)**	0.1 (–0.7, 0.8)	7.2 (5.2–9.2)**	0.2 (–1.3, 1.8)	0.2 (–0.7, 1.2)	0.8 (–0.3, 1.9)	–0.4 (–0.7, 0.02)

Abbreviations: CHF, chronic heart disease; COPD, chronic obstructive pulmonary disease; MDD, major depressive disorder.

Higher scores indicate better physical function. Entries in the table denote the least squares mean and 95% confidence interval, as estimated in the mixed models.

*P < 0.05, **P ≤ 0.001.

on the same T-score metric, across conditions, and within condition over time (Fig. 1). Most of those comparisons over time (other than cancer and RA) included a baseline physical function assessment before initiating standard treatment for the condition. This enabled a view into the clinical responsiveness of PROMIS physical function assessment across these conditions.

We hypothesized that PROMIS Physical Function scores would improve over time in clinical samples receiving an intervention or experiencing a clinical course likely to result in enhanced physical function, namely CHF, back pain, and exacerbation of COPD. The findings from the CHF and back pain samples provided support for this hypothesis, with notable improvements over time. Contrary to our hypothesis, patients with COPD exacerbation did not report significant change in physical function over time. This result may reflect on the PROMIS Physical Function measure’s ability to detect change in this clinical population, particularly if improvement was confined to a narrow

set of physical tasks (e.g., climbing stairs). It is also possible that patients with COPD exacerbation simply did not have measurable improvement in physical function during the 12 weeks of treatment. Finally, COPD exacerbations represent acute worsening in health in contrast with CHF and back pain which represent relatively stable chronic disability. It is therefore possible that patients over-report their physical function during the acute exacerbation because they reference their usual state rather than their acutely ill state.

The MDD, RA, and cancer samples showed stability in average PROMIS Physical Function scores over time, consistent with our hypotheses. Given the more psychological than physical nature of depression, treatment for depression would not necessarily affect physical function. RA is considered an autoimmune, inflammatory condition and much of the sample in this study was already receiving heterogeneous interventions at the time of enrollment. Similarly, the cancer sample comprised individuals with

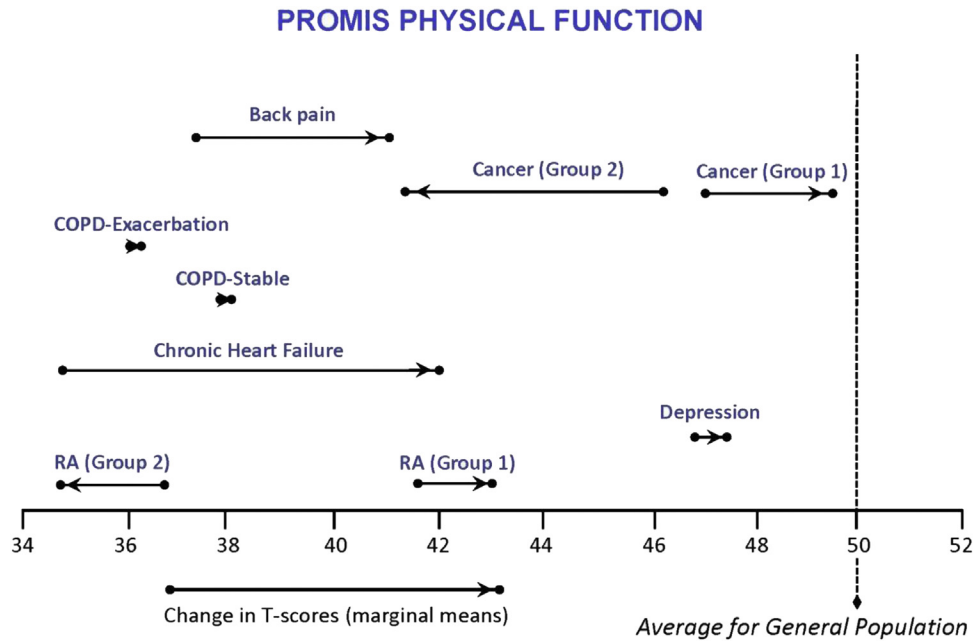


Fig. 1. PROMIS Physical Function scores in different clinical groups from baseline to follow-up. Higher scores indicate better physical function. A T-score of 50 reflects the mean (and 10 the standard deviation) in the US general population sample that was used to center the T-scores [33]. Cancer (group 1) and RA (group 1) patients reported better physical function on the domain-specific global change rating. Cancer (group 2) and RA (group 2) reported worse physical function on the domain-specific global change rating. COPD, chronic obstructive pulmonary disease; RA, rheumatoid arthritis.

Table 2. PROMIS Physical Function scores in cancer and RA subsamples

	Cancer-better ^a (n = 75)	Cancer-same (n = 114)	Cancer-worse ^b (n = 55)	RA-better ^a (n = 59)	RA-same (n = 262)	RA-worse ^b (n = 151)
Baseline	46.9 (44.7–49.1)	50.0 (48.4–51.6)	46.0 (43.8–48.2)	41.7 (39.6–43.7)	43.1 (42.0–44.2)	36.8 (35.6–37.9)
Follow-up	49.5 (47.4–51.7)	50.4 (48.8–52.0)	41.4 (39.2–43.6)	43.0 (40.9–45.0)	43.2 (42.1–44.3)	34.8 (33.7–36.0)
Change	2.7 (1.2–4.2)**	0.4 (–0.8, 1.5)	–4.6 (–6.3, 2.9)**	1.3 (0.1–2.6)*	0.1 (–0.4, 0.6)	–1.9 (–2.6, –1.2)**

Abbreviation: RA, rheumatoid arthritis.

Higher scores indicate better physical function. Entries in the table denote the least squares mean and 95% confidence interval, as estimated in the mixed models. Cancer subgroups were created using responses to the question, “Since the last time you filled out a questionnaire, your physical function is...” RA subgroups were created using responses to the question, “How has your ability to carry out your everyday physical activities such as walking, climbing stairs, carrying groceries, or moving a chair changed? ...,” assessed at follow-up.

* $P < 0.05$, ** $P \leq 0.001$.

^a Labeled “Group 1” in Fig. 1.

^b Labeled “Group 2” in Fig. 1.

diverse cancer diagnoses receiving a wide range of treatments. Some were likely to improve; some to worsen; and others to remain essentially the same, after a 2- to 3-month follow-up. The analysis of cancer and RA by subgroups reporting improvement or deterioration reflected that PROMIS Physical Function meaningfully distinguishes these subgroups. This pattern of effects has been found before in cancer patients [17]. Therefore, these findings indicate that the PROMIS Physical Function item bank

generally captured expected change and stability in scores across clinical samples.

This study also evaluated change in the PROMIS Physical Function scores in relation to global change anchors, both general health and domain-specific health. As hypothesized, positive and negative change on the anchors corresponded to positive and negative change on the PROMIS Physical Function measure in most samples. This was most clearly evident in the five samples for which we had global

Table 3. Change in PROMIS Physical Function by general health and physical function anchors

Subsample	Change in PROMIS physical function by general health anchor				Change in PROMIS physical function by general physical function anchor			
	n	SRM	T-score change	SD	n	SRM	T-score change	SD
Back pain								
Better	51	0.64	4.5	7.0	—	—	—	—
About the same	95	0.49	3.1	6.2	—	—	—	—
Worse	24	0.29	1.7	5.8	—	—	—	—
Cancer								
Better	85	0.42	2.3	7.1	74	0.44	2.5	5.8
About the same	129	0.00	0.0	5.6	114	0.08	0.5	5.9
Worse	50	–0.22	–1.6	5.6	54	–0.77	–4.5	5.7
Chronic heart failure								
Better	44	1.05	7.2	–6.9	43	1.20	7.8	–6.5
About the same	0	—	—	—	0	—	—	—
Worse	2	0.13	1.1	–8.3	3	–1.36	–4.3	–3.2
COPD—exacerbation								
Better	7	0.45	3.1	6.9	12	0.47	2.0	4.2
About the same	14	0.12	0.5	4.1	19	0.30	1.1	3.6
Worse	13	0.15	0.7	4.9	12	–0.38	–2.6	6.7
COPD—stable								
Better	13	0.70	2.5	3.5	20	0.11	0.4	3.4
About the same	40	0.02	0.1	4.4	31	0.07	0.3	4.1
Worse	14	–0.55	–1.5	2.7	18	0.02	0.1	4.9
Major depressive disorder								
Better	43	0.30	3.2	10.7	—	—	—	—
About the same	113	0.03	0.2	6.0	—	—	—	—
Worse	30	0.04	0.2	6.0	—	—	—	—
Rheumatoid arthritis								
Better	61	0.21	0.9	4.2	56	0.29	1.3	4.7
About the same	297	–0.12	–0.5	3.9	252	0.03	0.1	3.8
Worse	92	–0.19	–0.9	4.9	143	–0.46	–1.9	4.2

Abbreviations: COPD, chronic obstructive pulmonary disease; SRM, standardized response mean.

Positive change scores indicate better physical function.

Subsample sizes in this table do not add up to the baseline sample (Table 1) due to missing PROMIS PF or anchor data at follow-up.

physical function anchors. With the exception of COPD stable, change for these subgroups was typically medium to large for PROMIS Physical Function in the expected direction (SRMs ranging from 0.29 to 1.20).

For clinical groups anchored on global change in health, the effects generally followed the same pattern. Contrary to prediction, patients experiencing worsening health in two studies (back pain and COPD exacerbation) reported small improvements in PROMIS Physical Function. In both cases, however, the positive improvement in physical function was much higher in the better health than the worse health group (0.64 vs. 0.29 for back pain; 0.45 vs. 0.15 for COPD exacerbation), suggesting that the PROMIS instrument was nevertheless distinguishing these two subgroups. Finally, although the COPD-stable subgroups showed a sharp contrast in SRMs for better and worse health (0.70 vs. -0.55), the COPD-stable subgroups defined by the domain-specific anchor did not (0.11 vs. 0.02). The cause of this unexpected difference in SRMs between these two (using general vs. specific anchor) is unclear, but likely the result of sampling error, given the small subsamples ($n \leq 20$) in COPD.

The present study is not without limitations. First, although the use of global ratings of health or domain-specific change may be face valid and clinically relevant, they provide some methodological complications. Because retrospective ratings are assessed at follow-up, they are typically correlated more with follow-up (current) scores rather than pretest or change scores [34]. Second, sample sizes for COPD and CHF were modest; when anchored to different change groups, several of the subsamples for COPD and CHF were below 20. Consequently, results for these subgroups should be interpreted cautiously. In addition, the forms of the PROMIS PF instrument (CAT vs. SF) differed across groups in our study. Although existing research on CAT vs. SF suggests that scores obtained are highly comparable across forms [35], more research in multiple domains is needed to fully understand this, specifically focusing on longitudinal data and responsiveness. In addition, all PROMIS CAT forms of the instrument were based on the v1.0 bank, which contained 124 items. The current version of the bank (v1.2) contains 121 items (for details, see the [online appendix](#)). Finally, a key test of responsiveness is a comparison between established “legacy” instruments and the new PROMIS instruments. Although such an analysis has been reported on elsewhere [7] (using the same RA sample), future outcome studies need to incorporate multiple measures of similar constructs to enable this analysis.

The present study serves as an important step in the ongoing evaluation of the PROMIS Physical Function item bank. The examination of the PROMIS Physical Function item bank’s validity across six diverse clinical samples extended past psychometric evaluations of the measure by demonstrating that it is not only appropriately sensitive to change over time and in response to intervention but also able to meaningfully differentiate different clinical samples

on the basis of physical function scores. Thus, although validity is not a static concept, this study extends past research by examining the performance of the PROMIS Physical Function item bank in real-world clinical samples longitudinally. As such, the findings of the present study can be used to inform the incorporation of the PROMIS Physical Function item bank into future clinical research and practice of CHF, COPD, back pain, RA, and MDD groups. The findings can also inform definitions of treatment response and assist with interpretations of results in comparative effectiveness research.

Acknowledgments

PROMIS was funded with cooperative agreements from the National Institutes of Health (NIH) Common Fund Initiative (Northwestern University, PI: David Cella, PhD, U54AR057951, U01AR052177, R01CA60068; Northwestern University, PI: Richard C. Gershon, PhD, U54AR057943; American Institutes for Research, PI: Susan (San) D. Keller, PhD, U54AR057926; State University of New York, Stony Brook, PIs: Joan E. Broderick, PhD and Arthur A. Stone, PhD, U01AR057948, U01AR052170; University of Washington, Seattle, PIs: Heidi M. Crane, MD, MPH, Paul K. Crane, MD, MPH, and Donald L. Patrick, PhD, U01AR057954; University of Washington, Seattle, PI: Dagmar Amtmann, PhD, U01AR052171; University of North Carolina, Chapel Hill, PI: Harry A. Guess, MD, PhD (deceased), Darren A. DeWalt, MD, MPH, U01AR052181; Children’s Hospital of Philadelphia, PI: Christopher B. Forrest, MD, PhD, U01AR057956; Stanford University, PI: James F. Fries, MD, U01AR052158; Boston University, PIs: Alan Jette, PT, PhD, Stephen M. Haley, PhD (deceased), and David Scott Tulskey, PhD (University of Michigan, Ann Arbor), U01AR057929; University of California, Los Angeles, PIs: Dinesh Khanna, MD (University of Michigan, Ann Arbor) and Brennan Spiegel, MD, MSHS, U01AR057936; University of Pittsburgh, PI: Paul A. Pilkonis, PhD, U01AR052155; Georgetown University, PIs: Carol M. Moinpour, PhD (Fred Hutchinson Cancer Research Center, Seattle) and Arnold L. Potosky, PhD, U01AR057971; Children’s Hospital Medical Center, Cincinnati, PI: Esi M. Morgan DeWitt, MD, MSCE, U01AR057940; University of Maryland, Baltimore, PI: Lisa M. Shulman, MD, U01AR057967; and Duke University, PI: Kevin P. Weinfurt, PhD, U01AR052186). NIH Science Officers on this project have included Deborah Ader, PhD, Vanessa Ameen, MD (deceased), Susan Czajkowski, PhD, Basil Eldadah, MD, PhD, Lawrence Fine, MD, DrPH, Lawrence Fox, MD, PhD, Lynne Haverkos, MD, MPH, Thomas Hilton, PhD, Laura Lee Johnson, PhD, Michael Kozak, PhD, Peter Lyster, PhD, Donald Mattison, MD, Claudia Moy, PhD, Louis Quatrano, PhD, Bryce Reeve, PhD, William Riley, PhD, Peter Scheidt, MD, Ashley Wilder Smith, PhD,

MPH, Susana Serrate-Sztejn, MD, William Phillip Tonkins, DrPH, Ellen Werner, PhD, Tisha Wiley, PhD, and James Witter, MD, PhD. The contents of this article uses data developed under PROMIS. These contents do not necessarily represent an endorsement by the US Federal Government or PROMIS. See www.nihpromis.org for additional information on the PROMIS initiative.

Supplementary data

Supplementary data related to this article can be found at <http://dx.doi.org/10.1016/j.jclinepi.2015.08.039>.

References

- [1] Haley SM, Coster WJ, Binda-Sundberg K. Measuring physical disablement: the contextual Challenge. *Phys Ther* 1994;74:443–51.
- [2] Haley SM, McHorney CA, Ware JE Jr. Evaluation of the MOS SF-36 physical functioning scale (PF-10): I. Unidimensionality and reproducibility of the Rasch item scale. *J Clin Epidemiol* 1994;47:671–84.
- [3] Stewart AL, Kamberg C. Physical functioning. In: Stewart AL, Ware JE, editors. *Measuring functioning and well-being: the medical outcomes study approach*. Durham, NC: Duke University Press; 1992:86–142.
- [4] Wilson IB, Cleary PD. Linking clinical variables with health-related quality of life. A conceptual model of patient outcomes. *JAMA* 1995;273:59–65.
- [5] Fries JF, Bruce B, Bjorner J, Rose M. More relevant, precise, and efficient items for assessment of physical function and disability: moving beyond the classic instruments. *Ann Rheum Dis* 2006;65:16–21.
- [6] Cella D, Riley W, Stone A, Rothrock N, Reeve B, Yount S. 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.
- [7] Fries JF, Krishnan E, Rose M, Lingala B, Bruce B. Improved responsiveness and reduced sample size requirements of PROMIS physical function scales with item response theory. *Arthritis Res Ther* 2011;13(5):R147.
- [8] Stewart AL, Greenfield S, Hays RD, Wells K, Rogers WH, Berry SD. Functional status and well-being of patients with chronic conditions: results from the medical outcomes study. *JAMA* 1989;262:907–13.
- [9] Stewart AL, Hays RD, Wells KB, Rogers WH, Spritzer KL, Greenfield S. Long-term functioning and well-being outcomes associated with physical activity and exercise in patients with chronic conditions in the medical outcomes study. *J Clin Epidemiol* 1994;47:719–30.
- [10] Fries JF, Spitz PW, Young DY. The dimensions of health outcomes: the health assessment questionnaire, disability and pain scales. *J Rheumatol* 1982;9:789–93.
- [11] Bruce B, Fries JF. The Stanford Health Assessment Questionnaire: a review of its history, issues, progress, and documentation. *J Rheumatol* 2003;30:167–78.
- [12] Ware JE Jr, Sherbourne CD. The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. *Med Care* 1992;30:473–83.
- [13] Bruce B, Fries JF, Ambrosini D, Lingala B, Gandek B, Ware JE Jr. Better assessment of physical function: item improvement is neglected but essential. *Arthritis Res Ther* 2009;11(6):R191.
- [14] Rose M, Bjorner JB, Gandek B, Bruce B, Fries JF, Ware JE. The PROMIS Physical Function item bank was calibrated to a standardized metric and shown to improve measurement efficiency. *J Clin Epidemiol* 2014;67:516–26.
- [15] Hays RD, Spritzer KL, Amtmann D, Lai JS, DeWitt EM, Rothrock N. Upper extremity and mobility subdomains from the Patient-Reported Outcomes Measurement Information System (PROMIS) adult physical functioning item bank. *Arch Phys Med Rehabil* 2013;94:2291–6.
- [16] Walke LM, Gallo WT, Tinetti ME, Fried TR. The Burden of symptoms among community-dwelling older persons with advanced chronic disease. *Arch Intern Med* 2004;164:2321–4.
- [17] Cella D, Hahn EA, Dineen K. Meaningful change in cancer-specific quality-of-life scores: differences between improved and worsening. *Qual Life Res* 2002;11:207–21.
- [18] Anzueto A, Sethi S, Martinez FJ. Exacerbations of chronic obstructive pulmonary disease. *Proc Am Thorac Soc* 2007;4(7):554–64.
- [19] Rose M, Bjorner JB, Becker J, Fries JF, Ware JE. Evaluation of a preliminary physical function item bank supported the expected advantages of the Patient-Reported Outcomes Measurement Information System (PROMIS). *J Clin Epidemiol* 2008;61:17–33.
- [20] Hays RD, Spritzer KL, Fries JF, Krishnan E. Responsiveness and minimally important difference for the Patient-Reported Outcomes Measurement Information System (PROMIS) 20-item physical functioning short form in a prospective observational study of rheumatoid arthritis. *Ann Rheum Dis* 2015;74(1):104–7.
- [21] Yost KJ, Eton DT, Garcia SF, Cella D. Minimally important differences were estimated for six Patient-Reported Outcomes Measurement Information System–Cancer scales in advanced-stage cancer patients. *J Clin Epidemiol* 2011;64:507–16.
- [22] van der Linden WJ. Bayesian item selection criteria for adaptive testing. *Psychometrika* 1998;63(2):201–16.
- [23] Choi SW, Swartz RJ. Comparison of CAT item selection criteria for polytomous items. *Appl Psychol Meas* 2009;33(6):419–40.
- [24] Cook KF, Jensen SE, Schalet BD, Beaumont JL, Amtmann D, Czajkowski S. PROMIS measures of pain, fatigue, negative affect, physical function and social function demonstrate clinical validity across a range of chronic conditions. *J Clin Epidemiol* 2016; <http://dx.doi.org/10.1016/j.jclinepi.2015.08.038>. [Epub ahead of print].
- [25] Hedeker DR, Gibbons RD. *Longitudinal data analysis*. Hoboken, NJ: Wiley-Interscience; 2006.
- [26] Verbeke G, Molenberghs G. *Linear mixed models for longitudinal data*. New York, NY: Springer-Verlag; 2000.
- [27] Troxel AB, Fairclough DL, Curran D, Hahn EA. Statistical analysis of quality of life with missing data in cancer clinical trials. *Stat Med* 1998;17:653–66.
- [28] Little RJA, Rubin DB. *Statistical analysis with missing data*. Hoboken, NJ: John Wiley & Sons, Inc.; 2002.
- [29] Liang MH, Fossel AH, Larson MG. Comparisons of five health status instruments for orthopedic evaluation. *Med Care* 1990;28:632–42.
- [30] Fayers PM, Machin D. *Quality of life: the assessment, analysis and interpretation of patient-reported outcomes*. 2nd ed. Chichester, England; Hoboken, NJ: John Wiley & Sons; 2007.
- [31] Cohen J. *Statistical power analysis for the behavioral sciences*. 2nd ed. Hillsdale, NJ: L. Erlbaum Associates; 1988.
- [32] Yost KJ, Eton DT. Combining distribution- and anchor-based approaches to determine minimally important differences: the FACIT experience. *Eval Health Prof* 2005;28(2):172–91.
- [33] Liu H, Cella D, Gershon R, Shen J, Morales LS, Riley W. Representativeness of the PROMIS Internet panel. *J Clin Epidemiol* 2010;63:1169–78.
- [34] Norman GR, Stratford P, Regehr G. Methodological problems in the retrospective computation of responsiveness to change: the lesson of Cronbach. *J Clin Epidemiol* 1997;50:869–79.
- [35] Choi S, Reise S, Pilkonis P, Hays RD, Cella D. Efficiency of static and computer adaptive short forms compared to full-length measures of depressive symptoms. *Qual Life Res* 2010;19:125–36.