

SHORT REPORT

Open Access



# Development and validation of an interpretive guide for PROMIS scores

Nan E. Rothrock<sup>1\*</sup> , Dagmar Amtmann<sup>2</sup> and Karon F. Cook<sup>1</sup>

## Abstract

**Background:** Accurate score interpretation is required for the appropriate use of patient-reported outcome measures in clinical practice.

**Objective:** To create and evaluate figures (T-score Maps) to facilitate the interpretation of scores on Patient-Reported Outcome Measurement Information System (PROMIS) measures.

**Methods:** For 21 PROMIS® short forms, item-level information was used to predict the most probable responses to items for the range of possible scores on each short form. Predicted responses were then “mapped” graphically along the range of possible scores. In a previously conducted longitudinal study, 1594 adult participants with chronic conditions (e.g., multiple sclerosis) responded to four items each of a subset of these PROMIS short forms. Participants’ responses to these items were compared to those predicted by the T-score Maps. Difference scores were calculated between observed and predicted scores, and Spearman correlations were calculated.

**Results:** We constructed T-score Maps for 21 PROMIS short forms for adults and pediatric self- and parent-proxy report. For the clinical population, participants’ actual responses were strongly correlated with their predicted responses ( $r = 0.762$  to  $0.950$ ). The majority of predicted responses exactly matched observed responses (range 69.5% to 85.3%).

**Conclusion:** Results support the validity of the predicted responses used to construct T-score Maps. T-score Maps are ready to be tested as interpretation aids in a variety of applications.

**Keywords:** Patient-reported outcomes, PROMIS, Item response theory

## Introduction

Patient-reported outcome (PRO) measures are increasingly integrated into routine clinical practice to inform clinical decision making [1–3], monitor or screen for symptoms [4, 5], or meet treatment guidelines [6]. In order to base treatment decisions on the PRO scores, providers must be able to accurately interpret their resultant scores. Although guidance on score interpretation was identified by experts as a required component of implementation of PROs in clinical practice [7], a recent systematic review found that only 39% of oncology implementations included it [8]. Approaches to facilitate score interpretation have included identification of important severity thresholds [9–12] and construction of population-based norms reference data [13, 14].

Attributes of the Patient-Reported Outcome Measurement Information System® (PROMIS®) item banks offer potential to create new PRO score interpretation tools. First, in addition to being psychometrically sound [15], PROMIS item banks were developed to reflect how patients conceptualize important symptoms and functions as they apply in one’s day-to-day life. In developing these measures, investigators used mixed methods with substantial patient input [16]. This included identification of important components of a symptom or function to be assessed, as well as reliable and accurate interpretation of the meaning of items across patients [17, 18]. Second, PROMIS measures were constructed with item response theory (IRT) [15, 19]. In IRT, the most likely response to an item can be identified for each score. For example, patients with very poor function are most likely to respond “unable to do” for an item such as, “Are you able to run a short distance such as to catch a bus?” whereas patients with exceptional function are most likely to respond

\* Correspondence: [n-rothrock@northwestern.edu](mailto:n-rothrock@northwestern.edu)

<sup>1</sup>Department of Medical Social Sciences, Feinberg School of Medicine of Northwestern University, 625 N. Michigan Ave Suite 2700, Chicago, IL 60660, USA

Full list of author information is available at the end of the article

“without any difficulty.” For each item in an IRT-calibrated item bank, a *most likely* response can be identified for each level of the domain measured. This attribute of IRT-calibrated item banks has been used to construct vignettes comprised of subsets of items and responses reflecting different levels of severity [9]. Patients and clinicians have been successful in rank ordering these vignettes, supporting their validity as a tool to convey severity [10–12].

We used IRT-predicted responses for PROMIS item banks to construct figures (“T-score Maps”) that display the most likely responses for a subset of items. This translates numeric scores into language used by patients to describe their degree of severity or impairment in a given symptom or function. Then, we compared the IRT-predicted responses with actual responses in a de-identified archival clinical dataset. We hypothesized that IRT-predicted responses would correlate strongly with patients’ responses ( $r > 0.70$ ) and that the majority of actual responses would be the same as those predicted. We explore potential applications of these figures to facilitate PRO measure score interpretation.

## Methods

### Development of T-score maps

PROMIS measures generate T-scores. T-scores are standard scores with a mean of 50 and standard deviation of 10 in a reference population (usually U.S. general population). T-score Maps were constructed for 21 PROMIS short forms that comprise the PROMIS-57 Profile v2.1, PROMIS Pediatric-49 Profile v2.0, and PROMIS Parent Proxy-49 Profile v2.0 [20]. The profiles reflect multiple domains of health relevant across the general population and people with chronic conditions, and include highly informative items across mild to severe levels of symptoms and dysfunction. Domains include anxiety, depression, fatigue, physical function, pain interference, sleep disturbance, and social function. Longer short forms (7–10 items) were used in order to represent varied content, allow greater measurement specificity, and be printable on a single page. PROMIS items consist of a statement (e.g., “I feel fatigued”) with five response options (e.g., 1 = not at all, 2 = a little bit, 3 = somewhat, 4 = quite a bit, 5 = very much).

All PROMIS measures were previously calibrated using unidimensional IRT models for each domain [15, 19]. We used the item parameters derived in these calibrations to identify the most probable responses based on the item characteristic curves (ICCs) for each item. ICCs are probability curves that display the probabilities of each response as a function of respondents’ scores on the domain being measured; they are mathematically generated from the IRT model. In ICC plots, probability is plotted on the  $y$ -axis and scores are plotted on the  $x$ -axis. For any score on  $x$ , the response curve with the

highest value of  $y$  is the most probable response. We wrote computer code to identify these most probable responses by score. The code was written using the R program language [21] and is available from the authors. Note that although a response may be the most probable at a given level of severity, this does not necessarily mean that it has a very high probability. A person with a T-score of 60 on PROMIS Anxiety, for example, would have the following response probabilities ( $p$ ) for the item, “My worries overwhelmed me”: never,  $p = 0.089$ ; rarely,  $p = 0.442$ ; sometimes,  $p = 0.415$ ; often,  $p = 0.052$ ; and always,  $p = 0.002$ . The most likely response is “rarely” but there is an almost equal probability of answering “sometimes”. For a T-score of 61, the response of “sometimes” is the most likely response (never,  $p = 0.063$ ; rarely,  $p = 0.376$ ; sometimes,  $p = 0.484$ ; often,  $p = 0.073$ ; and always,  $p = 0.003$ ). Thus, the most probable response changes from “rarely” to “sometimes” between the T-scores of 60 and 61.

Once the most likely responses at each level of symptom severity or function were obtained for items in the 21 short forms, the results were “mapped” onto the PROMIS T-score continuum in a figure. Specifically, a band for each response option was constructed to indicate the range of scores for which it was the most likely response.

### Comparison of predicted and observed responses

#### Data

Scores predicted by ICCs were compared with observed responses in a de-identified archival clinical dataset. Data came from a survey of adults aging with muscular dystrophy, multiple sclerosis, post-polio syndrome, or spinal cord injury [22]. Individuals living with one of these chronic conditions completed a mailed self-report symptom survey every year for 7 years. Cross-sectional data from year 4 (collected 2012–2013) were used for this secondary analysis because they included the largest sample size for the domains of interest. The dataset included PROMIS v1.0 Fatigue, Anxiety, Depression, and Pain Interference 4a Short Forms (all of which comprise 4 items each). All items in 4a short forms are also included in the short forms displayed in the T-score Map. Of the 1814 surveys mailed, 1594 individuals (88%) completed it. Participants received \$25 for completing the survey. All research participants provided informed consent and all study procedures were approved by the University of Washington Human Subjects Division.

#### Analyses

We conducted descriptive analyses to evaluate the degree to which predicted responses matched responses observed in the clinical data. For every participant in the clinical study, we calculated PROMIS T-scores for

Fatigue, Anxiety, Depression, and Pain Interference based on their responses to the four administered items of each measure. These T-scores were then located on the appropriate T-score Map. We identified the predicted item response for each item associated with the calculated T-score. We then obtained “difference scores” by subtracting the number associated with their predicted response (1 to 5) from the number associated with their observed response (1 to 5). For example, an individual with a PROMIS Anxiety Score of 60 is predicted to respond “rarely” to, “My worries overwhelmed me.” A response of “rarely” has a numerical value of 2. A respondent who answered “sometimes” (response value of 3), would have a difference score for this item of + 1. Respondents with a T-score of 60 on Anxiety who answered “never” (response value of 1), would have a difference score of - 1. In addition, we calculated the Spearman Correlation Coefficient between predicted and observed responses for each of the 16 items targeted in the study.

**Results**

**T-score maps**

We constructed 21 T-score Maps for adult, pediatric, and parent-proxy PROMIS short forms (see Fig. 1). For a given short form, each item was displayed underneath a ruler showing the PROMIS T-score metric. The ranges in which

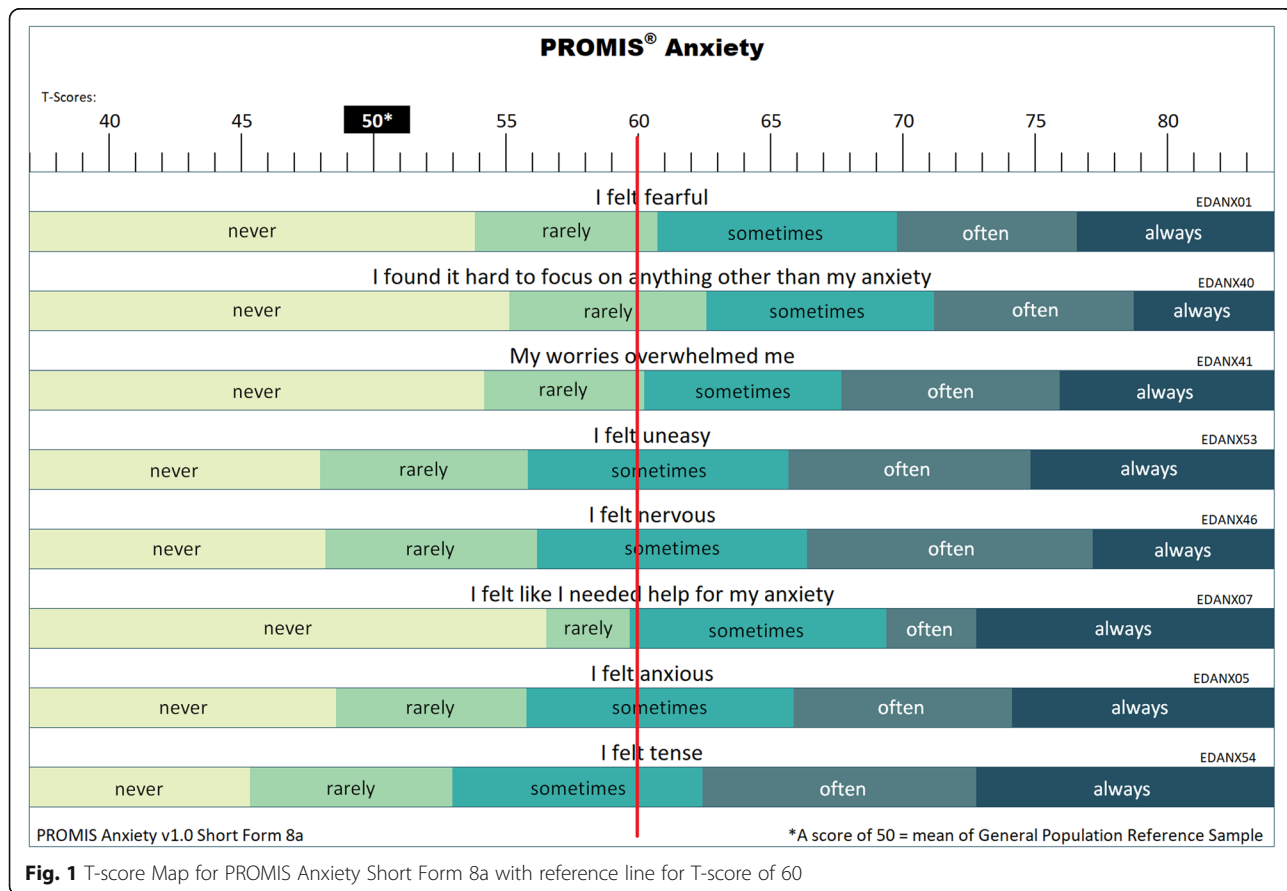
each response category was the most likely response were displayed as shaded bands. As the Fig. 1 Map shows, at T = 60, the most likely response to the item “My worries overwhelmed me” is “rarely;” the most likely response to the item “I felt uneasy” is “sometimes.” All T-score Maps are available at <http://www.healthmeasures.net/score-and-interpret/interpret-scores/promis/t-score-maps>.

**Sample characteristics**

The mean age of the clinical sample was 59.3 years (SD = 13.0), with a mean time since diagnosis of 29.0 years (SD = 21.6). Participants were primarily female (63.8%), non-Hispanic white (91.2%), and had received a college degree or greater (56.7%; Table 1).

**Comparison of predicted and observed responses**

The majority of predicted responses matched the observed responses for each of the 16 items and were consistent across the 4 domains: Fatigue (70.8% to 81.3%), Anxiety (69.5% to 82.0%), Depression (70.5% to 84.9%), and Pain Interference (78.2% to 85.3%). In cases where participants did not select the predicted response, they usually selected the adjacent response reflecting more severity (6.0% to 20.8%) or the adjacent response reflecting less severity (2.5% to 17.1%). These findings were



**Table 1** Participant Characteristics

Participant Characteristics (n = 1594)	
	Mean [SD]
Participant age (years)	59.3 [13.0]
Years since diagnosis	29.0 [21.6]
PROMIS v1.0 Short Form 4a T-scores	
Fatigue	55.4 [10.4]
Anxiety	51.7 [8.9]
Depression	50.1 [8.8]
Pain Interference	55.0 [9.7]
	n (%)
Diagnosis	
Multiple sclerosis	509 (31.9)
Muscular dystrophy	282 (17.7)
Post-polio syndrome	389 (24.4)
Spinal cord injury	414 (26.0)
Gender	
Female	1017 (63.8)
Male	576 (36.1)
Did not respond	1 (0.1)
Race/ethnicity	
Non-Hispanic White	1454 (91.2)
Other	128 (8.0)
Did not respond	12 (0.8)
Education	
Some high school or less	23 (1.4)
High school grad/GED	189 (11.9)
Some college/vocational or technical degree	476 (29.9)
College degree	499 (31.3)
Graduate or professional degree	405 (25.4)
Did not respond	2 (0.1)

consistent across domains. The IRT-predicted responses displayed in the T-score Maps were strongly correlated with participants' actual responses to PROMIS short form items ( $r = 0.762$  to  $0.950$ , see Table 2). A higher bar to consider is the number of participants whose predicted responses perfectly matched their observed responses across all items of a short form. This level of congruence occurred about half the time with 51.7%, 42.6%, 47.3%, and 55.2% of Fatigue, Anxiety, Depression, and Pain Interference responses matching perfectly across all items of a scale.

## Discussion

PROMIS T-score Maps were constructed for 21 short forms. Each Map displays the most likely responses for possible measure scores. In a follow-up study, predicted responses for a subset of items were compared to

responses observed for these items in a clinical dataset and were found to be strongly correlated. This supports the validity of the predicted responses.

Because T-score Maps transform a numeric value to a series of statements about the real-world experience of a symptom or function, they have multiple potential applications. First, they may aid in conveying the meaning of a mean or range of outcomes for various treatments. For example, a clinical trial may identify mean scores for control and intervention groups (e.g.,  $T = 61$  versus  $T = 53$ ). Using Anxiety as an example, with a T-score Map this difference can be conveyed as a "My worries *sometimes* overwhelmed me" to "My worries *never* overwhelmed me." A clinician and patient can use this information to better understand the expected outcome of a given intervention and inform treatment decisions. A second potential application is to use a T-score Map to set a threshold (e.g., for inclusion in a study, for clinical action). For example, in oncology, collecting PROs for emotional distress is part of standard care. Guidelines state that patients with moderate or severe distress should be provided appropriate referrals for care [23]. T-score Maps for depression and anxiety short forms could be used by mental health experts to aid in identifying thresholds an organization should utilize for referrals. Third, T-score Maps could be utilized as a tool for setting goals for care. For example, a physical therapist may ask patients to identify what level of function the patient hopes to achieve by the end of treatment on a T-score Map. Short form items may be particularly helpful in achieving consensus on treatment expectations because of their ability to convey a range of intensity (e.g., without any difficulty, with a little difficulty, with some difficulty, with much difficulty, unable to do) through their response options. Finally, using T-score Maps to compare two scores could be a helpful tool in creating new methods for identifying what amount of change is meaningful to patients.

This study has three notable limitations. First, the de-identified archival clinical dataset only included four domains (fatigue, anxiety, depression, pain interference) that overlapped with the T-score Map domains. All were adult measures. Although the concordance between IRT-predicted and actual responses was consistent across domains, the extent to which our findings can be generalized to other adult domains or pediatric and parent proxy respondents is untested. Second, the T-score Maps were constructed using primarily 8-item short forms whereas the de-identified archival clinical dataset included 4-item short forms. Although all 4 items were included in the longer short form and the patterns of predicted and actual responses were consistent across items, the extent to which other items from an item

**Table 2** Differences scores (observed - predicted response category) and Spearman correlations between observed and predicted responses

Differences Scores (Observed Response Category - Predicted Response Category) for 16 PROMIS Items and Spearman Correlations between Observed and Predicted Responses

PROMIS Items	Response Category Differences		Item Content and Correlations							
	Difference	N	Percent	N	Percent	N	Percent	N	Percent	
Fatigue	Observed Response Category - Predicted Response Category		I feel fatigued $r = .906$		I have trouble starting things because I am tired $r = .874$		How fatigued were you on average? $r = .931$		How run-down did you feel on average? $r = .921$	
	-4	0	0.0	2	0.1	0	0.0	0	0.0	
	-3	0	0.0	5	0.3	0	0.0	1	0.1	
	-2	4	0.3	51	3.2	0	0.0	4	0.3	
	-1	164	10.3	238	15.0	62	3.9	113	7.1	
	<b>0</b>	<b>1174</b>	<b>74.0</b>	<b>1123</b>	<b>70.8</b>	<b>1272</b>	<b>80.1</b>	<b>1290</b>	<b>81.3</b>	
	1	240	15.1	163	10.3	251	15.8	175	11.0	
	2	5	0.3	5	0.3	3	0.2	4	0.3	
	3	0	0.0	0	0.0	0	0.0	0	0.0	
	4	0	0.0	0	0.0	0	0.0	0	0.0	
	Total N	1587		1587		1588		1587		
Anxiety	Difference		I felt fearful $r = 0.762$		I found it hard to focus on anything other than my anxiety $r = 0.849$		My worries overwhelmed me $r = 0.814$		I felt uneasy $r = 0.848$	
	-4	0	0.0	0	0.0	0	0.0	0	0.0	
	-3	1	0.1	0	0.0	0	0.0	0	0.0	
	-2	7	0.4	1	0.1	2	0.1	7	0.4	
	-1	132	8.3	53	3.3	95	6.0	271	17.1	
	<b>0</b>	<b>1103</b>	<b>69.5</b>	<b>1302</b>	<b>82.0</b>	<b>1238</b>	<b>78.0</b>	<b>1181</b>	<b>74.4</b>	
	1	330	20.8	209	13.2	232	14.6	123	7.7	
	2	15	0.9	22	1.4	20	1.3	5	0.3	
	3	0	0.0	1	0.1	0	0.0	1	0.1	
	4	0	0.0	0	0.0	1	0.1	0	0.0	
	Total N	1588		1588		1588		1588		
Depression	Difference		I felt worthless $r = 0.847$		I felt helpless $r = 0.798$		I felt depressed $r = 0.846$		I felt hopeless $r = 0.826$	
	-4	0	0.0	0	0.0	0	0.0	0	0.0	
	-3	0	0.0	0	0.0	0	0.0	0	0.0	
	-2	12	0.8	5	0.3	4	0.3	1	0.1	
	-1	130	8.2	66	4.2	154	9.7	40	2.5	
	<b>0</b>	<b>1349</b>	<b>84.9</b>	<b>1212</b>	<b>76.3</b>	<b>1120</b>	<b>70.5</b>	<b>1323</b>	<b>83.3</b>	
	1	96	6.0	256	16.1	302	19.0	211	13.3	
	2	2	0.1	47	3.0	8	0.5	13	0.8	
	3	0	0.0	2	0.1	0	0.0	0	0.0	
	4	0	0.0	0	0.0	0	0.0	0	0.0	
	Total N	1589		1588		1588		1588		

**Table 2** Differences scores (observed - predicted response category) and Spearman correlations between observed and predicted responses (*Continued*)

Differences Scores (Observed Response Category - Predicted Response Category) for 16 PROMIS Items and Spearman Correlations between Observed and Predicted Responses

PROMIS Items	Response Category Differences	Item Content and Correlations							
		How much did pain interfere with your day to day activities? $r = 0.922$		How much did pain interfere with work around the home? $r = 0.950$		How much did pain interfere with your ability to participate in social activities? $r = 0.916$		How much did pain interfere with your household chores? $r = 0.920$	
	Difference	N	Percent	N	Percent	N	Percent	N	Percent
	-4	0	0.0	0	0.0	0	0.0	0	0.0
	-3	0	0.0	0	0.0	0	0.0	0	0.0
	-2	0	0.0	2	0.1	6	0.4	7	0.4
	-1	114	7.2	104	6.6	136	8.7	167	10.7
	<b>0</b>	<b>1253</b>	<b>79.5</b>	<b>1334</b>	<b>85.3</b>	<b>1230</b>	<b>78.2</b>	<b>1222</b>	<b>78.2</b>
	1	200	12.7	121	7.7	185	11.8	151	9.7
	2	9	0.6	3	0.2	13	0.8	14	0.9
	3	1	0.1	0	0.0	2	0.1	1	0.1
	4	0	0.0	0	0.0	0	0.0	0	0.0
	Total N	1577		1564		1572		1562	

bank would produce similar results is untested. Finally, all observed responses were provided by individuals with chronic conditions. Additional comparisons with other samples, particularly those with more emotional health concerns, would clarify the generalizability of our results.

In conclusion, the need for aids in interpreting the meaning of PRO scores is significant. T-score Maps are ready to be tested as interpretation aids in a variety of applications. T-score Maps need not be limited to 4 items and, in fact, those developed for [HealthMeasures.net](http://www.healthmeasures.net) include 7–10 items. T-score Maps that showed predicted responses for all items would be unwieldy because of the number of items that comprise item banks. An interesting line of future study would be to identify items of most relevance to particular patient populations and target these in developing T-score Maps.

#### Abbreviations

IRT: Item response theory; PRO: Patient-reported outcome; PROMIS: Patient-reported Outcomes Measurement Information System

#### Acknowledgements

The authors would like to thank Rana Salem for generating the de-identified dataset with measure scores utilized for this study.

#### Authors' contributions

Conception and development of the manuscript (all authors); data analysis (KC); data interpretation and manuscript preparation (all authors). All authors read and approved the final manuscript.

#### Funding

Generating and evaluating T-score maps was supported by a grant from the National Cancer Institute (U2C CA186878). The initial data collection that generated the de-identified archival dataset used to evaluate T-score Maps was supported in part by grant number 90RT5023-01-00, from the National Institute on Disability, Independent Living, and Rehabilitation Research

(NIDILRR). NIDILRR is a Center within the Administration for Community Living (ACL), Department of Health and Human Services (HHS).

#### Availability of data and materials

The dataset used in this study is available as a supplemental file. All PROMIS T-score Maps are available at <http://www.healthmeasures.net/score-and-interpret/interpret-scores/promis/t-score-maps>.

R code used to generate response probabilities is available from the authors.

#### Ethics approval and consent to participate

Data collection was approved by the University of Washington Human Subjects Institutional Review Board. This work utilized a de-identified dataset.

#### Consent for publication

Not applicable.

#### Competing interests

The authors declare that they have no competing interests.

#### Author details

<sup>1</sup>Department of Medical Social Sciences, Feinberg School of Medicine of Northwestern University, 625 N. Michigan Ave Suite 2700, Chicago, IL 60660, USA. <sup>2</sup>Department of Rehabilitation Medicine, University of Washington, Seattle, WA, USA.

Received: 9 September 2019 Accepted: 14 February 2020

Published online: 28 February 2020

#### References

- Baumhauer, J. F. (2017). Patient-reported outcomes—Are they living up to their potential? *The New England Journal of Medicine*, 377(1), 6–8.
- Gerhardt, W. E., Mara, C. A., Kudel, I., Morgan, E. M., Schoettker, P. J., Napora, J., et al. (2018). Systemwide implementation of patient-reported outcomes in routine clinical care at a children's hospital. *Joint Commission Journal on Quality and Patient Safety*, 44(8), 441–453.
- Biber, J., Ose, D., Reese, J., Gardiner, A., Facelli, J., Spuhl, J., et al. (2018). Patient reported outcomes—experiences with implementation in a university health care setting. *Journal of Patient-Reported Outcomes*, 2(1), 34.
- Basch, E., Deal, A. M., Kris, M. G., Scher, H. I., Hudis, C. A., Sabbatini, P., et al. (2015). Symptom monitoring with patient-reported outcomes during

- routine cancer treatment: A randomized controlled trial. *Journal of Clinical Oncology*, 34(6), 557–565.
5. Wagner, L. I., Schink, J., Bass, M., Patel, S., Diaz, M. V., Rothrock, N., et al. (2015). Bringing PROMIS to practice: Brief and precise symptom screening in ambulatory cancer care. *Cancer*, 121(6), 927–934.
  6. Singh, J. A., Saag, K. G., Bridges Jr., S. L., Akl, E. A., Bannuru, R. R., Sullivan, M. C., et al. (2016). 2015 American College of Rheumatology Guideline for the treatment of rheumatoid arthritis. *Arthritis & Rheumatology*, 68(1), 1–26.
  7. Chan, E. K. H., Edwards, T. C., Haywood, K., Mikles, S. P., & Newton, L. (2018). Implementing patient-reported outcome measures in clinical practice: A companion guide to the ISOQOL user's guide. *Quality of Life Research*. <https://doi.org/10.1007/s11136-018-2048-4>.
  8. Anatchkova, M., Donelson, S. M., Skalicky, A. M., McHorney, C. A., Jagun, D., & Whiteley, J. (2018). Exploring the implementation of patient-reported outcome measures in cancer care: Need for more real-world evidence results in the peer reviewed literature. [journal article]. *Journal of Patient-Reported Outcomes*, 2(1), 64.
  9. Cook, K. F., Reeve, B. B., & Cella, D. (2019). PRO-bookmarking to estimate clinical thresholds for patient-reported symptoms and function. *Medical Care*, 57(Suppl 5), S13–S17.
  10. Cook, K. F., Victorson, D. E., Cella, D., Schalet, B. D., & Miller, D. (2015). Creating meaningful cut-scores for Neuro-QOL measures of fatigue, physical functioning, and sleep disturbance using standard setting with patients and providers. *Quality of Life Research*, 24(3), 575–589.
  11. Nagaraja, V., Mara, C., Khanna, P. P., Namas, R., Young, A., Fox, D. A., et al. (2018). Establishing clinical severity for PROMIS® measures in adult patients with rheumatic diseases. *Quality of Life Research*, 27(3), 755–764.
  12. Cella, D., Choi, S., Garcia, S., Cook, K. F., Rosenbloom, S., Lai, J.-S., et al. (2014). Setting standards for severity of common symptoms in oncology using the PROMIS item banks and expert judgment. *Quality of Life Research*, 23(10), 2651–2661.
  13. Paradowski, P. T., Bergman, S., Sundén-Lundius, A., Lohmander, L. S., & Roos, E. M. (2006). Knee complaints vary with age and gender in the adult population. Population-based reference data for the knee injury and osteoarthritis outcome score (KOOS). *BMC Musculoskeletal Disorders*, 7, 38.
  14. Hays, R. D., Spritzer, K. L., Thompson, W. W., & Cella, D. (2015). U.S. general population estimate for “excellent” to “poor” self-rated health item. *Journal of General Internal Medicine*, 30(10), 1511–1516.
  15. Reeve, B. B., Hays, R. D., Bjorner, J. B., Cook, K. F., Crane, P. K., Teresi, J. A., Thissen, D., Revicki, D. A., Weiss, D. J., & Hambleton, R. K. (2007). Psychometric evaluation and calibration of health-related quality of life item banks: Plans for the patient-reported outcomes measurement information system (PROMIS). *Medical Care*, 45(5), S22–S31.
  16. Cella, D., Riley, W., Stone, A., Rothrock, N., Reeve, B., Yount, S., et al. (2010). The patient-reported outcomes measurement information system (PROMIS) developed and tested its first wave of adult self-reported health outcome item banks: 2005–2008. *Journal of Clinical Epidemiology*, 63(11), 1179–1194.
  17. Irwin, D. E., Varni, J. W., Yeatts, K., & DeWalt, D. A. (2009). Cognitive interviewing methodology in the development of a pediatric item bank: A patient reported outcomes measurement information system (PROMIS) study. *Health and Quality of Life Outcomes*, 7, 3.
  18. DeWalt, D. A., Rothrock, N., Yount, S., & Stone, A. A. (2007). Evaluation of item candidates: The PROMIS qualitative item review. *Medical Care*, 45(5 Suppl 1), S12–S21.
  19. Hansen, M., Cai, L., Stucky, B. D., Tucker, J. S., Shadel, W. G., & Edelen, M. O. (2013). Methodology for developing and evaluating the PROMIS® smoking item banks. *Nicotine & Tobacco Research*, 16(Suppl 3), S175–S189.
  20. Cella, D., Choi, S. W., Condon, D. M., Schalet, B., Hays, R. D., Rothrock, N. E., et al. (2019). PROMIS® adult health profiles: Efficient short-form measures of seven health domains. *Value in Health*, 22(5), 537–544.
  21. Team, R. C. (2008). *R: A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing.
  22. Battalio, S. L., Jensen, M. P., & Molton, I. R. (2019). Secondary health conditions and social role satisfaction in adults with long-term physical disability. *Health Psychology*, 38, 445–454.
  23. Commission on Cancer. (2015). *Cancer Program Standards: Ensuring Patient-Centered Care* (2016th ed.). Chicago: American College of Surgeons.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Submit your manuscript to a SpringerOpen® journal and benefit from:**

- Convenient online submission
- Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

---

Submit your next manuscript at ► [springeropen.com](https://www.springeropen.com)

---