



Published in final edited form as:

Qual Life Res. 2018 March ; 27(3): 755–764. doi:10.1007/s11136-017-1709-z.

Establishing Clinical Severity for PROMIS® Measures in Adult Patients with Rheumatic Diseases

Vivek Nagaraja, MD^{1,2}, Constance Mara, PhD³, Puja P. Khanna, MD, MPH^{1,4}, Rajaie Namas, MD¹, Amber Young, MD^{1,2}, David A Fox, MD¹, Timothy Laing, MD¹, William J. McCune, MD¹, Carol Dodge, OTRL, CHT¹, Debra Rizzo¹, Maha Almackenzie, MD⁵, and Dinesh Khanna, MD, MS^{1,2}

¹Division of Rheumatology, University of Michigan, Ann Arbor, MI

²University of Michigan Scleroderma Program, University of Michigan, Ann Arbor, MI

³Behavioral Medicine and Clinical Psychology, Cincinnati Children's Hospital Medical Center, Cincinnati, OH

⁴Ann Arbor Veterans Affairs Medical Center Ann Arbor, MI

⁵Department of Medicine, University of Toledo, Toledo, OH

Abstract

Purpose—Different patient reported outcome (PRO) measures are used for rheumatic diseases (RD). The aims of this study are – (1) Identify PROMIS® domains most relevant to care of patients with RD, (2) Collect T-Score metrics in patients with RD, and (3) Identify clinically meaningful cut-points for these domains.

Corresponding Author: Dinesh Khanna, MD MS, Division of Rheumatology, University of Michigan, 300 North Ingalls Building, Room 7C27, Ann Arbor, MI 48109-5422, Phone 734-936-5561 / Fax 734-763-1253, khannad@med.umich.edu.

Other authors:

Vivek Nagaraja MD, Division of Rheumatology, University of Michigan, 300 North Ingalls Building, Room 7C27, Ann Arbor, MI 48109-5422, Phone 734-936-5561 / Fax 734-763-1253, nagarajv@med.umich.edu

Constance Mara PhD, Behavioral Medicine and Clinical Psychology, Cincinnati Children's Hospital Medical Center, 3333 Burnet Ave. MLC 7039, Cincinnati, OH 45229, Constance.Mara@cchmc.org

Puja P. Khanna MD, MPH, Division of Rheumatology, University of Michigan, 300 North Ingalls Building, Room 7C27, Ann Arbor, MI 48109-5422, pkhanna@med.umich.edu

Rajaie Namas MD, Division of Rheumatology, University of Michigan, 300 North Ingalls Building, Room 7C27, Ann Arbor, MI 48109-5422, rnamas@med.umich.edu

Amber Young MD, Division of Rheumatology, University of Michigan, 300 North Ingalls Building, Room 7C27, Ann Arbor, MI 48109-5422, yamber@med.umich.edu

David A Fox MD, Division of Rheumatology, University of Michigan, 300 North Ingalls Building, Room 7C27, Ann Arbor, MI 48109-5422, dfox@med.umich.edu

Timothy Laing MD, Division of Rheumatology, University of Michigan, 300 North Ingalls Building, Room 7C27, Ann Arbor, MI 48109-5422, tlaing@med.umich.edu

William J. McCune MD, Division of Rheumatology, University of Michigan, 300 North Ingalls Building, Room 7C27, Ann Arbor, MI 48109-5422, jmccune@med.umich.edu

Carol Dodge OTRL, CHT, Division of Rheumatology, University of Michigan, 300 North Ingalls Building, Room 7C27, Ann Arbor, MI 48109-5422, cdjr@med.umich.edu

Debra Rizzo, Division of Rheumatology, University of Michigan, 300 North Ingalls Building, Room 7C27, Ann Arbor, MI 48109-5422, rizzod@med.umich.edu

Maha Almackenzie MD, Division of Rheumatology, University of Toledo, 3000 Arlington Ave, mailstop 1186, Toledo, OH 43614, Maha.Almackenzie@utoledo.edu

Compliance with Ethical Standards

Disclosure of potential conflicts of interest

The authors declare that they have no conflict of interest.

Methods—A convenience sample of RD patients was recruited consecutively during clinic visits, and asked to complete computer-adaptive tests on thirteen Patient-Reported Outcomes Measurement Information System (PROMIS®) instruments. Based on discussion with clinical providers, four measures were chosen to be relevant and actionable (from rheumatologists' perspective) in RD patients. Data from RD patients was used to develop clinical vignettes across a range of symptom severity. Vignettes were created based on most likely item responses at different levels on the T-score metric (mean = 50; SD = 10) and anchored at 5-point intervals (0.5 SDs). Patients with RD (N=9) and clinical providers (N=10) participated as expert panelists in separate one-day meetings using a modified educational standard setting method.

Results—Four domains (physical function, pain interferences, sleep disturbance, depression) that are actionable at the point-of-care were selected. For all domains, patients endorsed cut points at lower levels of impairment than providers by 0.5 to 1 SD (e.g., severe impairment in physical function was defined as a T-score of 35 by patients and 25 by providers).

Conclusions—We used a modified educational method to estimate clinically relevant cut points to classify severity for PROMIS measures. This allows for meaningful interpretation of PROMIS® measures in a clinical setting of RD population.

Keywords

Health status; patient-reported outcomes; rheumatic diseases; PROMIS®; clinically meaningful cut points

Purpose

Chronic medical conditions, such as rheumatic diseases (RD), have a detrimental effect on self-reported physical, mental, and social health, i.e., health related quality of life (HRQOL) [1; 2]. RDs are diverse with variable impact on HRQOL may be variable, can fluctuate over time, and may mirror disease flares [3–5]. A patient reported outcome (PRO) is any report of the status of a patient's health condition that comes directly from the patient, without interpretation of the patient's response by a clinician or anyone else [6]. PROs can supplement clinical decision making by aiding assessment and management of these conditions. There is increased enthusiasm within the rheumatology research community to integrate PRO measures with clinical assessments [7; 8].

The National Institutes of Health Patient-Reported Outcomes Measurement Information System (PROMIS®) Roadmap initiative (available at www.nihpromis.org) is a cooperative research program designed to develop, evaluate, and standardize item banks to measure PROs across patients with varying medical conditions and in a cross-section of the US population [9]. The aim of PROMIS® is to use item response theory (IRT) to develop reliable and valid item banks that can be administered as short forms and computerized adaptive tests (CAT) [10; 11]. As it improves measurement precision, it lessens burden on the patient. PROMIS short forms are available for incorporation in the electronic medical records [12].

As PROMIS measures move from psychometric development and validation stage to bedside, there is an urgent need to assess on what to measure and how to interpret these in

clinical care. Therefore, the aims of this study are – (1) Identify PROMIS health domains relevant to the care of patients with RD, (2) Collect T-Score information for these domains across patients with varied RD, and (3) Identify clinically meaningful cut-points for these domains.

Methods

This study was approved by the Institutional Review Board at University of Michigan (UM) prior to participant enrollment. The study was conducted in three phases.

Phase 1

We conducted four focus group discussions among practicing rheumatologists to obtain their input on various formats of a PROMIS report card, to optimize ease of interpretation and track utility in longitudinal clinical care. We held multiple sessions to incorporate all the clinicians who volunteered to participate in this phase. Each session started with an overview of PROMIS® followed by a patient score report of representative PROMIS CAT measures. The report contained a heat map indicating the T-scores for each measure in comparison to the general US population, and description of the individual scores as T-scores with standard error of mean. We asked the following questions to the panelists: (i) What do you think of this report? (ii) Do you understand what this report is communicating? (iii) Is there anything you would change about the format of the report?

Phase 2

We recruited a convenience sample of RD patients aged 18 and over during routine clinic visits. A study titled ‘PROMIS in rheumatology’ was created on the Assessment CenterSM, an online data collection tool that enables researchers to create study-specific websites for capturing participant data securely. Thirteen PROMIS CAT measures were chosen – anger, anxiety, depression, fatigue, pain behavior, pain interference, physical function, physical function with mobility aid, satisfaction in roles and activities, sleep disturbance, sleep-related impairment, ability to participate in social activities, and social isolation. In recent studies, PROMIS measures can identify the impact of different RDs across a range of domains of physical, mental, and social health, which prompted us to use these 13-PROMIS CAT measures [8]. A unique login and password were provided to the subjects at the time of registration, and the subjects were consented electronically. Eligible subjects had to be able to read and interpret English, possess computer skills (to access the study site and complete the questionnaire), and be willing to give consent for study participation.

Each participant completed the demographic information, self-reported rheumatic diagnoses, and the PROMIS measures. Subjects completed the PROMIS measures using either a hand-held device or a desktop in a clinic examination room, or accessed the study through their home computer. Each subject was assigned a unique identification number using a computer-based system. Only de-identified data were used in the analyses.

Phase 3

The aim in this phase was to create meaningful cut-points for chosen PROMIS measures in patients with RD. The clinical providers who participated in Phase 1 were asked to choose domains that should be incorporated for routine clinical care of RD patients. The providers were informed that we plan to include these domains in clinical practice in near future. The providers agreed on limiting the assessment to four domains, keeping in mind the workflow of a busy rheumatology practice. Hence, providers, based on consensus, agreed on the four (of the thirteen PROMIS domains from Phase 2) – physical function, pain interference, sleep disturbance, and depression. They felt these domains were actionable and could be incorporated in the first stage of dissemination in clinical practice. On the other hand, they felt that some domains such as fatigue are multifactorial and difficult to address in clinical practice, so was chosen not to be incorporated as a domain in this initial stage. The rheumatologists acknowledged that fatigue is common in RD and should be considered in future.

Creating clinical vignettes

We used real-person data collected from patients with RD as part of phase 2. The scores in PROMIS measures are computed to a T-score metric where the mean of 50 represents the average level of the domain for US General Population and 10 is the standard deviation. It is important to note that US General Population is a representative population and not 'healthy' population. For positively-worded concepts, a T-score of 60 is one standard deviation (SD) better than average; on the other hand, for negatively-worded concepts, a T-score of 60 is one SD worse than average. Thus, the metric provides a normative context for scores; for example, a T-score of 70 indicates a level of outcome that is two standard deviations above the mean in the reference sample. Numerical T-score values are generated for the different PROMIS measures. However, in RD patients these different T-scores for the chosen domains, have not been uniformly defined and classified in terms of representing clinical severity of RD.

We adapted the 'Bookmark standard-setting procedure' in creating clinical vignettes at different cut-points. This method is routinely used in educational and psychological testing. An essential feature of this method is the use of IRT to "map" items onto a proficiency distribution where cut scores (standards) are set [13]. We implemented this concept to 'map' items for the four PROMIS measures onto a severity distribution around the mean T-score. Recently, the bookmark method has been used in creating meaningful cut-scores in patients with multiple sclerosis (MS), juvenile inflammatory arthritis (JIA), and in oncology setting [14–16].

Using the real-patient data from phase 2, we created clinical vignettes based on symptom severity (Supplementary file) taking into consideration the minimum and maximum scores for each measure from the cohort. We identified target locations on the T-score metric that were five points (0.5 SD units) apart. The vignettes were set at 2.5 and 7.5 T-score levels, so that bookmarks were at 0.0 and 0.5 T-score units. For each target location, we identified predicted responses for every item in each bank. We chose a wide range of PROMIS items from each bank for the development of the vignettes to prevent panelists from comparing

vignettes' items side by side, and determining severity by comparing responses to the same items across vignettes. Each vignette had a clinical narrative and predicted responses for five items. These vignettes were created around the target locations on the T-score metric with a difference of 0.5 SD between them.

Pre-workshop assignment

We identified two different panelists – patients and clinical providers. We contacted a random sample of patients with different RDs, who were under the care of rheumatologists at University of Michigan. We provided detailed information about the study and patient panel participation in focus group discussions. Most of the clinical providers in the panel had previously participated in Phase 1, and other providers were sent electronic mail messages requesting their participation. Both panels were provided with the details of the study and informed consent was obtained. Prior to the in-person workshop, we mailed panelists a packet of materials. This packet included: (a) clinical vignettes color-coded for each PROMIS domain, (b) Scoring sheets for each set of vignettes. Written instructions were provided that asked panelists to – (a) work on one set of vignettes at a time, (b) rank the vignettes in the order of severity, and (c) mail the rank-order list to the investigators.

In-person workshop

We conducted 2-day expert panel meetings – patients on day 1 and providers on day 2. The panel meetings began with introductions, followed by a warm-up exercise to acquaint participants with the bookmarking method, and then review of the clinical vignettes. Patient panelists began by completing their demographic information and PROMIS global health short form. For each domain, panelists were provided with clinical vignettes (the same vignettes had been mailed to them), and rank order list of the vignettes was presented, based on sequential order of the T-scores for each of the vignettes. Next, working individually, panelists placed the vignettes in front of them from the one representing the least to the one representing the most severe difficulty in a given domain. In successive steps, they placed bookmarks at thresholds between no problems and mild problems, mild and moderate problems, and, moderate and severe problems. The thresholds for levels of severity were calculated. These were the mean scores of the locations of the two vignettes that bordered the bookmark location. Panelists were then encouraged to discuss and evaluate the consequences of their cut-scores, and were allowed to change their bookmark locations. The same cycle was repeated for each domain.

Results

Phase 1

Eleven practicing rheumatologists participated in the phase 1 of the study. In total, we held four focus group sessions. We summarized the data and presented our questions to the panelists (described in the previous section). All rheumatologists agreed on the following: (i) T-scores of the PROMIS domains should be presented as thermal graphs or “heat maps” for easy interpretation, (ii) for better comparability, average T-scores of two different reference population should be provided for each domain – US general population (PROMIS standard) and rheumatology patients seen at the University of Michigan, (iii) it would be useful to

include an easily recognizable pictorial depiction of the most bothersome domain and an appropriate actionable item, and (iv) present the “concerning zone” where a rheumatologist may consider an action (Figures 1A and 1B), similar to recent guidance from PCORI on integrating PROs in EMR [17].

Phase 2

In this phase of the study, 217 patients with RD were recruited as a convenience sample. The majority of patients were women (77%) and Caucasian (82%; Table 1). Table 1 also shows the mean T-score metrics for this convenience sample. Except for physical function (with and without mobility aid), satisfaction with social roles and activities, and ability to participate in social activities, the rest of the domains are negatively-worded. The mean T-scores for physical function with (42) and without mobility aid (40), fatigue (58), and pain interference (58), were 0.8–1 SD worse than the US general population.

Phase 3

Clinical vignettes—In our cohort, the PROMIS measures had different range of T-scores noted in previous studies – (i) for depression, T-score range 42.5 to 82.5 (nine vignettes), (ii) for pain interference, T-score range 47.5 to 82.5 (eight vignettes), (iii) for physical function, T-score range 12.5 to 72.5 (twelve vignettes), and (iv) for sleep disturbance, T-score range 32.5 and 82.5 (ten vignettes)[18]. Cut scores for different severity levels were assigned the value of the mean of the upper and lower vignette scores delimiting the cut-point (e.g., if the cut-point was chosen by panelists between the two vignettes corresponding to T-Scores of 32.5 and 37.5, then the severity cut score would be 35). The location of vignettes was chosen so that the mean would be an integer value, for ease of use. It is important to note that mean T-score for these domains crosses that of general US population (50), with the range of scores around this mean.

Focus group discussion—Nine patients and ten providers participated in the expert panel meetings. The baseline demographics and RD diagnosis of the patient panelists are shown in Table 2.

Eight of the nine patients were women. Patients began the workshop by completing a PROMIS global health short form. The mean (SD) of the physical health T-score was 38 (± 7), and mental health T-score was 46 (± 10). The ten clinical providers comprised of practicing rheumatologists (6), rheumatology nurse practitioners (2), rheumatology fellow-in-training (1), and an occupational therapist with expertise in RD. Based on provider consensus in Phase 1 and T-scores of PROMIS measures from the Phase 2, we chose the following four PROMIS measures for phase 3 – physical function, pain interference, sleep disturbance, and depression. Clinical vignettes were created for these measures based on the symptom severity.

Table 3 summarizes the consensus-derived cut-scores and table 4 shows the ranking of the vignettes for the four domains and classification into different severity categories, by the patient and provider panelists.

Across all measures, patients classified the categories with a T-score distribution that was within 0.5-1 SD, with the exceptions of 'mild impairment' in physical function (65-45) and 'moderate impairment' in sleep (45-60). The pattern was similar with providers, with the exception of mild impairment (60-45) and moderate impairment (45-25) in physical function. The patients selected cut-points at lower level of impairments than the providers across all measures. Except for pain interference, patients classified lower T-score as indicative of 'no problem'. For example, the cut-score in the patient classification indicating 'no problem' in physical function was >65, compared to >60 in provider classification (sleep disturbance <35 vs <45, depression <45 vs <55). The same trends were observed across other categories in all PROMIS measures. For example, patients judged sleep disturbance to be moderately impaired at lower scores (45-60) than did providers (55-65). Likewise, patients judged physical function to be severely impaired at a higher score (<35) than did providers (<25). When comparing results across measures, there was more agreement on the cut-points for pain interference across panelists. The panelists agreed on selected cut-points for all categories, except for severe category in pain interference, where the cut-points were within 0.5 SDs of each other (Table 3).

We considered the proportion of patients in the UM cohort that would fall into each severity category according to the panelists' cut-scores (Table 3). Based on these, a majority of patients (75-96%) would be classified as having mild to moderate problems with physical function. There was large variability in patients classified as having severity impairment in physical function. Across all measures, patient panelists categorized a higher percentage of the scenarios as indicating severe impairment, compared to the providers – physical function (24% vs 2%), pain interference (27% vs 11%), sleep disturbance (33% vs 14%), depression (22% vs 8%). For pain interference, 53% of patients were categorized to have none or mild problems, while 47% had moderate or severe problems. With regard to sleep disturbance, considerable variability was observed at higher severity levels categorized as moderate and severe problems (86% by patient cut-point, vs 51% by provider cut-point). With respect to depression, there was variability in classification of patients based on the panel groups. Based on the patient cut-point, 14% of patients were classified as not having depression (vs 67% by provider cut-point), 53% were classified to have mild depression (vs 11% by provider cut-point), and 22% had severe depression (vs 8% by provider cut-point).

Discussion

Using an iterative process, we have used focus group consensus and a modified educational standard setting method to choose PROMIS domains that are actionable in clinical care and estimated clinically relevant cut-scores for PROMIS measures using clinical vignettes developed from patient with different RD. There is increased interest in integrating PRO measures to electronic health records (EHR) and many barriers / steps to accomplish this goal have been identified [17]. Rather than prescribing one 'right way' in achieving this aim, the consensus is to consider different options depending on the organization and the context for this integration. Some of the key questions to facilitate PRO integration to EHR include – (i) which outcomes are important to measure for a given population?, (ii) how should the responses be interpreted, (iii) how should PRO data be displayed, and (iv) how will the

providers act upon. We have attempted to answer some of these questions which is relevant to RD population.

Our rheumatologists chose four PROMIS domains: physical function, pain interference, sleep disturbance, and depression, based on their importance in published literature and they were considered actionable at the point-of-care. Recently published literature supports our choice of the PROMIS domains [19]. In a prospective cohort of RA patients who completed PROMIS questionnaires across 11 domains, significant impairment was reported in physical function, pain and fatigue scales, sleep disturbance and emotional distress [20]. An international group of RA patients and providers agreed on following as essential domains for describing RA flare - pain, physical function, sleep, emotional distress and fatigue (latter three not endorsed by the providers) [21]. Further, the American College of Rheumatology (ACR) in collaboration with the National Committee for Quality Assurance, endorsed the measurement of functional status in rheumatoid arthritis as part of the Physician Quality Reporting System. ACR recommends the PROMIS® physical function scale as one of the measures to assess functional status [7]. In a cross sectional study of SLE patients, mean PROMIS T-scores for fatigue, pain interference, sleep disturbance, and physical function were worse than the the general US population[22; 23]. These findings are similar to the results from phase 2 of our study, where mean PROMIS T-scores for physical function, pain interference, sleep disturbance, fatigue, and depression were worse than the general US population. Fatigue is a commonly endorsed symptom by patients with RDs [24–26]. Despite a high mean T-score in the phase 2, fatigue was excluded as a domain for phase 3 as the clinicians felt it was multifactorial with many putative causes, and hence, difficult to intervene at different levels of severity in a majority of patients with RD [27; 28]. A critical aspect of our study was that clinicians had to choose domains across patients with different RDs, keeping in mind how they could intervene in routine practice if these domains were impaired. This may have prompted them to exclude fatigue.

A goal of this study is to incorporate chosen PROMIS measures into electronic health record (EHR) to enable practicing clinicians to use these measures at the point of care in the RD population. This step in turn requires the inclusion of actionable options to guide clinicians in the management of RD patients who report varying severity of impairment in chosen PROMIS Measures. Based on input from the providers, we have developed a PROMIS report card that alerts providers if the scores are in the “concerning zone”, usually reflecting the severe problem scores in Table 3, where a provider may decide to intervene (such as referral for physical therapy, increase immunosuppressive therapy, referral for joint replacement). A similar approach has recently been described in gynecological cancer patients receiving chemotherapy [29]. These patients completed PROMIS CAT assessments and psychosocial needs assessment on a secure website up to 3 days prior to a clinic visit. Patients who were unable to complete the assessment were asked to complete a survey in the office on the day of appointment. PROs were automatically scored and saved in the patient’s EHR. Scores that exceed a pre-determined and validated threshold for severity were flagged within the EHR, and generated an automated message to appropriate staff. This approach has been recently endorsed in yet another oncology study and also in the Patient-Centered Outcomes Research Institute (PCORI) funded users’ guide [17; 30]. In the study, PRO data was circulated by internet survey to cancer patients/survivors, oncology clinicians, and PRO

researchers. The data was color coded – normal score in green and concerning core in red, with red threshold lines between normal and concerning. The interpretation by survey responders was more accurate when PRO data was presented with threshold lines indicating normal versus concerning scores.

In our study, we observed a consistent pattern of divergence in the cut-points across the different domains as scored by patients and providers, in that patients ranked cut-points at lower level of impairments than did providers. Except for pain interference, there was some divergence across all severity levels in sleep disturbance and depression, and at higher severity levels in physical function. These findings are consistent with other reports of discrepancy in symptom ratings between patients and clinicians [31]. This divergence can be partly explained by the fact that sleep disturbance, depression and functional impairment are problems which many RD patients experience and live with, yet providers may be less familiar with these problems. Patients and providers may emphasize different aspects of the symptoms or different latent constructs of a given domain. Hence, the responses may be contradictory. Since the providers' focus group was conducted on the day after patients' group, we asked the providers why there may be differences in the cut points. Providers felt that higher specificity was needed for cut-points that warranted definite intervention (as in moderate and severe problems in physical function), for which red-flags would be raised in the electronic medical records. However, when our study is compared to studies establishing cut-points in MS and JIA cohort, which employed a similar 'bookmarking' methodology, there are observable differences. In the MS study, congruence in the cut-points was noted between patients and providers across all chosen domains [14]. In the JIA study, congruence in estimated cut-scores across panel groups (patients, parent and providers) was noted for upper extremity function and fatigue; however, some divergence was seen for mobility and pain interference. When there was divergence, patients chose cut-scores at highest dysfunction and parents at the lowest dysfunction for severity classifications. [16]. The second observation is in the 'mild problem' category where the scores are less impaired in our study. For example, in the PROMIS physical function measure, the T-score range in 'mild problem' category in our study is 65- 45 (patients) and 60-45 (providers). In MS study, T-score range is 50-40 (patients and clinicians), and in JIA study, the T-score range is 40-30 (patients) and 45-30 (clinicians). Both these observations could be due to a principal difference between these studies and our study - the cut-points were decided by patients and providers from the lens of an actionable intervention in clinical practice (in our study). In an oncology setting, although a parallel methodology was not employed (patient and physician paper-based surveys with a different scoring system), interesting observations may be noted [31]. When reporting of symptom severity between patients and physicians was compared, the agreement was higher for symptoms that could be observable directly and needed active management, such as vomiting and diarrhea, than for subjective symptoms, such as fatigue and dyspnea. So, the important question to consider in future research is whether patients and providers choose cut-points differently when the anchor of intervention is taken into account.

There is another observation which seeks clarification. Some of the scores in the 'mild problem' category are below (denoting better scores) the standardized mean of the US general population. For example, the T-score range for sleep disturbance in the 'mild

problem' category is 35-45 (patients) and 45-55 (providers). The score range by patients is below the population mean of 50. Firstly, the US general population is a 'representative' population with a mix of both healthy subjects and patients with chronic medical conditions. Secondly, the clinical vignettes were created based on symptom severity obtained from real-patient data in phase 2, and taking into consideration the minimum and maximum scores for each measure. Our RD patients felt that the sleep disturbance vignettes representing a score range of 45-60 needed a higher level of intervention compared to those in the score range of 35-45. Again, the intervention angle may have influenced to include a larger score range in the 'moderate problem' category.

This study has many strengths. We followed an iterative consensus methodology which was data-driven. Thus, our study serves as launching pad for PROMIS measures in a large academic setting with EMR capabilities. Second, a wide range of PROMIS CAT measures were administered in patients with various RD. Third, a representative patient sample with a diversity of RD was enrolled and real-person data was used to develop our scenarios. A mix of providers involved in the care of RD patient was enrolled for participation in the provider group (rheumatologists, nurse practitioners, occupational therapist). Finally, this is one of the few studies to establish cut-points at different levels of severity for chosen PROMIS measures in such a diverse patient population.

The study is not without limitations. First, we did not seek input from patients regarding the choice of domains and development of scenarios for phase 3. We based our decision on which domains are important on the judgment of the clinicians, who considered what was within the realm of rheumatology care and actionable in day-to-day practice. The choice of domains was based on published data indicating the relevance of these domains in RDs [20–22]. The PROMIS physical function measure has been endorsed by the American College of Rheumatology for functional status assessment of RA patients at least once a year [7]. Second, we did not capture disease activity and severity of the patients who participated in the focus group and how their current status may have affected their ranking.

Conclusions

In conclusion, this study describes the reporting of different PROMIS CAT measures in the RD population. Parallel exercises identified the cut points from the perspectives of patients with RD, and the clinical providers who treat rheumatic diseases. This allows for meaningful interpretation of PROMIS® measures in a clinical setting in RD population. Further work is focused on incorporating these cut points into clinical practice and its impact on clinical care.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The authors thank National Institutes of Health and National Institute of Arthritis and Musculoskeletal and Skin Diseases for funding the study and Dr. D. Khanna (NIH/NIAMS K24 AR 063120), and Dr. Young (NIH/NIAMS T32-AR007080-38).

References

1. Khanna D, Krishnan E, Dewitt EM, Khanna PP, Spiegel B, Hays RD. The future of measuring patient-reported outcomes in rheumatology: Patient-Reported Outcomes Measurement Information System (PROMIS). *Arthritis Care Res (Hoboken)*. 2011; 63(Suppl 11):S486–490. [PubMed: 22588770]
2. Prevention, C. f. D. C. a. Measuring healthy days: Population assessment of health-related quality of life. Atlanta: CDC; 2000. p. 4-6.
3. Devilliers H, Amoura Z, Besancenot JF, Bonnotte B, Pasquali JL, Wahl D, Maurier F, Kaminsky P, Pennaforte JL, Magy-Bertrand N, Arnaud L, Binquet C, Guillemin F, Bonithon-Kopp C. Responsiveness of the 36-item Short Form Health Survey and the Lupus Quality of Life questionnaire in SLE. *Rheumatology (Oxford)*. 2015; 54(5):940–949. [PubMed: 25361539]
4. Husted JA, Gladman DD, Farewell VT, Cook RJ. Health-related quality of life of patients with psoriatic arthritis: a comparison with patients with rheumatoid arthritis. *Arthritis Rheum*. 2001; 45(2):151–158. [PubMed: 11324779]
5. Uhlig T, Loge JH, Kristiansen IS, Kvien TK. Quantification of reduced health-related quality of life in patients with rheumatoid arthritis compared to the general population. *J Rheumatol*. 2007; 34(6): 1241–1247. [PubMed: 17516624]
6. Health, U. S. D. o., Human Services, F. D. A. C. f. D. E., Research, Health, U. S. D. o., Human Services, F. D. A. C. f. B. E., Research, Health, U. S. D. o., Human Services, F. D. A. C. f. D., & Radiological, H. Guidance for industry: patient-reported outcome measures: use in medical product development to support labeling claims: draft guidance. *Health Qual Life Outcomes*. 2006; 4:79. [PubMed: 17034633]
7. Singh JA, Saag KG, Bridges SL Jr, Akl EA, Bannuru RR, Sullivan MC, Vaysbrot E, McNaughton C, Osani M, Shmerling RH, Curtis JR, Furst DE, Parks D, Kavanaugh A, O'Dell J, King C, Leong A, Matteson EL, Schousboe JT, Drevlow B, Ginsberg S, Grober J, St Clair EW, Tindall E, Miller AS, McAlindon T. 2015 American College of Rheumatology Guideline for the Treatment of Rheumatoid Arthritis. *Arthritis Rheumatol*. 2016; 68(1):1–26.
8. Witter JP. The Promise of Patient-Reported Outcomes Measurement Information System—Turning Theory into Reality: A Uniform Approach to Patient-Reported Outcomes Across Rheumatic Diseases. *Rheum Dis Clin North Am*. 2016; 42(2):377–394. [PubMed: 27133496]
9. Cella D, Yount S, Rothrock N, Gershon R, Cook K, Reeve B, Ader D, Fries JF, Bruce B, Rose M, Group, P. C. 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):S3–S11.
10. Reeve BB, Hays RD, Bjorner JB, Cook KF, Crane PK, Teresi JA, Thissen D, Revicki DA, Weiss DJ, Hambleton RK, Liu H, Gershon R, Reise SP, Lai JS, Cella D, Group, P. C. Psychometric evaluation and calibration of health-related quality of life item banks: plans for the Patient-Reported Outcomes Measurement Information System (PROMIS). *Med Care*. 2007; 45(5 Suppl 1):S22–31. [PubMed: 17443115]
11. Hays RD, Liu H, Spritzer K, Cella D. Item response theory analyses of physical functioning items in the medical outcomes study. *Med Care*. 2007; 45(5 Suppl 1):S32–38. [PubMed: 17443117]
12. Khullar OV, Rajaei MH, Force SD, Binongo JN, Lasanajak Y, Robertson S, Pickens A, Sancheti MS, Lipscomb J, Gillespie TW, Fernandez FG. Pilot Study to Integrate Patient Reported Outcomes After Lung Cancer Operations Into The Society of Thoracic Surgeons Database. *Ann Thorac Surg*. 2017; 104(1):245–253. [PubMed: 28483154]
13. Karantonis A, Sireci SG. The bookmark standard-setting method: A literature review. *Educational Measurement: Issues and Practice*. 2006; 25(1):4–12.
14. Cook KF, Victorson DE, Cella D, Schalet BD, Miller D. Creating meaningful cut-scores for Neuro-QOL measures of fatigue, physical functioning, and sleep disturbance using standard setting with patients and providers. *Qual Life Res*. 2015; 24(3):575–589. [PubMed: 25148759]
15. Cella D, Choi S, Garcia S, Cook KF, Rosenbloom S, Lai JS, Tatum DS, Gershon R. Setting standards for severity of common symptoms in oncology using the PROMIS item banks and expert judgment. *Qual Life Res*. 2014; 23(10):2651–2661. [PubMed: 24938431]

16. Morgan EM, Mara CA, Huang B, Barnett K, Carle AC, Farrell JE, Cook KF. Establishing clinical meaning and defining important differences for Patient-Reported Outcomes Measurement Information System (PROMIS(R)) measures in juvenile idiopathic arthritis using standard setting with patients, parents, and providers. *Qual Life Res.* 2017; 26(3):565–586. [PubMed: 27913986]
17. University, J. H. Users' Guide to Integrating Patient-Reported Outcomes in Electronic Health Records. 2017. Retrieved July 29, 2017, from <https://www.pcori.org/document/users-guide-integrating-patient-reported-outcomes-electronic-health-records>
18. Cella D, Riley W, Stone A, Rothrock N, Reeve B, Yount S, Amtmann D, Bode R, Buysse D, Choi S, Cook K, Devellis R, DeWalt D, Fries JF, Gershon R, Hahn EA, Lai JS, Pilkonis P, Revicki D, Rose M, Weinfurt K, Hays R, Group, P. C. 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(11):1179–1194. [PubMed: 20685078]
19. Idzerda L, Rader T, Tugwell P, Boers M. Can we decide which outcomes should be measured in every clinical trial? A scoping review of the existing conceptual frameworks and processes to develop core outcome sets. *J Rheumatol.* 2014; 41(5):986–993. [PubMed: 24584917]
20. Bartlett SJ, Orbai AM, Duncan T, DeLeon E, Ruffing V, Clegg-Smith K, Bingham CO 3rd. Reliability and Validity of Selected PROMIS Measures in People with Rheumatoid Arthritis. *PLoS One.* 2015; 10(9):e0138543. [PubMed: 26379233]
21. Bartlett SJ, Hewlett S, Bingham CO 3rd, Woodworth TG, Alten R, Pohl C, Choy EH, Sanderson T, Boonen A, Bykerk V, Leong AL, Strand V, Furst DE, Christensen R, Group, O. R. F. W. Identifying core domains to assess flare in rheumatoid arthritis: an OMERACT international patient and provider combined Delphi consensus. *Ann Rheum Dis.* 2012; 71(11):1855–1860. [PubMed: 22772326]
22. Mahieu MA, Ahn GE, Chmiel JS, Dunlop DD, Helenowski IB, Semanik P, Song J, Yount S, Chang RW, Ramsey-Goldman R. Fatigue, patient reported outcomes, and objective measurement of physical activity in systemic lupus erythematosus. *Lupus.* 2016
23. Jensen RE, Rothrock NE, DeWitt EM, Spiegel B, Tucker CA, Crane HM, Forrest CB, Patrick DL, Fredericksen R, Shulman LM, Cella D, Crane PK. The role of technical advances in the adoption and integration of patient-reported outcomes in clinical care. *Med Care.* 2015; 53(2):153–159. [PubMed: 25588135]
24. Basu N, Jones GT, Fluck N, MacDonald AG, Pang D, Dospinescu P, Reid DM, Macfarlane GJ. Fatigue: a principal contributor to impaired quality of life in ANCA-associated vasculitis. *Rheumatology (Oxford).* 2010; 49(7):1383–1390. [PubMed: 20400759]
25. Druce KL, Jones GT, Macfarlane GJ, Basu N. Patients receiving anti-TNF therapies experience clinically important improvements in RA-related fatigue: results from the British Society for Rheumatology Biologics Register for Rheumatoid Arthritis. *Rheumatology (Oxford).* 2015; 54(6):964–971. [PubMed: 25313148]
26. Lai JS, Beaumont JL, Jensen SE, Kaiser K, Van Brunt DL, Kao AH, Chen SY. An evaluation of health-related quality of life in patients with systemic lupus erythematosus using PROMIS and Neuro-QoL. *Clin Rheumatol.* 2017; 36(3):555–562. [PubMed: 27848056]
27. Druce KL, Jones GT, Macfarlane GJ, Basu N. Determining Pathways to Improvements in Fatigue in Rheumatoid Arthritis: Results From the British Society for Rheumatology Biologics Register for Rheumatoid Arthritis. *Arthritis Rheumatol.* 2015; 67(9):2303–2310. [PubMed: 26108189]
28. Hifinger M, Putrik P, Ramiro S, Keszzi AP, Hmamouchi I, Dougados M, Gossec L, Boonen A. In rheumatoid arthritis, country of residence has an important influence on fatigue: results from the multinational COMORA study. *Rheumatology (Oxford).* 2016; 55(4):735–744. [PubMed: 26683195]
29. Wagner LI, Spiegel D, Pearman T. Using the science of psychosocial care to implement the new american college of surgeons commission on cancer distress screening standard. *J Natl Compr Canc Netw.* 2013; 11(2):214–221. [PubMed: 23411387]
30. Snyder CF, Smith KC, Bantug ET, Tolbert EE, Blackford AL, Brundage MD, Board PRODPSA. What do these scores mean? Presenting patient-reported outcomes data to patients and clinicians to improve interpretability. *Cancer.* 2017; 123(10):1848–1859. [PubMed: 28085201]

31. Basch E, Iasonos A, McDonough T, Barz A, Culkin A, Kris MG, Scher HI, Schrag D. Patient versus clinician symptom reporting using the National Cancer Institute Common Terminology Criteria for Adverse Events: results of a questionnaire-based study. *Lancet Oncol.* 2006; 7(11): 903–909. [PubMed: 17081915]

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Research involving Human Participants

Ethics approval

The study was approved by the University of Michigan Institutional Review Board. All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional review board, and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

Informed consent

Informed consent was obtained from all individuals who participated in the phase 2 of the study, and from patients and providers who participated in the panel group discussions.

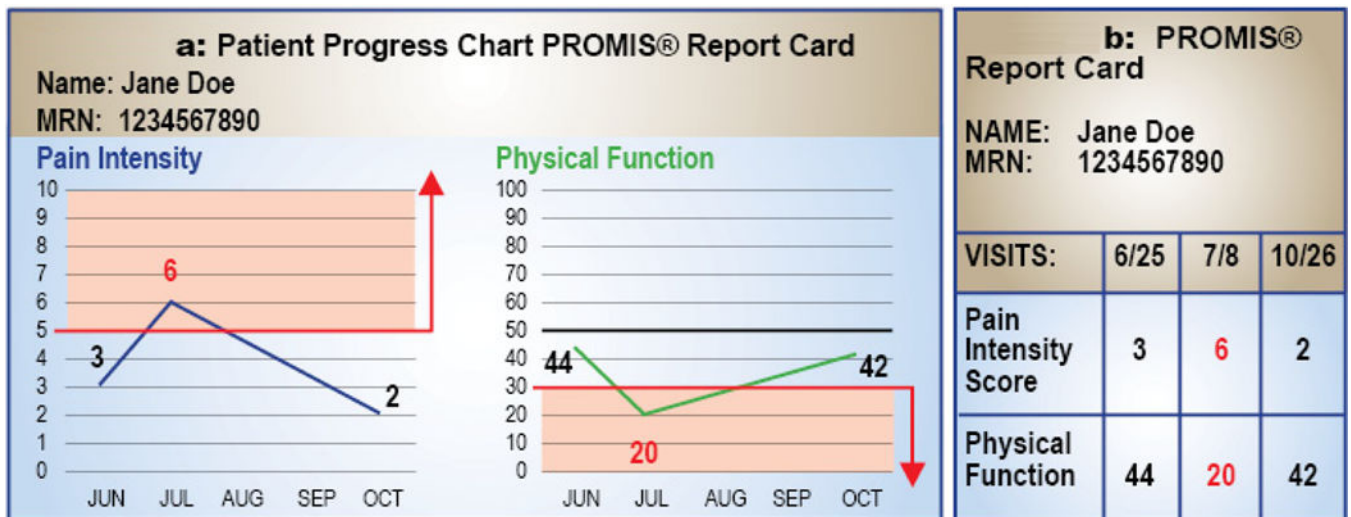


Figure 1.

*A score of greater than or equal to 5 is indicative of severe pain (Concerning zone – shaded grey area). Consider work-up of underlying pain and appropriate treatment / referral.

** The mean score for US population is 50. A higher score indicates better physical function. A score less than 30 is indicative of severe physical limitations (see Concerning zone – shaded grey area). Consider optimizing therapy – such as physical therapy or change in medication – or re-evaluation at follow-up visit.

Table 1

Baseline demographic data of the convenience sample

		N	%
Total patients		217	
Age in years, Mean \pm SD		53 \pm 14	
Gender, N(%)	Females	169	78
	Males	48	22
Race, N(%)	Caucasian	178	82
	African American	11	5
	Asian	2	1
	Not provided	26	12
Self-reported Diagnoses, (N)%			
	Rheumatoid arthritis	48	
	Systemic Sclerosis	35	
	Systemic Lupus Erythematosus	24	
	Overlap Connective Tissue Diseases	20	
	ANCA associated vasculitis	13	
	Osteoarthritis	11	
	Undifferentiated Connective Tissue Diseases	10	
	Fibromyalgia	8	
	Myositis	6	
	Psoriatic arthritis	6	
	Other Inflammatory arthropathy	6	
	Other *	30	
Mean T-score metrics across 13 PROMIS domains, Mean \pm SD	PROMIS Domain	T-score	
	Anger	52 [\pm 9]	
	Anxiety	54 [\pm 9]	
	Depression	53 [\pm 8]	
	Fatigue	58 [\pm 10]	
	Pain behavior	55 [\pm 8]	
	Pain interference	58 [\pm 10]	
	Physical function	40 [\pm 8]	
	Physical function with mobility aid	42 [\pm 10]	
	Satisfaction with social roles and activities	44 [\pm 10]	
	Sleep disturbance	55 [\pm 9]	
	Sleep related behavior	56 [\pm 9]	
	Social isolation	47 [\pm 9]	
	Ability to participate in social activities	47 [\pm 9]	

SD standard deviation, ANCA Anti-Neutrophilic Cytoplasmic Antibodies,

* Other conditions (n): Spondyloarthritis (4), Sjogren's syndrome (4), Gout (3), Morphea (3), Inflammatory bowel disease associated arthritis (2), Interstitial Lung Disease (2), Low back pain (2), Reactive arthritis (2), Relapsing polychondritis (2), Adult-onset asthma and periocular xanthogranuloma (1), Behcet's disease (1), Cicatricial pemphigoid (1), Giant cell arteritis (1), Polymyalgia rheumatic (1), Urticarial vasculitis (1)

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 2

Patient panelists – demographics, diagnoses and PROMIS Global scores

Patient panelists	Age in years	Gender	Race	Diagnosis	PROMIS Global short form	
					Physical health T-score	Mental health T-score
Patient 1	44	Female	Caucasian/ White	Sjogren's syndrome	32	51
Patient 2	59	Male	Caucasian/ White	Psoriatic arthritis, osteoarthritis, Gout	32	36
Patient 3	69	Female	Caucasian/ White	Rheumatoid arthritis	54	62
Patient 4	53	Female	Caucasian/ White	Osteoarthritis	40	56
Patient 5	53	Female	Caucasian/ White	Fibromyalgia	40	34
Patient 6	50	Female	Caucasian/ White	Gout	30	43
Patient 7	62	Female	Caucasian/ White	Idiopathic Inflammatory Myopathy	35	34
Patient 8	63	Female	Caucasian/ White	Vasculitis	35	51
Patient 9	58	Female	Caucasian/ White	Systemic Lupus Erythematosus	42	48
					Mean [± SD] 38 [± 7]	Mean [± SD] 46 [± 10]

Table 3

Consensus cut-scores by domains and expert panel

Domain	Categories	Cut scores		Severity level when applied to UM cohort			
		Patient classification	Provider classification	Patient classification		Provider classification	
				%	N	%	N
Physical Function	No problem	>65	>60	0.4	1	1	3
	Mild problem	65-45	60-45	27	63	26	61
	Moderate problem	45-35	45-25	48	113	70	164
	Severe problem	<35	<25	24	56	2	5
Pain Interference	No problem	<50	<50	14	32	14	32
	Mild problem	50-60	50-60	39	90	39	90
	Moderate problem	60-65	60-70	20	46	36	84
	Severe problem	>65	>70	27	64	11	26
Sleep disturbance	No problem	<35	<45	2	5	14	32
	Mild problem	35-45	45-55	12	27	35	80
	Moderate problem	45-60	55-65	53	124	37	87
	Severe problem	>60	>65	33	76	14	33
Depression	No problem	<45	<55	14	33	67	157
	Mild problem	45-55	55-60	53	124	11	26
	Moderate problem	55-60	60-65	11	26	14	34
	Severe problem	>60	>65	22	52	8	18

Table 4

Ranking of case vignettes by patients and provider panelists

Physical function												
Vignettes	1	2	3	4	5	6	7	8	9	10	11	12
Patients	>65		65 thru 45			45 thru 35			<35			
	None		Mild			Moderate			Severe			
Providers	>60			60 thru 45			45 thru 25			<25		
	None			Mild			Moderate			Severe		
Pain interference												
Vignettes	1	2	3	4	5	6	7	8				
Patients	<50			50 thru 60		60 thru 65		>65				
	None			Mild		Moderate		Severe				
Providers	<50			50 thru 60		60 thru 70			>70			
	None			Mild		Moderate			Severe			
Sleep disturbance												
Vignettes	1	2	3	4	5	6	7	8	9	10		
Patients	<35		35 thru 45		45 thru 60			>60				
	None		Mild		Moderate			Severe				
Providers	<45			45 thru 55		55 thru 65		>65				
	None			Mild		Moderate		Severe				
Depression												
Vignettes	1	2	3	4	5	6	7	8	9			
Patients	<45		45 thru 55		55 thru 60		>60					
	None		Mild		Moderate		Severe					
Providers	<55			55 thru 60		60 thru 65		>65				
	None			Mild		Moderate		Severe				