



Published in final edited form as:

*Arthritis Care Res (Hoboken)*. 2017 March ; 69(3): 393–402. doi:10.1002/acr.22937.

## Patient Reported Outcomes Measurement Information System® (PROMIS®) Tools for Collecting Patient-Reported Outcomes in Children with Juvenile Arthritis

Timothy G. Brandon, MPH<sup>1</sup>, Brandon D. Becker, MPH<sup>2</sup>, Katherine B. Bevans, PhD<sup>2,3</sup>, and Pamela F. Weiss, MD, MSCE<sup>1,3</sup>

<sup>1</sup>Division of Rheumatology, Children's Hospital of Philadelphia, Philadelphia, PA

<sup>2</sup>Division of General Pediatrics, Children's Hospital of Philadelphia, Philadelphia, PA

<sup>3</sup>University of Pennsylvania Perelman School of Medicine, Philadelphia, Pennsylvania

### Abstract

**Objective**—To evaluate the precision and construct validity of pediatric Patient Reported Outcomes Measurement Information System® (PROMIS®) instruments in a population of juvenile idiopathic arthritis (JIA) patients and parent proxies.

**Methods**—A convenience sample of JIA patients and parents of JIA patients completed PROMIS instruments for eight domains: anger, anxiety, depressive symptoms, fatigue, mobility, pain interference, peer relationships, and upper extremity function. Short form and computerized adaptive test (CAT) scores were derived from item bank responses. Raw scores were translated to standardized T-scores with corresponding standard errors (SEs). Discrimination between inactive versus active disease was evaluated as an indicator of each measure's construct validity. SEs were plotted to evaluate each instrument's relative precision. Patient-parent concordance was assessed using intraclass correlations (ICC).

**Results**—228 patients and 223 parents participated, providing 71–78 responses per domain. Patient- and parent-reported anger, fatigue, mobility, and pain interference scores significantly differed between those with inactive and active disease. Anxiety, depressive symptoms, and peer relationships differed by disease activity levels for parent-report only. Short forms and CATs provided comparable reliability to the full item banks across the full range of each outcome. Patient-parent agreement ranged from ICC=0.3 to 0.8. CAT did not reduce the number of items for any domain compared to the short form.

**Conclusion**—Precision and discriminatory abilities of PROMIS instruments depend on health domain and report type (self-report versus parent proxy-report) for children with JIA. Varying levels of patient-parent concordance reinforces the importance of considering both perspectives in comprehensive health outcomes assessments.

---

Child/parent perspectives on their own/their child's health are highly pertinent to disease management. Physician and child or parent ratings of pain(1), global disease severity(2), and

inactive disease(3) are often discordant in juvenile idiopathic arthritis (JIA). These discrepancies highlight that health professionals often perceive the health status of children with JIA differently than patients and their families(4, 5). As such, there has been increasing attention to integrating patient reported outcomes (PROs) into routine clinical care and research. The growing importance of PRO assessment in clinical research is highlighted by the Food and Drug Administration mandate to use PROs to support medical product labeling claims(6), the US government's Patient Protection and Affordable Care Act that created the Patient-Centered Outcomes Research Institute(7), and the creation of the NIH Patient Reported Outcomes Measurement Information System® (PROMIS®) initiative(8). Existing PRO tools for JIA include questionnaires focused on function/health related quality of life and visual analogue scales (VAS) for the assessment of overall well-being and pain intensity(9–16). Currently, the only multidimensional PRO tool for JIA is the Juvenile Arthritis Multidimensional Assessment Report, a validated tool that includes items that assess well-being, pain, function, quality of life, morning stiffness, disease activity, medication side effects, and overall satisfaction(17). These items represent a combination of legacy PRO instruments that were designed for routine clinical use, not research.

PROMIS is a collection of patient-reported health status tools available for children and adults that were developed to be disease non-specific(18). Children aged 8–17 years can complete self-report instruments and parents of children aged 5–17 years can complete parent proxy-report instruments. These tools can be administered to healthy children as well as to children with a variety of chronic health conditions. At study inception, the available pediatric PROMIS domains included those assessing anger(19), anxiety(20), depressive symptoms(20), fatigue(21), mobility(22), pain interference(23), peer relationships(24), and upper extremity function(22). Each PROMIS domain is composed of a collection of purposefully assembled questions called an “item bank” that, as a group, encompass the full range of the trait being measured. Items in an item bank are calibrated on a common scale for comparability across different populations with varying degrees of the trait. Item banks are used to derive separate assessment options that do not require all questions in a domain to be asked called short forms and computerized adaptive tests (CATs). A “short form” is a static selection of items from an item bank that represent a domain's item bank with fewer questions. The benefit of short forms is they can be administered with or without a computer. A “CAT” is a flexible option that enables the answer from one question to inform the choice of the most informative next question. Therefore, each child completing a CAT instrument could conceivably answer a distinct set of questions for a particular domain to arrive at their score. CATs require participants to have access to a computer but are designed to provide more precise measurement than short forms. PROMIS uses item response theory, which allows the validation of the item bank to be applied to all combinations of subsets of its items, creating a great deal of flexibility when selecting an instrument for use.

All raw scores generated from PROMIS instruments are translated into standardized T-scores with a population mean of 50 and standard deviation of 10. The population mean refers to the mean of the calibration sample, which, for pediatric and parent proxy instruments, is composed of a higher percentage of patients with chronic illness. It is important to note that higher scores in a domain represent more of the trait being measured. In this study, a higher T-score indicates a worse outcome in the following domains: anger,

anxiety, depressive symptoms, fatigue, and pain interference. Lower T-scores indicate a worse outcome in the remaining domains: mobility, peer relationships, and upper extremity function.

This is the first clinical validation study of multiple PROMIS instruments involving children with JIA. The use of PRO measures adds value to physician-based instruments as they assess a broad range of outcomes that are highly valued and best reported by patients, but not routinely considered by physicians. The assessment of these outcomes over time in children with JIA will add breadth to the overall assessment and care plan and may enlighten pediatric rheumatologists about areas (e.g. fatigue, anxiety) that need to be addressed through clinical care.

## **MATERIALS AND METHODS**

### **Human subjects protection**

The protocol for the conduct of this study was reviewed and approved by the Children's Hospital of Philadelphia Committee for the Protection of Human Subjects.

### **Study population**

This was a cross-sectional survey of children with JIA and parents of children with JIA. Participants were a convenience sample enrolled during routine rheumatology clinic visits between April 2012 and October 2014. Children eligible to complete PROMIS patient-report instruments were 8–17 years old and had JIA according to the International League of Associations for Rheumatology criteria(25). Parents or guardians of children 5–17 years old with an existing JIA diagnosis were eligible to complete the PROMIS parent-report item banks. Children or parents who were non-English speaking or reading, or who had a developmental delay, were excluded.

### **Data collection and survey instruments**

Demographics, clinical characteristics, inflammatory markers, and patient reported outcomes (physician and parent/patient global assessment) from the most recent clinic visit were abstracted from the hospital electronic health record. Disease activity was evaluated using the Juvenile Arthritis Disease Activity Score 3 (JADAS3) and categorized as inactive or active disease (with active disease encompassing mild, moderate, and high activity)(26). Cutoffs for disease activity categorization varied depending on if the patient had an oligoarthritic or polyarthritic course of disease, with higher cutoffs for each category in the latter(26). The formula for the JADAS3, active joint count (AJC) (max 10) + physician global (10cm VAS) + parent global (10cm VAS), is the same as the full JADAS calculation except it does not include erythrocyte sedimentation rate (ESR)(27). Scores can range from 0–30 with higher scores corresponding to more disease activity.

The NIH PROMIS initiative supports the PROMIS Assessment Center, a free data collection and management tool that provides an online resource for gathering study subject data securely(28). PROMIS users have a variety of options for administering and scoring their instruments. For this study, we administered the questionnaires directly through PROMIS

Assessment Center for real-time scoring and data storage. Full banks, short forms, and CATs are available through this option. For those interested in scoring just short forms, PROMIS Assessment Center has a Scoring Service where users may administer questionnaires outside of the PROMIS Assessment Center and then upload a file of the responses to still effectively utilize IRT scoring. Another option for scoring short forms is to use the score conversion tables available in the scoring manuals on the PROMIS Assessment Center webpage. This option is advantageous in that it does not require an internet-enabled device for scoring but is less informative because it does not take full advantage of IRT. This survey was administered electronically and all questionnaire content was identical to that of the paper and pencil mode of administration. Studies have found that there are no significant differences between electronic versus paper and pencil modes of administration in pediatric populations(29–31). Options for survey completion included using study team resources to access the online survey portal (study team laptop/tablet) or accessing the portal on their own with a provided reference ID and URL to the PROMIS Assessment Center. PROMIS Assessment Center was used to collect consent and survey responses electronically through unique patient login identification. In all, eight PROMIS item banks were administered to patients and parents: anger, anxiety, depressive symptoms, fatigue, mobility, pain interference, peer relationships, and upper extremity function. Each patient and/or parent completed one of 3 randomly assigned forms containing 1 to 4 PROMIS full item banks (Form 1=46 questions: anger, anxiety, depressive symptoms, and peer relationships; Form 2=13 questions: pain interference; Form 3=75 questions: fatigue, mobility, and upper extremity function). The number of questions for full banks, short forms, and CATs are listed for pediatric self-report and parent-report in Table 1.

The full item bank responses were used to generate the short form scores using the applicable questions for each PROMIS instrument. One domain, anger, was only available as a 5-item short form for the pediatric instrument. Full item bank responses were used to perform CAT simulations using the computer program Firestar(32). The CAT simulations conducted using Firestar were programmed using the default item selection parameters employed by the PROMIS Assessment Center and the following stopping criteria: Minimum: 4, Maximum: 12, and Standard Error: 0.3. These stopping conditions mirror those used in the PROMIS Assessment Center except we lowered the minimum number of items from 5 to 4 and required higher reliability by decreasing the standard error criteria from the default 0.4 to 0.3. The anger domain could not be simulated with CAT because anger was not available as a full bank pediatric instrument.

### Statistical analysis

Differences in clinical and demographic characteristics between JIA categories were compared using the Kruskal-Wallis or chi-squared test, as appropriate. Full bank and short form T-scores were calculated for each domain using the Bayesian Expected A Posteriori (EAP) estimation procedures in the PROMIS Assessment Center. The CAT simulations also used EAP estimation procedures to generate theta scores in R using Firestar (v 1.2.2)(32). Theta scores from the CAT simulations were transformed to make them comparable to the full bank and short form T-scores through a linear transformation by multiplying scores by 10 and adding 50.

The Mann-Whitney test was used to evaluate each measure's discrimination in children with inactive versus active disease as defined by the JADAS3. Standard errors were plotted across the full range of T-scores for each reporter (patient and parent) and assessment option (full bank, short form, and CAT) to assess measurement precision. Reliability estimates range from 0–1 and are calculated as  $1-SE^2$ , with greater values corresponding to higher reliability(33). SE of 3.2 translates to a reliability coefficient of approximately 0.9, which is the minimum acceptable reliability recommended for individual comparisons(33). The associations between patient and parent dyad responses, as well as across assessment options (full item banks versus short forms and CATs), were assessed using absolute-agreement intraclass correlations (ICC) for two-way random-effects models. Qualitative interpretation of ICCs is complicated by the variability in the different formulations of the ICC as well as the levels of variability between subjects within the data(34). A general guideline for interpretation of ICCs from the literature is that reproducibility of an ICC<0.40 is poor, 0.40 ICC<0.75 is fair to good, and ICC 0.75 is excellent(35).

## RESULTS

### Participant characteristics

228 patients diagnosed with JIA completed pediatric self-report forms and 223 parents of children diagnosed with JIA completed parent proxy forms for a total of 265 unique patients (Table 2). 185 patient-parent dyads participated, providing 59–63 paired responses per domain. Participants' demographic and disease activity characteristics are summarized in Table 2. There were no differences in the sex, race, ethnicity, JIA category, or disease activity scores between the patient self-reporters and the patients represented by parent proxy-reporters. There were also no significant differences between the patient characteristics across the three different forms (Table 2). Overall, the most prevalent diagnosis was oligoarticular JIA (32.1%) (Table 2). Approximately half the children (52.6%) had inactive disease (Table 2); 12.1% of patients had high disease activity according to JADAS3 standards. The three components of the JADAS3 were available for 93% of subjects, with non-calculable JADAS3 scores secondary to 0.4% missing physician global VAS and 6.8% missing patient global VAS (Table 2).

### Error in full item banks, short forms, and simulated CAT

The full bank, short form, and CAT instruments provided comparable measurement precision across the full range of each outcome. Both patient- and parent-report had levels of standard error that met or exceeded the minimum acceptable reliability coefficient of 0.9 at the population mean of 50 and at least one standard deviation in the direction of clinical interest (e.g. poorer peer relationships or elevated anxiety) for all assessment options in the depressive symptoms and pain interference domains (Figure 1). Anxiety, fatigue, and peer relationships only met these criteria for parent-report (Figure 1, 2). Instrument error levels for the full item banks, short forms, and CATs did not reach the minimum standard recommended for individual assessments at the population mean in the anger, mobility, or upper extremity function domains for either patient- or parent-report (Figure 2).

For all domains, except fatigue parent-report, CAT required more questions, on average, to reach the assigned stopping criteria than the fixed number of items in the corresponding short form (Table 1). The mean and range of items for each CAT are listed in Table 3.

### **Discriminative ability of instruments across disease activity levels**

Ability to discriminate scores between patients with inactive versus active disease did not differ by assessment option (full item bank, short form, and CAT). Figure 3 is a graphical display of the short form and CAT T-scores by disease activity category for each domain and respondent type. Parent-report scores differed significantly ( $p<0.05$ ) between disease activity states in all domains except upper extremity function (Figure 3). Patient-report scores were not as effective in discriminating disease states, with only the anger, fatigue, mobility, and pain interference domains differing significantly ( $p<0.05$ ; Figure 3).

### **Correlation of patient and parent dyad responses**

Correlation between patient- and parent-report responses was assessed for the full item banks, short forms, and CATs (Table 1). The highest pairwise correlations between patient and parent dyads for full item bank scores were seen in pain interference,  $ICC=0.80$ , and fatigue,  $ICC=0.71$  (Table 1). The lowest correlation coefficients in the full item banks were observed in peer relationships,  $ICC=0.42$  and anxiety,  $ICC=0.52$  (Table 1).

Near perfect correlations were observed between the full item banks and short form/CAT instruments for patient and parent T-scores in all domains. Full item bank T-scores were highly correlated ( $ICC = 0.96$ ) to short form and CAT T-scores in all domains for both patient self-report and parent proxy-report. Standard error agreement between assessment options was also high, though not as high ( $ICC = 0.83$ ) for both patient self-report and parent proxy-report.

## **DISCUSSION**

This is the first clinical validation study in children with JIA using multiple PROMIS domains. None of these domains have been previously validated in children with JIA and we hypothesized that each domain would have “poorer” outcomes with more disease activity. As evident in Figure 3, the domains generally behaved as expected, with patients and parents of patients having active disease reporting worse outcomes. The PROMIS pediatric short forms and CATs for anxiety, depressive symptoms, fatigue, mobility, pain interference, and peer relationships were discriminative between active disease and inactive disease for patient-, parent-report, or both. The short form in the anger domain was also discriminative between disease activity levels for both patient and parent reporters. Upper extremity function was the only domain unable to significantly differentiate between inactive and active disease for either respondent type in any instrument. While this is an important health domain in JIA, it is unique in that, unlike the other domains, there are situations where patients could have “active” disease that does not affect their upper extremity function, such as knee arthritis.

We demonstrated that all short forms and CATs had comparable error to the full item banks. The majority of PROMIS domains had acceptable standard error at the population mean and



at one standard deviation or greater in the direction of clinical interest for either one or both survey respondent types; the exceptions to this statement include the anger, mobility, and upper extremity function domains. Mobility and upper extremity function are particularly relevant domains for JIA patients and warrant further scrutiny to determine if they are capable of performing adequately in this population. These domains that did not show reliability at the population mean (anger, mobility, and upper extremity function) have shown similar reliability patterns in PROMIS reference populations(36). This amount of error also seems expected when one examines the short form score conversion tables (an alternative scoring method briefly mentioned in the materials and methods section) available in the scoring manuals on the Assessment Center webpage. The three domains with low reliability at the population mean have SEs ranging from 4–5.4 for the T-scores closest to 50. In comparison, the other domains range from SEs of 2–3.8 using the same criteria. It is important to note that although the mobility and upper extremity function domains did not meet the minimum reliability level at the population mean, they exhibited acceptable reliability in the direction of clinical importance (i.e., less mobility, reduced upper extremity function). Nonetheless, future work may be needed to augment the reliability of these domains in the JIA population.

Surprisingly, the number of items required for CAT was on average greater than the number of items in the short forms. We can attribute this to the more stringent stopping criteria employed in our study than the default criteria used in the PROMIS Assessment Center. Additionally, CAT required more items to converge to the assigned stopping criteria in patient reporters versus parent reporters.

Patient-parent dyad T-score correlations ranged from fair to excellent for all instruments in the measured domains, with only the peer relationships short form and CAT ICCs dropping down to poor agreement. Although we observed wide-ranging levels of agreement between children and their parents across content areas, it remains important to collect both perspectives in many situations. Patients with JIA may be too young or impaired to complete the questionnaires and, further, it is often the parental perception of health that drives healthcare utilization(37, 38). While pediatric self-reports provide the most insight into a patient's wellbeing, comprehensive health outcomes assessments will also consider parent or guardian proxy-reports when necessary as useful tools in patient evaluation.

As new drug therapies continue to emerge and the proportion of patients achieving inactive and low disease activity increase, the importance of PROs may become paramount in therapeutic decision-making. Domains that may be of considerable interest in this population of children include many of the domains covered by PROMIS including anxiety and peer relationships. Reliability across all instruments exhibited patterns seen in other studies(36) — lower measurement precision in the direction of better health or functioning. This is of little consequence when measuring these outcomes in a chronic disease population because we are generally more interested in measuring disability.

Our results should be considered in light of several limitations. First, the sample of JIA patients and parent reporters were drawn from a single center and may not be fully reflective of patients seen in other geographic areas. Our institution, however, is a relatively large

tertiary care facility and the sample included all categories of JIA with varying levels of disease activity. Second, the sample of patients was a convenience sample of children with predominantly well-controlled disease. The proportion of children with well-controlled disease in this study, however, is likely to be reflective of the composition at other tertiary care centers. Ideally, the PROMIS short forms and CATs will perform equally well, if not better, in children with more disease activity. Third, although our sample size was adequate to assess standard error and discrimination in JIA as a whole, our sample size was not sufficient to allow for validation or comparison of error or discrimination in each of the seven JIA categories. Future studies will be needed to assess differences in these PROs across these categories. Fourth, this was a cross-sectional study and, consequently, responsiveness of the instruments to changes in disease activity over time could not be assessed.

In summary, precision and discriminatory abilities of PROMIS instruments depend on health domain and report type (self-report versus parent proxy-report) for children with JIA. Future studies may be warranted to optimize the reliability of the upper extremity and mobility domains in the JIA population given their very high relevance to this condition. The decision to use a short form or CAT should rest with the provider and will likely depend upon whether Internet access is available at the time of PRO assessment. Increased attention to PROs is likely not only to enrich the assessment of disease activity, therapeutic tolerance and acceptability, but also to increase patient and parent satisfaction with their rheumatologic care.

## Acknowledgments

**Funding:** Dr. Weiss' work was supported by the NIH/National Institute of Arthritis and Musculoskeletal and Skin Diseases (grant 1-K23-AR-059749-01A1).

Special thanks to Jenna Tress and the CHOP Rheumatology Research Core team of Kelley Collier, Valerie Levy, and Janille Diaz for their recruitment efforts. We also thank the families and patients who completed PROMIS forms to make this study possible.

PROMIS® was funded with cooperative agreements from the National Institutes of Health (NIH) Common Fund Initiative (Northwestern University, PI: David Cella, PhD, U54AR057951, U01AR052177; 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](http://www.nihpromis.org) for additional information on the PROMIS® initiative.

## References

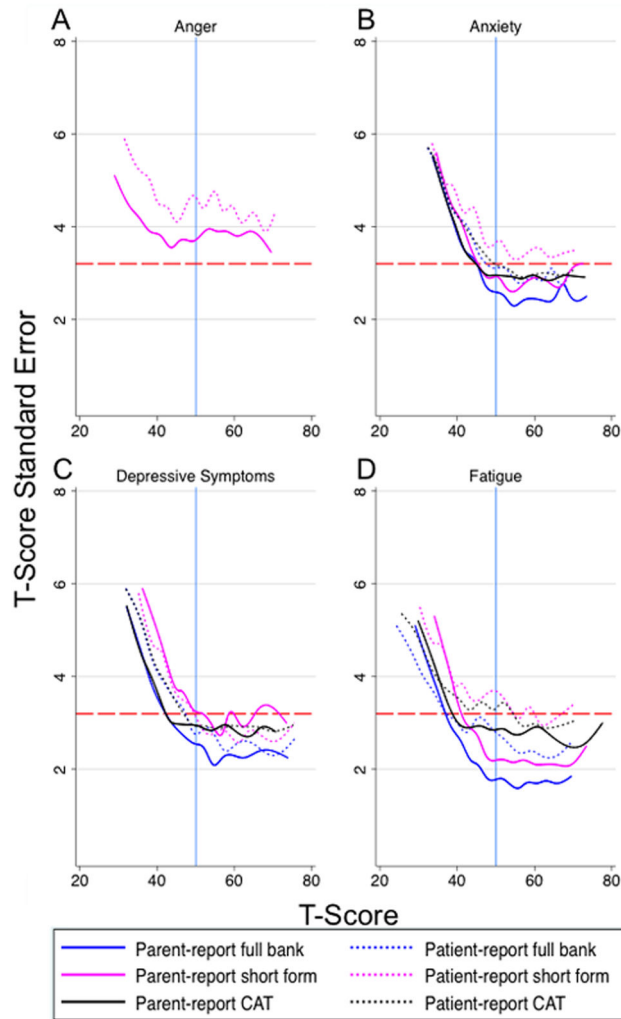
1. Garcia-Munitis P, Bandeira M, Pistorio A, Magni-Manzoni S, Ruperto N, Schivo A, et al. Level of agreement between children, parents, and physicians in rating pain intensity in juvenile idiopathic arthritis. *Arthritis Rheum.* 2006; 55(2):177–83. [PubMed: 16583392]
2. Sztajn bok F, Coronel-Martinez DL, Diaz-Maldonado A, Novarini C, Pistorio A, Viola S, et al. Discordance between physician's and parent's global assessments in juvenile idiopathic arthritis. *Rheumatology (Oxford).* 2007; 46(1):141–5. [PubMed: 16782733]
3. Consolaro A, Vitale R, Pistorio A, Lattanzi B, Ruperto N, Malattia C, et al. Physicians' and parents' ratings of inactive disease are frequently discordant in juvenile idiopathic arthritis. *J Rheumatol.* 2007; 34(8):1773–6. [PubMed: 17611978]
4. Taxter AJ, Wileyto EP, Behrens EM, Weiss PF. Patient-reported Outcomes across Categories of Juvenile Idiopathic Arthritis. *J Rheumatol.* 2015; 42(10):1914–21. [PubMed: 26329337]
5. Weiss PF, Beukelman T, Schanberg LE, Kimura Y, Colbert RA. Enthesitis-related arthritis is associated with higher pain intensity and poorer health status in comparison with other categories of juvenile idiopathic arthritis: the Childhood Arthritis and Rheumatology Research Alliance Registry. *J Rheumatol.* 2012; 39(12):2341–51. [PubMed: 23070991]
6. U.S. Department of Health and Human Services, Food and Drug Administration. Guidance for industry: Patient-reported outcome measures: Use in medical product development to support labeling claims. 2009.
7. Clancy C, Collins FS. Patient-Centered Outcomes Research Institute: the intersection of science and health care. *Sci Transl Med.* 2010; 2(37):37cm18.
8. Ader DN. Developing the Patient-Reported Outcomes Measurement Information System (PROMIS). *Med Care.* 2007; 45(5, Suppl 1):S1–S2. [PubMed: 18027399]
9. Howe S, Levinson J, Shear E, Hartner S, McGirr G, Schulte M, et al. Development of a disability measurement tool for juvenile rheumatoid arthritis. The Juvenile Arthritis Functional Assessment Report for Children and their Parents. *Arthritis Rheum.* 1991; 34(7):873–80. [PubMed: 2059234]
10. Singh G, Athreya BH, Fries JF, Goldsmith DP. Measurement of health status in children with juvenile rheumatoid arthritis. *Arthritis Rheum.* 1994; 37(12):1761–9. [PubMed: 7986222]
11. Wright FV, Kimber JL, Law M, Goldsmith CH, Crombie V, Dent P. The Juvenile Arthritis Functional Status Index (JASI): a validation study. *J Rheumatol.* 1996; 23(6):1066–79. [PubMed: 8782142]
12. Iglesias MJ, Cuttica RJ, Herrera Calvo M, Micelotta M, Pringe A, Brusco MI. Design and validation of a new scale to assess the functional ability in children with juvenile idiopathic arthritis (JIA). *Clin Exp Rheumatol.* 2006; 24(6):713–8. [PubMed: 17207392]
13. Filocamo G, Sztajn bok F, Cespedes-Cruz A, Magni-Manzoni S, Pistorio A, Viola S, et al. Development and validation of a new short and simple measure of physical function for juvenile idiopathic arthritis. *Arthritis Rheum.* 2007; 57(6):913–20. [PubMed: 17665481]
14. Duffy CM, Arsenault L, Duffy KN, Paquin JD, Strawczynski H. The Juvenile Arthritis Quality of Life Questionnaire--development of a new responsive index for juvenile rheumatoid arthritis and juvenile spondyloarthritis. *J Rheumatol.* 1997; 24(4):738–46. [PubMed: 9101511]
15. Varni JW, Seid M, Smith Knight T, Burwinkle T, Brown J, Szer IS. The PedsQL in pediatric rheumatology: reliability, validity, and responsiveness of the Pediatric Quality of Life Inventory Generic Core Scales and Rheumatology Module. *Arthritis Rheum.* 2002; 46(3):714–25. [PubMed: 11920407]
16. Filocamo G, Schiappapietra B, Bertamino M, Pistorio A, Ruperto N, Magni-Manzoni S, et al. A new short and simple health-related quality of life measurement for paediatric rheumatic diseases: initial validation in juvenile idiopathic arthritis. *Rheumatology (Oxford).* 2010; 49(7):1272–80. [PubMed: 20338888]

17. Filocamo G, Consolaro A, Schiappapietra B, Dalpra S, Lattanzi B, Magni-Manzoni S, et al. A new approach to clinical care of juvenile idiopathic arthritis: the Juvenile Arthritis Multidimensional Assessment Report. *J Rheumatol*. 2011; 38(5):938–53. [PubMed: 21362761]
18. Cella D, Yount S, Rothrock N, Gershon R, Cook K, Reeve B, 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(5 Suppl 1)
19. Irwin DE, Stucky BD, Langer MM, Thissen D, DeWitt EM, Lai JS, et al. PROMIS Pediatric Anger Scale: an item response theory analysis. *Qual Life Res*. 2012; 21(4):697–706. [PubMed: 21785833]
20. Irwin DE, Stucky B, Langer MM, Thissen D, Dewitt EM, Lai JS, et al. An item response analysis of the pediatric PROMIS anxiety and depressive symptoms scales. *Qual Life Res*. 2010; 19(4): 595–607. [PubMed: 20213516]
21. Lai JS, Stucky BD, Thissen D, Varni JW, DeWitt EM, Irwin DE, et al. Development and psychometric properties of the PROMIS((R)) pediatric fatigue item banks. *Qual Life Res*. 2013; 22(9):2417–27. [PubMed: 23378106]
22. DeWitt EM, Stucky BD, Thissen D, Irwin DE, Langer M, Varni JW, et al. Construction of the eight-item patient-reported outcomes measurement information system pediatric physical function scales: built using item response theory. *J Clin Epidemiol*. 2011; 64(7):794–804. [PubMed: 21292444]
23. Varni JW, Stucky BD, Thissen D, Dewitt EM, Irwin DE, Lai JS, et al. PROMIS Pediatric Pain Interference Scale: an item response theory analysis of the pediatric pain item bank. *J Pain*. 2010; 11(11):1109–19. [PubMed: 20627819]
24. Dewalt DA, Thissen D, Stucky BD, Langer MM, Morgan Dewitt E, Irwin DE, et al. PROMIS Pediatric Peer Relationships Scale: development of a peer relationships item bank as part of social health measurement. *Health Psychol*. 2013; 32(10):1093–103. [PubMed: 23772887]
25. Petty RE, Southwood TR, Manners P, Baum J, Glass DN, Goldenberg J, et al. International League of Associations for Rheumatology classification of juvenile idiopathic arthritis: second revision, Edmonton, 2001. *J Rheumatol*. 2004; 31(2):390–2. [PubMed: 14760812]
26. Consolaro A, Negro G, Gallo MC, Bracciolini G, Ferrari C, Schiappapietra B, et al. Defining Criteria for Disease Activity States in Nonsystemic Juvenile Idiopathic Arthritis Based on a Three-Variable Juvenile Arthritis Disease Activity Score. *Arthritis Care Res (Hoboken)*. 2014; 66(11): 1703–9. [PubMed: 24980508]
27. McErlane F, Beresford MW, Baildam EM, Chieng SE, Davidson JE, Foster HE, et al. Validity of a three-variable Juvenile Arthritis Disease Activity Score in children with new-onset juvenile idiopathic arthritis. *Ann Rheum Dis*. 2012
28. Gershon RC, Rothrock N, Hanrahan R, Bass M, Cella D. The use of PROMIS and assessment center to deliver patient-reported outcome measures in clinical research. *J Appl Meas*. 2010; 11(3): 304–14. [PubMed: 20847477]
29. Raat H, Mangunkusumo RT, Landgraf JM, Kloek G, Brug J. Feasibility, reliability, and validity of adolescent health status measurement by the Child Health Questionnaire Child Form (CHQ-CF): internet administration compared with the standard paper version. *Qual Life Res*. 2007; 16(4):675–85. [PubMed: 17286197]
30. Salaffi F, Gasparini S, Ciapetti A, Gutierrez M, Grassi W. Usability of an innovative and interactive electronic system for collection of patient-reported data in axial spondyloarthritis: comparison with the traditional paper-administered format. *Rheumatology (Oxford)*. 2013; 52(11):2062–70. [PubMed: 23955646]
31. Young NL, Varni JW, Snider L, McCormick A, Sawatzky B, Scott M, et al. The Internet is valid and reliable for child-report: An example using the Activities Scale for Kids (ASK) and the Pediatric Quality of Life Inventory (PedsQL). *J Clin Epidemiol*. 2009; 62(3):314–20. [PubMed: 18834710]
32. FIRESTAR: Computerized Adaptive Testing (CAT) Simulation Program. 1.2.2 ed. Department of Medical Social Sciences Northwestern University's Feinberg School of Medicine: Seung W. Choi, PhD; 2009.

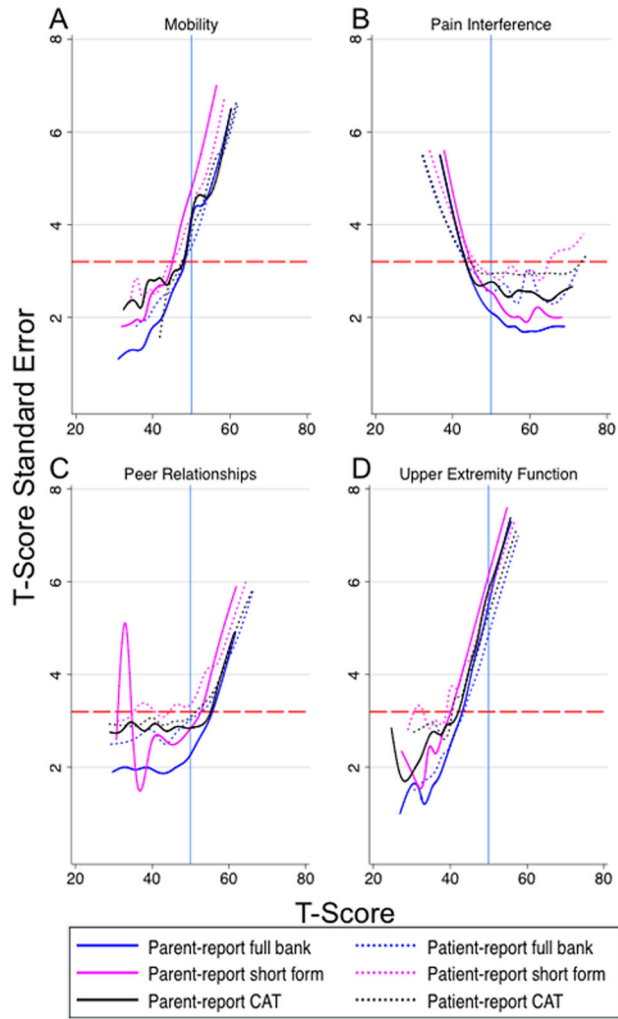
33. NIH. Instrument Development and Validation - Scientific Standards: Version 2.0. In: Hays, R.Moinpour, C., et al., editors. Appendix 13 Reliability. National Institutes of Health; 2013. p. 3
34. Weir JP. Quantifying Test-Retest Reliability Using the Intraclass Correlation Coefficient and the SEM. *J Strength Cond Res.* 2005; 19(1):231. [PubMed: 15705040]
35. Rosner, B. *Fundamentals of Biostatistics.* 6. Belmont, CA: Thomson-Brooks/Cole; 2006.
36. Varni JW, Magnus B, Stucky BD, Liu Y, Quinn H, Thissen D, et al. Psychometric properties of the PROMIS® pediatric scales: precision, stability, and comparison of different scoring and administration options. *Qual Life Res.* 2013; 23(4):1233–43. [PubMed: 24085345]
37. Irwin DE, Gross HE, Stucky BD, Thissen D, DeWitt E, Lai J, et al. Development of six PROMIS pediatrics proxy-report item banks. *Health Qual Life Outcomes.* 2012; 10(1):22. [PubMed: 22357192]
38. Janicke DM, Finney JW, Riley AW. Children’s health care use: a prospective investigation of factors related to care-seeking. *Med Care.* 2001; 39(9):990–1001. [PubMed: 11502956]

**SIGNIFICANCE AND INNOVATION**

- This is the first study to evaluate PROMIS tools for measurement of patient-reported outcomes in children with JIA.
- The PROMIS pediatric short forms and CATs for anger, anxiety, depressive symptoms, fatigue, mobility, pain interference, and peer relationships discriminated between inactive and active disease, as defined by JADAS3, for parent-report or both patient- and parent-report.
- Administration of CAT did not reduce the number of items compared to the static short form for all but one domain.
- Varying levels of child-parent concordance reinforces the importance of considering both perspectives in comprehensive health outcomes assessments.

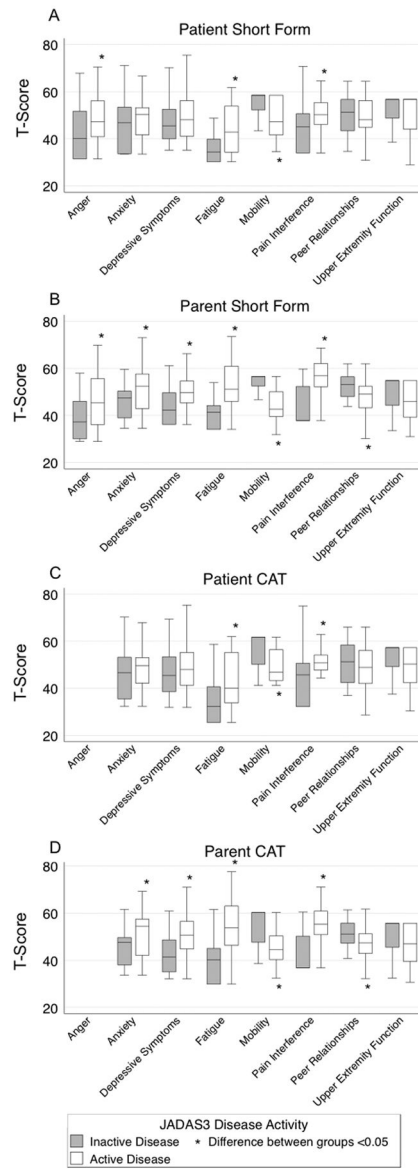


**Figure 1.** Standard error in the anger, anxiety, depressive symptoms, and fatigue domains across the full range of T-scores for each assessment option administered. Change in standard error (SE) in the (A) ‘Anger’, (B) ‘Anxiety’, (C) ‘Depressive Symptoms’, and (D) ‘Fatigue’ domains. Full item bank, short form, and computerized adaptive test (CAT) instruments shown. A T-score of ‘50’ (solid, blue, vertical line) represents the population mean score with standard deviation equal to  $\pm 10$ . The dashed, red, horizontal line corresponds with a reliability score of 0.9 (SE=3.2). T-score reliability increases as SE approaches zero. Scores below the reference line at SE=3.2 (reliability 0.9) have acceptable reliability for individual assessment according to the Patient Reported Outcomes Measurement Information System (PROMIS) scientific standards.



**Figure 2.** Standard error in the mobility, pain interference, peer relationships, and upper extremity function domains across the full range of T-scores for each assessment option administered. Change in standard error (SE) in the (A) ‘Mobility’, (B) ‘Pain Interference’, (C) ‘Peer Relationships’, and (D) ‘Upper Extremity Function’ domains. Full item bank, short form, and computerized adaptive test (CAT) instruments shown. A T-score of ‘50’ (solid, blue, vertical line) represents the population mean score with standard deviation equal to  $\pm 10$ . The dashed, red, horizontal line corresponds with a reliability score of 0.9 (SE=3.2). T-score reliability increases as SE approaches zero. Scores below the reference line at SE=3.2 (reliability = 0.9) have acceptable reliability for individual assessment according to the Patient Reported Outcomes Measurement Information System (PROMIS) scientific standards.





**Figure 3.**

Discrimination of short forms and computerized adaptive tests (CATs) between JADAS3 disease activity levels.

(A) Patient-report short form T-scores, (B) Parent-report short form T-scores, (C) Patient-report CAT T-scores, and (D) Parent-report CAT T-scores. A higher T-score indicates a worse outcome in the following domains: Anger, Anxiety, Depressive Symptoms, Fatigue, and Pain Interference. Lower T-scores indicate a worse outcome in the remaining domains: Mobility, Peer Relationships, and Upper Extremity Function.

**Table 1**

Number of items available for full item banks and short forms in each measured domain, the mean number of items required to complete the corresponding computerized adaptive tests (CAT), and the intraclass correlation coefficients (ICCs) for patient-parent dyad T-scores.

Domain	Patient-report			Parent-report			ICC coefficients for patient-parent dyad T-scores		
	Full bank	Short form	CAT	Full bank	Short form	CAT	Full bank N (ICC)	Short form N (ICC)	CAT N (ICC)
Anger	--	5	--	--	5	--	63 (0.54)	--	--
Anxiety	13	8	11.8	13	8	9.1	63 (0.52)	63 (0.51)	63 (0.52)
Depressive symptoms	13	8	10.1	13	6	8.7	61 (0.57)	61 (0.55)	61 (0.55)
Fatigue	23	10	11.9	23	10	8.2	62 (0.71)	62 (0.7)	62 (0.64)
Mobility	23	8	10.0	23	8	9.1	63 (0.66)	63 (0.57)	63 (0.65)
Pain interference	13	8	10.0	13	8	8.2	59 (0.8)	59 (0.77)	59 (0.79)
Peer relationships	15	8	10.9	15	7	7.7	62 (0.42)	62 (0.33)	62 (0.35)
Upper extremity function	29	8	11.9	29	8	8.2	62 (0.59)	62 (0.47)	62 (0.56)

Values listed for CAT are the mean number of items required for the CAT simulation to reach the assigned stopping criteria for patient- and parent-report.

**Table 2**

Patient clinical and demographic characteristics at time of survey

	All* N=265	Form 1 <sup>¶</sup> N=91	Form 2 <sup>¶</sup> N=85	Form 3 N=90
Age in years, M (IQR)	12 (9, 15)	11 (9, 14)	13 (10, 16)	12 (9, 15)
Sex: Female, N (%)	185 (69.8%)	66 (72.5%)	57 (67.1%)	63 (70.0%)
Race, N (%)				
White	225 (84.9%)	77 (84.6%)	73 (85.9%)	76 (84.4%)
Black/African American	18 (6.8%)	7 (7.7%)	4 (4.7%)	7 (7.8%)
Asian	3 (1.1%)	2 (2.2%)	1 (1.2%)	0 (0.0%)
Other	16 (6.0%)	4 (4.4%)	6 (7.1%)	6 (6.7%)
More than one race	3 (1.1%)	1 (1.1%)	1 (1.2%)	1 (1.1%)
JIA subtype, N (%)				
Enthesitis-related arthritis	58 (21.9%)	16 (17.6%)	24 (28.2%)	18 (20.0%)
Oligoarticular arthritis	85 (32.1%)	27 (29.7%)	27 (31.8%)	31 (34.4%)
Polyarticular rheumatoid factor –	57 (21.5%)	22 (24.2%)	17 (20.0%)	19 (21.1%)
Polyarticular rheumatoid factor +	4 (1.5%)	0 (0.0%)	1 (1.2%)	3 (3.3%)
Psoriatic arthritis	24 (9.1%)	9 (9.9%)	7 (8.2%)	8 (8.9%)
Systemic	12 (4.5%)	5 (5.5%)	2 (2.4%)	5 (5.6%)
Undifferentiated	25 (9.4%)	12 (13.2%)	7 (8.2%)	6 (6.7%)
JADAS3 score, M (IQR)	1 (0, 4)	1 (0, 4)	1 (0, 4)	1 (0, 4)
JADAS3 disease state <sup>§</sup> , N (%)				
Inactive Disease	130 (52.6%)	43 (51.2%)	43 (54.4%)	45 (52.9%)
Active Disease	117 (47.4%)	41 (48.8%)	36 (45.6%)	40 (47.1%)
Survey Responses <sup>¶</sup> , N (%)	N=452	154 (34.1%)	145 (32.1%)	153 (33.8%)
Patient-reported	229 (50.7%)	77 (50.0%)	74 (51.0%)	78 (51.0%)
Parent-reported	223 (49.3%)	77 (50.0%)	71 (49.0%)	75 (49.0%)
Patient-parent dyads	185	63	59	63

Abbreviations: JADAS 3: Juvenile Arthritis Disease Activity Score 3; JIA: Juvenile Idiopathic Arthritis; M (IQR): Median (Interquartile range).

\* All is a reference to all unique patients in the study (excludes repeated patient information from dyads and also the repeated information from the patient who completed 2 forms);

<sup>¶</sup>One patient completed two forms (forms 1 and 2).

<sup>§</sup>JADAS3 disease activity states were consolidated from 4 categories (inactive, low, moderate, and high disease activity) to 2 categories by combining low, moderate and high disease activity into a new category called “active disease”.

Form 1=Anger, Anxiety, Depressive Symptoms, and Peer Relationships; Form 2=Pain Interference; Form 3=Fatigue, Mobility, Upper Extremity Function. No significant differences ( $p<0.05$ ) in participant characteristics existed between Form 1, Form 2, and Form 3 as tested by Kruskal-Wallis or chi-squared test, as appropriate.

**Table 3**

Patient- and parent-reported PROMIS T-scores by assessment option

	N	Full item bank Mean (SD)	Short form Mean (SD)	CAT Mean (SD)
<b>Patient-report</b>				
Anger	77	--	46.1 (11.1)	--
Anxiety	77	47.3 (10.9)	47.6 (10.6)	47.4 (10.8)
Depressive symptoms	75	47.0 (10.4)	47.8 (9.6)	47.2 (10.4)
Fatigue	78	38.0 (12.3)	40.2 (10.5)	38.9 (11.8)
Mobility	78	52.2 (8.9)	51.8 (8.1)	52.9 (7.8)
Pain interference	74	46.8 (10.3)	47.6 (9.9)	47.0 (10.5)
Peer relationships	76	49.0 (9.5)	49.4 (9.4)	49.2 (9.5)
Upper extremity function	77	51.3 (8.3)	51.5 (7.8)	51.2 (7.8)
<b>Parent-report</b>				
Anger	77	--	43.5 (10.9)	--
Anxiety	77	48.7 (9.8)	49.0 (9.7)	49.1 (10.1)
Depressive symptoms	77	47.0 (10.0)	47.3 (9.3)	47.1 (9.9)
Fatigue	74	43.9 (11.5)	45.6 (10.3)	45.4 (12.2)
Mobility	75	49.8 (8.2)	49.3 (7.7)	49.7 (8.4)
Pain interference	71	48.5 (11.3)	48.9 (10.6)	48.0 (10.8)
Peer relationships	77	48.9 (8.0)	49.6 (8.2)	48.9 (7.9)
Upper extremity function	75	48.0 (8.6)	47.9 (8.2)	48.1 (8.5)

Abbreviations: PROMIS: Patient Reported Outcomes Measurement Information System; SD: Standard Deviation; CAT: Computerized Adaptive Test; A higher T-score indicates a worse outcome in the following domains: Anger, Anxiety, Depressive Symptoms, Fatigue, and Pain Interference. Lower T-scores indicate a worse outcome in the remaining domains: Mobility, Peer Relationships, and Upper Extremity Function.