HHS Public Access

Author manuscript

Arthritis Care Res (Hoboken). Author manuscript; available in PMC 2023 January 01.

Published in final edited form as:

Arthritis Care Res (Hoboken). 2022 January ; 74(1): 151–160. doi:10.1002/acr.24433.

Measurement of minimal disease activity in psoriatic arthritis using PROMIS-Physical Function or the Health Assessment Questionnaire-Disability Index

Erin Chew, MD[1,2], **Jamie Perin, PhD**[3], **Thomas Grader-Beck, MD**[1], **Ana-Maria Orbai, MD, MHS**[1]

[1]Johns Hopkins University School of Medicine Division of Rheumatology, Baltimore, MD

[2]Johns Hopkins Hospital. Baltimore, MD

[3]Johns Hopkins University School of Public Health, Department of International Health, Baltimore

Abstract

Objective: To assess the interchangeability of the Health Assessment Questionnaire-Disability Index (HAQ-DI) with the Patient-Reported Outcomes Measurement Information System-Physical Function (PROMIS-PF) in the calculation of minimal disease activity (MDA) in psoriatic arthritis (PsA).

Methods: Comprehensive PsA disease activity was collected concomitantly with HAQ-DI and PROMIS-PF measures in a PsA cohort. PROMIS-based MDA definitions were built using the existing cross-walk between the scores: HAQ-DI≤0.5 equivalent to a PROMIS-PF Tscore≥41.3. We assessed agreement between MDA (MDA HAQ-DI) and PROMIS-PF MDA definitions (MDA PROMIS-PF4a, MDA PROMIS-PF Bank) at each visit and longitudinally (MDA state changes between consecutive visits) through the kappa statistic. The predictive value of MDA PROMIS-PF for MDA HAQ-DI was assessed using ROC curve analysis.

Results: One hundred participants contributed 352 observations with up to five visits. Mean (SD) age was 52 (12) years, 60% were female, and 43% were in MDA at baseline. Kappa statistic for PROMIS-PF based MDA reflected excellent agreement with HAQ-DI MDA: kappa=0.94 (95% CI 0.90-0.97) for MDA PROMIS-PF Bank, and kappa=0.90 (95% CI 0.80-0.95) for MDAPROMIS-PF4a. Higher longitudinal agreement was seen between MDA HAQ-DI and MDA PROMIS-PF Bank versus MDA PROMIS-PF4a between consecutive visits: kappa ranged between 0.81-0.94 versus 0.72-0.84, respectively. Area under ROC curve for predicting MDA HAQ-DI was 0.97 for MDA PROMIS-PF Bank and 0.95 for MDA PROMIS-PF4a.

Conclusion: Excellent agreement was seen between HAQ-DI and PROMIS-based MDA definitions statically and longitudinally. The PROMIS-PF Bank and PROMIS-PF4a are accurate replacements for the HAQ-DI in calculating MDA state in PsA.

Corresponding Author: Ana-Maria Orbai, MD, MHS, Assistant Professor of Medicine, Director Psoriatic Arthritis Program, Johns Hopkins University School of Medicine, Division of Rheumatology, 5200 Eastern Avenue, Baltimore, MD 21224, aorbai1@jhmi.edu, Phone: 410-55-8231.

Keywords

Psoriatic Arthritis; PROMIS; HAQ-DI; Patient Reported Outcomes; Treat-to Target

Psoriatic arthritis (PsA) is an autoimmune disease that affects up to 1% of the US population and about 1:3 people living with the skin disease psoriasis. PsA is heterogeneous in pathophysiology, affecting the joints, entheses, digits, spine, skin, and skin appendages. Its impact on quality of life is equally broad and manifests with symptoms of pain, fatigue, depression/anxiety, as well as decreased physical function and social participation, disability, and work loss.[1,2]

The PsA treat-to-target (T2T) state was established through consensus among international experts,[3] and provisionally endorsed by the American College of Rheumatology and the National Psoriasis Foundation[4]. Minimal disease activity (MDA) is a PsA T2T state defined by meeting pre-specified criteria for disease activity across PsA pathophysiologic manifestations (swollen and tender joints, enthesitis, psoriasis) and patient reported outcomes (PRO) (physical function, pain, and patient global assessment of psoriatic disease).[5] The original MDA criteria capture patient reported physical function through the Health Assessment Questionnaire-Disability Index (HAQ-DI).[6]

The Patient-Reported Outcomes Measurement Information System (PROMIS)[7] is a library of PRO instruments developed using state-of-the art psychometric science, and normed to the general US population (a T-score of 50 represents the US population mean and the standard deviation is 10 points).[8] PROMIS measures are increasingly available in medical records and can be incorporated into routine care.[8] Through the PROsetta Stone® project,[9] “walkways” have been developed between PROMIS scores and commonly used legacy instruments such as linking PROMIS Physical Function (PF) scores with HAQ-DI scores.[10] Specifically, for assessing physical function, it is increasingly relevant to adapt PRO to include activities that reflect therapeutic advances in rheumatology.[11] PROMIS measures have been developed using qualitative research as the basis of item content and item formulation for each questionnaire/item bank and assessments are focused on each individual’s physical function ability, not merely on the lack of disability and/or frequency of tasks performed.[12] From this perspective, the PROMIS Physical function items cover the basic activities of daily living (walking, dressing) and also complex activities (dancing, jogging, taking part is sports, and strenuous activities); items are formulated in the present tense using simple syntax for clarity and comprehension, and each item has four or five response options instead of three, as is the case with the HAQ-DI, to reduce floor and ceiling effects and provide greater discrimination. [12]

Thus, while physical function has been assessed for a long time using the HAQ-DI, it is of interest to transition to more current population normed instruments such as PROMIS-PF.[10] Schalet et al. conducted a single-group design study using a large standardization sample centered on the 2000 US census and linked legacy physical function PRO to the PROMIS-PF scale. Thus, there now exists a common reporting metric that can support transition from legacy instruments to PROMIS-PF scales.

In our study, we sought to determine whether PROMIS-PF can replace the HAQ-DI as a measure of physical function to accurately classify the MDA T2T state in PsA. The objective was to compare agreement between the routine HAQ-DI based MDA definition and PROMIS-based MDA definitions in a PsA cohort where we conducted longitudinal PsA-specific disease status assessments.

Patients and Methods

The Johns Hopkins Psoriatic Arthritis Cohort

The Johns Hopkins Psoriatic Arthritis cohort is approved by The Johns Hopkins Institutional Review Board (IRB00063222). All study subjects signed written informed consent prior to participating in the study. Research visits were conducted every 3-6 months in conjunction with guideline-based rheumatologic care for PsA.

Adult patients with rheumatologist-diagnosed PsA were eligible to participate if they met Classification Criteria for Psoriatic Arthritis (CASPAR).[13] At each visit, PsA specific measures included: tender joint count (range 0–68), swollen joint count (0–66),[14] enthesitis count using the Leeds Enthesitis Index (0–6),[15] active tender dactylitis count (0–20), percent body surface area affected by active psoriasis (0–100%), Pain numeric rating scale (NRS) (0–10), Patient global psoriatic disease NRS (0–10), Patient global psoriatic arthritis NRS (0–10) and the HAQ-DI (0–3).

In addition, we collected PROMIS PF measures including the PROMIS bank v1.2 -Physical Function, and the PROMIS Short Form v2.0 Physical Function 4a (collected as part of the PROMIS Profile-29 v1.0)[8]. Participants completed all questionnaires in the clinic room, prior to the rheumatology visit with the physician, through self-report and without assistance.

Measures

HAQ-DI.—The HAQ-DI is a legacy physical function PRO developed by the Stanford Arthritis Center for Rheumatoid Arthritis.[6] It has been used in every PsA randomized controlled trial as part of the American College of Rheumatology response criteria and is also included in the MDA definition. The HAQ-DI consists of 20 questions in eight categories (Dressing and Grooming, Hygiene, Arising, Reach, Eating, Grip, Walking, Outside Activities). Each item has four response options, ranging from “No difficulty” to “Unable to do,” corresponding to scores from 0 to 3.[6] Lower HAQ-DI scores mean better function. Minimally important difference, or the smallest improvement considered to be clinically important in PsA, is defined by a longitudinal improvement in HAQ-DI of 0.35 points.[16] The PsA minimal disease activity criterion for HAQ-DI is met by a score of ≤0.5.[5] The HAQ-DI was administered on paper clinical research forms concomitantly with the pain and patient global NRS.

PROMIS Instruments.—The PROMIS instruments were developed using item response theory by the National Institute of Health (NIH).[7] PROMIS scores are normed to the US population, and expressed as T-scores with a mean of 50 (representing the US population mean for the measure as the reference) and a standard deviation of 10. Higher PROMIS-PF

scores mean better function. The PROMIS-PF measures can be administered either as fixed-content short forms or as a computer adaptive (CAT) test that selects items from the entire physical function item bank (PROMIS-PF Bank). For CAT administration, items from the PROMIS Physical Function bank are dynamically selected based on a patient's prior response to precisely capture each patient's functional status.[17,18] The equivalent PROMIS-PF T-score for a HAQ-DI score ≤0.5 has been defined as ≥41.3.[10] Participants completed the fixed PROMIS-PF short form 4a which includes PROMIS items: PFA11, PFA21, PFA23, and PFA53 (Supplement Figure 1). They also completed the PROMIS Bank v1.2 - Physical Function, administered on a tablet through the Assessment Center platform (www.assessmentcenter.net) using CAT limited to maximum 8 items, followed by the PROMIS-PF 4a. The short form was programmed without repetition with the CAT. English language versions, developed for adult participants, were used in the study. Scoring was performed automatically through the Assessment Center platform and results downloaded. Reporting of study results is being done in accordance with the recently published "Reporting checklist for ASCQ-Me, Neuro-QoL, NIH Toolbox Emotion, and PROMIS Measures" [19].

Treat-to-target (T2T) states.—The MDA criteria are listed in Supplement Table 1. If five out of the seven criteria are met [tender joint count (0-68)≤1, swollen joint count (0-66) ≤1, enthesitis (0-6)≤1, pain NRS (0-10)≤1.5, patient global assessment psoriatic disease (0-10)≤2, HAQ-DI (0-3)≤0.5, body surface area affected by psoriasis (0-100)≤3%), then PsA disease activity corresponds to the MDA state and the T2T objective has been achieved. We also examined the very low disease activity state (VLDA) defined as all seven MDA criteria being met.[5] In addition, we calculated the clinical Disease Activity in Psoriatic Arthritis (cDAPSA) score, defined as the sum of the tender (0-68) and swollen (0-66) joint counts, pain NRS (0-10) and patient global psoriatic arthritis NRS (0-10). Clinical DAPSA disease activity thresholds were defined as remission (REM≤4), low (LDA>4 & ≤13), moderate (MoDA>13 & ≤27), and high disease activity (HDA>27).[20] A cDAPSA score of ≤13 is considered as an alternate T2T state to MDA.[3]

Statistical analysis

Descriptive analyses for PsA disease characteristics, disease activity, and demographics were calculated. The established crosswalk tables between the HAQ-DI criterion (HAQ-DI≤0.5) and PROMIS-PF corresponding cutoffs (T score ≥41.3)[10] were used to build PROMIS-based MDA definitions. For each patient, we assessed MDA using the usual method with the HAQ-DI≤0.5 criterion (HAQ-DI MDA) and alternate MDA definitions using PROMIS-PF ≥41.3 as a replacement for the HAQ-DI criterion, with all other MDA criteria except PF/HAQ-DI being kept constant (PROMIS-PF4a MDA and PROMIS-PF CAT MDA).

To measure agreement between the original HAQ-DI MDA definition with the PROMIS MDA definitions (MDA state met, MDA state not met), we used the kappa statistic with the following interpretation: ≤0.2 slight; 0.2≤0.4 fair; 0.4≤0.6 moderate; 0.6≤0.8 substantial; >0.8 excellent agreement.[21] We calculated kappa statistic at each visit, globally across all visits, and longitudinally for state changes in MDA between consecutive visits. We used bootstrapping of individual patients with 2000 repetitions to calculate bias corrected

95% confidence intervals for the kappa statistic.[22] Additionally, we conducted sensitivity analyses by estimating kappa in subgroups including gender, age, pain level, patient global assessment, T2T state, and levels of physical function and disability. To further assess the validity of PROMIS-based MDA definitions we calculated agreement of all MDA definitions with the alternative definition of T2T state using the clinical Disease Activity Score in Psoriatic Arthritis (cDAPSA) cutoff of cDAPSA ≤13.[20] Where the number of available observations was less than 50 the kappa statistic was not calculated.

We calculated the sensitivity and specificity of PROMIS-MDA for the HAQ-DI based MDA by assessing the area under the curve from logistic regression, modeled to predict the HAQ-DI MDA using each PROMIS MDA definition by visit and globally across all visits. In an exploratory analysis we build a receiver operator characteristic (ROC) curve to compare different thresholds of the PROMIS scores to identify which cutoff best approximates patients with HAQ-DI less than 0.5. Using each measured value of the PROMIS score as a cutoff, we plotted the true positive rate (sensitivity) against the false positive rate (1 – specificity).[23] In order to define the most favorable cutoff among this PsA cohort, we calculated Youden's Index (sensitivity+specificity-1) for each cutoff of the PROMIS score, and chose the cutoff with the highest value.[24]

To assess longitudinal construct validity of PROMIS-based MDA definitions, we calculated agreement of change in PROMIS MDA with change in HAQ-DI MDA. To accomplish this, each participant was evaluated for longitudinal change in their MDA status at consecutive visits. MDA state change was determined between consecutive visits for each MDA definition (MDA HAQ-DI change, MDA PROMIS-PF4a change, and MDA PROMIS-PF CAT change). Participants were categorized as either "improved" if they transitioned from non-MDA to MDA at consecutive visits; "worsened" if they transitioned from MDA to non-MDA; or "unchanged" if their MDA category remained stable across consecutive visits.

For all analyses specified above, when the number of observations was sufficient, we also assessed the kappa statistic between HAQ-DI VLDA and PROMIS VLDA definitions for static and change states.

Results

Participant characteristics

One hundred patients contributed 352 total observations with up to five visits. Mean age (SD) was 52 (12) at cohort enrollment, 60% were female, 92% were white, 4% African-American, and 4% Asian. Majority of participants were working full time (56%) while 15% were retired and 14% were on disability. Among participants, 93% had at least 2 consecutive visits, 84% had at least 3, 71% had at least 4, and 4% had 5. The average time intervals between consecutive visits were 18.4, 12.1, 13.3, and 17.2 weeks, respectively. Participants completed an average number of PROMIS-PF Bank items (SD) as follows: 4.39 (1.07) at baseline, 4.39 (1.09) at the second visit, 4.37 (1.04) at the third visit, and 4.35 (0.99) at the fourth visit.

At baseline, mean (SD) tender (out of 68) and swollen (out of 66) joint counts were 3.23 (4.87) and 3.08 (3.74). Seven percent had enthesitis, and 3% had active dactylitis. Mean (SD) pain NRS was 3.61 (2.87), patient global psoriatic disease 3.77 (3.18), HAQ-DI 0.71 (0.76), PROMIS-PF 4a T-score 43.03 (9.39), and PROMIS-PF CAT T-score 43.76 (10.29). At baseline 43% met HAQ-DI based MDA/ PROMIS-PF 4a based MDA/PROMIS-PF CAT based MDA, and 53% met cDAPSA (≤13). Majority (56%) were treated with biologicals alone or in DMARD combination and 25% were treated with DMARD alone (Table 1). At baseline, 25% of participants had a HAQ-DI score of zero (floor effect) and at subsequent visits percentages ranged from 23% to 30% with a HAQ-DI score of zero (Table 1). This floor effect did not occur with the PROMIS-PF scores. More so, the PROMIS-PF CAT scores were approximately normally distributed in the PsA population (Supplement Figure 2).

Missing data on all variables used are summarized in Supplement Table 1.

Agreement among MDA/VLDA definitions

Kappa statistic for HAQ-DI and PROMIS- PF based MDA definitions reflected excellent agreement (kappa>0.8) consistently at each visit: kappa ranged between 0.83-0.93 for PROMIS-PF 4a based MDA and between 0.91-0.98 for PROMIS-PF CAT based MDA (Table 2). Kappa values for VLDA were consistent with MDA, and ranged between 0.81-0.88 for PROMIS-PF 4a based VLDA and between 0.76-0.91 for PROMIS-PF CAT based VLDA (Table 2).

Agreement between physical function equivalence thresholds for HAQ-DI and PROMIS-PF reflected substantial agreement (kappa>0.6) at each visit and overall, across visits: kappa=0.73 (95% CI 0.65-0.80) for PROMIS-PF4a and kappa=0.75 (95% CI 0.67-0.81) for PROMIS-PF CAT (Supplement Table 2).

Sensitivity analyses in subgroups

Agreement between MDA HAQ-DI and MDA PROMIS-PF CAT was generally greater than between MDA HAQ-DI and MDA PROMIS-PF4a among subgroups of male and female, age≤51 and ag >51, high and low pain as defined by Pain NRS median, high and low global psoriatic disease as defined by the Patient global psoriatic disease NRS median, and T2T state subgroups using cDAPSA (≤13/>13) (Supplement Table 3).

Agreement between MDA HAQ-DI and MDA PROMIS-PF CAT reflected excellent agreement in gender groups: kappa=0.93 (95% CI 0.86-0.98) for females and kappa=0.95 (95% CI 0.87-1.00) for males. Agreement was slightly higher between MDA HAQ-DI and MDA PROMIS-PF4a in females: kappa=0.91 (95% CI 0.84-0.96) compared to males: kappa=0.86 (95% CI 0.77-0.95). Agreement was higher between MDA HAQ-DI and both PROMIS-MDA states in patients who were younger than the median age of 51 years: kappa= 0.91 (95% CI 0.84-0.97) for MDA PROMIS-PF4a and kappa=0.96 (95% CI 0.91-1.00) for MDA PROMIS-PF CAT, compared to those older than 51 years: kappa=0.89 (95% CI 0.80-0.96) for MDA-PROMIS-PF4a and kappa=0.92 (95% CI 0.84-0.97) for MDA PROMIS-PF CAT. There was higher agreement between MDA HAQ-DI and both PROMIS MDA states in participants with lower pain: kappa=0.82 (95% CI 0.68-0.93) for MDA

PROMIS-PF4a and kappa=0.90 (95% CI 0.79-0.98) for MDA PROMIS-PF CAT compared to higher pain: kappa=0.76 (95% CI 0.55-0.91) for MDA PROMIS-PF4a and kappa=0.84 (95% CI 0.66-0.96) for MDA PROMIS-PF CAT (Supplement Table 3)

Taking physical function as the grouping criterion, agreement between MDA HAQ-DI and PROMIS MDA was higher in those with HAQ-DI scores ≤0.5: kappa=0.86 (95% CI 0.77-0.93) for MDA PROMIS-PF4a and kappa=0.94 (95% CI 0.88-0.99) for MDAPROMIS-PF CAT) compared to those with worse HAQ-DI scores >0.5: kappa=0.76 (95% CI 0.53-0.95) for MDA PROMIS-PF4a and kappa=0.76 (95% CI 0.53-0.95) for MDAPROMIS-PF CAT. Analysis using grouping defined by T scores for physical function (≥41.3 vs <41.3) yielded similar results (Table 3).

Findings for agreement between VLDA definitions were consistent with the findings for MDA except when physical function ability was grouped by PROMIS-PF CAT scores: agreement was slightly higher in the subgroups with PROMIS-PF CAT T score <41.3 versus ≥41.3, although in the excellent range for all subgroups (Table 3).

Validity of PROMIS-MDA using DAPSA T2T states and area under the curve

We calculated agreement between clinical cDAPSA T2T state (cDAPSA≤13) with each of the MDA definitions and found substantial agreement for each. As seen in Table 4, kappa (95%CI) was 0.70 (0.62-0.77), 0.67 (0.59-0.75) and 0.71 (0.63-0.78) for MDA HAQ-DI, MDA PROMIS-PF4a, and MDA PROMIS-PF CAT, respectively. Figure 1 represents agreement among the three MDA definitions with cDAPSA T2T using Venn diagrams and reflects almost overlapping agreement among the MDA definitions with the cDAPSA, while confirming cDAPSA as a more generous T2T classification compared to any of the MDA definitions (Figure 1).

VLDA agreement with cDAPSA remission (cDAPSA≤4) was substantial with kappa (95%CI) values of 0.65 (0.52-0.76), 0.68 (0.59-0.79) and 0.68 (0.56-0.78) for MDA HAQ-DI, MDA PROMIS-PF4a, and MDA PROMIS-PF CAT, respectively (Supplement Table 4).

Areas under ROC curve to predict HAQ-DI based MDA using MDA PROMIS-PF4a or MDA PROMIS-PF CAT across all visits were 0.95 and 0.97, respectively. These calculations were consistent at each visit. (Supplement Figure 3).

The best cutoffs of the PROMIS PF T-scores to represent whether the HAQ-DI is less than 0.5 in this PsA cohort, were based on the Youden's index and are presented in comparison with the performance of the external cutoff, as exploratory analyses. For the PROMIS-PF4a, the best cutoff in the PsA dataset was represented by a T-score of 39.8 which had a sensitivity of 81.6%, a specificity of 95.4%, and a corresponding Youden's index of 0.770. Comparatively, the external cutoff determined by Schalet B.D., et al[10], a T-score of 41.3, had a sensitivity of 85.8%, a specificity of 87.3%, and a corresponding Youden's index of 0.731, for representing a HAQ-DI less than 0.5 in the cohort (Supplement Figure 4). For the PROMIS-PF CAT, the best cutoff was a T-score of 40.2, which had a sensitivity of 77.3%, a specificity of 97%, and a corresponding Youden's index of 0.742. Comparatively the cutoff represented by a T-score of 41.3 had a sensitivity of 79.4%, a specificity of 92.4%, and

a corresponding Youden's index of 0.719, for representing a HAQ-DI less than 0.5 in the cohort (Supplement Figure 5).

Longitudinal validity of PROMIS-based MDA definitions

Kappa (95% CI) between MDA HAQ-DI change and MDA PROMIS-PF4a change was 0.75 (0.47-0.95), 0.84 (0.58-1.00), and 0.72 (0.37-0.94) across consecutive visits 1-2, 2-3 and 3-4. Kappa (95%CI) between MDA HAQ-DI change and MDA-PROMIS-PF CAT change was 0.81 (0.49-1.00), 0.94 (0.75-1.00), and 0.84 (0.48-1.00) across consecutive visits (Table 5). Agreement between transitions was similarly in the substantial to excellent range for VLDA definitions (Table 5).

Discussion

Measuring patient outcomes efficiently and accurately is crucial to evaluating therapies and monitoring disease progression in PsA. Patients who are receiving newer biologic agents are now functioning above average. Early detection of deterioration in clinical status in those with high levels of physical function is essential to providing optimal clinical care. Because the HAQ-DI focuses on assessing degree of disability, it performs well in disabled populations but not as well in those with average or above average physical function.[25] Thus, those classified as having no disability and at T2T as defined by HAQ-DI MDA criteria may have an overestimated measure of physical functioning. Similarly, deterioration or improvement within the range of no disability may not captured. Compared to the HAQ-DI, the PROMIS-PF item Bank is expanded to include items assessing higher levels of physical functioning (i.e. strenuous and vigorous exercise such as running and weight lifting). Thus, PROMIS scales are designed to focus on ability and are more sensitive than the legacy HAQ-DI in detecting clinical improvement or deterioration on newer therapies.[17,26,27] PROMIS instruments are also designed to be less taxing on patients and offer higher precision in assessing physical function than the legacy HAQ-DI with fewer questions,[17,26] especially with CAT administration.

Our study is the first to compare agreement of PROMIS and HAQ-DI MDA definitions in a PsA cohort, based on the equivalency of a HAQ-DI score of 0.5 to a PROMIS-PF T score of 41.3.[10] Our findings suggest that PROMIS is an accurate replacement for HAQ-DI given substantial to excellent agreement between PROMIS and HAQ-DI based MDA definitions, in cross-sectional as well as longitudinal analyses.

In our cohort, agreement was always higher between HAQ-DI and PROMIS-PF CAT based MDA states than between HAQ-DI and PROMIS-PF4a based MDA states which may be explained by higher precision and reduced ceiling and floor effects with the use of CAT. On average, the kappa statistic was 0.90, indicating near perfect agreement between the HAQ-DI based MDA and both PROMIS-based MDA states in this guideline-based treated PsA population. Furthermore, area under the curve for predicting the HAQ-DI based MDA using either PROMIS definition was consistently greater than 0.90 at each visit supporting excellent accuracy for both PROMIS-based MDA definitions. Agreement of MDA definitions with DAPSA T2T state was substantial, and in the ranges observed in other studies[28].

Substantial agreement (kappa > 0.6) was observed between HAQ-DI and PROMIS scores at the equivalence cutoff used, compared to excellent agreement (kappa > 0.8) between HAQ-DI and PROMIS based MDA states. At baseline, agreement between HAQ-DI and each PROMIS-PF based MDA definition remained high despite lower agreement between HAQ-DI and each PROMIS score. Since the physical function score was the only changing variable in calculation of HAQ-DI versus PROMIS-PF based MDA state, perfect agreement of raw physical function scores may not be necessary for high MDA state agreement. However, at lower states of physical functioning, MDA agreement dropped from excellent (at high PF level) to moderate-substantial, reflecting that differences in the physical function score were more likely to change the MDA status in this group, likely because other MDA criteria were not met.

Excellent agreement between HAQ-DI and PROMIS based MDA states was maintained within subgroups of gender, age, and pain levels. Agreement was overall higher between MDA HAQ-DI and MDA PROMIS-PF CAT than MDA PROMIS-PF4a; it was also generally higher in men versus women, in participants younger versus lower than age 51 years, lower versus higher pain/patient global assessment, those at treat to target state versus not, lower HAQ-DI scores, and higher PROMIS scores. In sum, agreement between HAQ-DI and PROMIS-PF4a based MDA was highest in those doing well on multiple MDA criteria. The observed differences in agreement between the MDA HAQ-DI with PROMIS-PF4a definitions among people doing well versus people doing not so well, attenuated significantly with the use of PROMIS-PF CAT based MDA. These findings were also observed for VLDA.

We provided an additional anchor for T2T state, cDAPSA low disease activity, for greater generalizability of our results to other cohorts. We confirmed cDAPSA T2T was easier to achieve than MDA. For all MDA definitions, whether PROMIS or HAQ-DI were used as measures for physical function, agreement with cDAPSA T2T was substantial and similar between definitions.

Our study also provides comparative performance results for the external standard cutoff, a T-score of 41.3[10], and the best cutoff determined through an exploratory analysis in the PsA cohort. The best T-score cutoff was numerically very close to the external cutoff for both PROMIS instruments. The external cutoff had slightly higher sensitivity (an increase by 4% for PROMIS-PF4a, and 2% for PROMIS-PF CAT) that came with a trade-off in specificity of 8% for PROMIS-PF4a, and 5% for PROMIS-PF CAT in this PsA dataset. However, cutoffs used in the dataset in which they were derived would bias toward higher agreement than using other standard measures. The analysis was conservative in using the external cutoff for agreement and supports the validity of this external cutoff in the PsA population.

Finally, there was substantial to excellent longitudinal agreement between HAQ-DI and PROMIS-PF based MDA states over time. However, as we may expect, PROMIS-PF CAT was more sensitive to MDA change as exhibited by higher kappa values compared to PROMIS-PF4a. These findings remained consistent when we examined VLDA definitions.

Our study findings may encourage clinicians who administer HAQ-DI based measures to switch to PROMIS-PF. Additionally, institutions that collect PROMIS-PF need not re-collect HAQ-DI for the purposes of calculating MDA and T2T determination. Given interchangeability of PROMIS and HAQ-DI in determining MDA, the use of PROMIS-PF offers the advantage of capturing a broader range of physical functioning more efficiently. PROMIS-PF CAT measures disability just as well as the HAQ-DI, as the item bank still contains questions focused on limited functioning (i.e. opening jars) without requiring completion of an extensive questionnaire. Further, PROMIS-PF CAT is able to measure maximum functional capacity for each patient regardless of whether they meet MDA, as higher scores correlate with greater performance status.

The characteristics of our guideline-based treatment cohort may limit generalizability to other cohorts, as most patient were Caucasian, slightly more than half were treated with biologic DMARDs, and about 50% were at treatment targets. As discussed by Schalet et al., validity of the crosswalk table may be sensitive to population differences and weaker at extreme ends of the physical function continuum.[10] Consistent with this observation, we observed a drop from excellent agreement when physical function was good, to moderate and substantial agreement in participants with low physical function ability. Finally, a limitation to crosswalk tables is that they are based on summed raw scores and can only be used when there are no missing values. However, the HAQ-DI has similar limitations and cannot be computed unless at least one item in a category score has been completed. Strengths of our study are collection of HAQ-DI, PROMIS-PF4a and PROMIS-PF CAT concomitantly at each study visit in addition to comprehensive PsA-specific phenotype and disease activity data. We performed analyses by visit which showed stability of our findings longitudinally and by subgroups of interest (gender, T2T, and physical function ability). Results were consistent when we triangulated methods of agreement (kappa) with prediction (ROC analysis).

In conclusion, we demonstrated interchangeability of the HAQ-DI threshold of ≤0.5 with a PROMIS-PF threshold of ≥41.3 in the calculation of PsA MDA and VLDA, which provides supportive data towards the validity of this cross-walk between HAQ-DI and PROMIS-PF scores in the PsA population. Results from our study demonstrate agreement between legacy HAQ-DI and PROMIS based MDA definitions statically, longitudinally, and within demographic, disease activity, functioning, and symptom subgroups. Thus, PROMIS-PF can replace HAQ-DI in calculating MDA state in PsA, and cohorts switching from HAQ-DI to PROMIS-PF can convert scores longitudinally on the physical function scale of their choice.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Funding:

The project was supported by a Rheumatology Research Foundation Scientist Development award (PI: AM Orbai). AMO is a Jerome L. Greene Foundation Scholar. AMO and JP are supported in part by a research grant from the

National Institute of Arthritis and Musculoskeletal and Skin Diseases (NIAMS) of the National Institutes of Health (NIH) under award number P30-AR070254 (Core B).

References

1. Ritchlin CT, Colbert RA, Gladman DD. Psoriatic Arthritis. N Engl J Med. 2017;376(10):957–970. doi:10.1056/NEJMra1505557 [PubMed: 28273019]
2. Orbai A- M, de Wit M, Mease P, et al. International patient and physician consensus on a psoriatic arthritis core outcome set for clinical trials. Ann Rheum Dis. 2017;76(4):673–680. doi:10.1136/annrheumdis-2016-210242 [PubMed: 27613807]
3. Smolen JS, Schöls M, Braun J, et al. Treating axial spondyloarthritis and peripheral spondyloarthritis, especially psoriatic arthritis, to target: 2017 update of recommendations by an international task force. Ann Rheum Dis. 2018;77(1):3–17. doi:10.1136/annrheumdis-2017-211734 [PubMed: 28684559]
4. Singh JA, Guyatt G, Ogdie A, et al. Special Article: 2018 American College of Rheumatology/National Psoriasis Foundation Guideline for the Treatment of Psoriatic Arthritis. Arthritis Care Res. 2019;71(1):2–29. doi:10.1002/acr.23789
5. Coates LC, Fransen J, Helliwell PS. Defining minimal disease activity in psoriatic arthritis: a proposed objective target for treatment. Ann Rheum Dis. 2010;69(01):48–53. doi:10.1136/ard.2008.102053 [PubMed: 19147615]
6. Bruce B, Fries JF. The Stanford Health Assessment Questionnaire: dimensions and practical applications. Health Qual Life Outcomes. 2003;1:20. doi:10.1186/1477-7525-1-20 [PubMed: 12831398]
7. 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. doi:10.1016/j.rdc.2016.01.007 [PubMed: 27133496]
8. Home. http://www.healthmeasures.net/. Accessed July 5, 2020.
9. Kaat AJ, Schalet BD, Rutsohn J, Jensen RE, Cella D. Physical function metric over measure: An illustration with the Patient-Reported Outcomes Measurement Information System (PROMIS) and the Functional Assessment of Cancer Therapy (FACT). Cancer. 2018;124(1):153–160. doi:10.1002/cncr.30981 [PubMed: 28885707]
10. Schalet BD, Revicki DA, Cook KF, Krishnan E, Fries JF, Cella D. Establishing a Common Metric for Physical Function: Linking the HAQ-DI and SF-36 PF Subscale to PROMIS® Physical Function. J Gen Intern Med. 2015;30(10):1517–1523. doi:10.1007/s11606-015-3360-0 [PubMed: 25990189]
11. Bukhari M, Kent A. How rheumatologists assess disability in the current era needs an overhaul: focus on the Health Assessment Questionnaire. Rheumatology. doi:10.1093/rheumatology/kez423
12. Bruce B, Fries JF, Ambrosini D, et al. Better assessment of physical function: item improvement is neglected but essential. Arthritis Res Ther. 2009;11(6):R191. doi:10.1186/ar2890 [PubMed: 20015354]
13. Taylor W, Gladman D, Helliwell P, et al. Classification criteria for psoriatic arthritis: development of new criteria from a large international study. Arthritis Rheum. 2006;54(8):2665–2673. doi:10.1002/art.21972 [PubMed: 16871531]
14. Duarte-García A, Leung YY, Coates LC, et al. Endorsement of the 66/68 Joint Count for the Measurement of Musculoskeletal Disease Activity: OMERACT 2018 Psoriatic Arthritis Workshop Report. J Rheumatol. 2019;46(8):996–1005. doi:10.3899/jrheum.181089 [PubMed: 30770518]
15. Healy PJ, Helliwell PS. Measuring clinical enthesitis in psoriatic arthritis: assessment of existing measures and development of an instrument specific to psoriatic arthritis. Arthritis Rheum. 2008;59(5):686–691. doi:10.1002/art.23568 [PubMed: 18438903]
16. Mease PJ, Woolley JM, Bitman B, Wang BC, Globe DR, Singh A. Minimally Important Difference of Health Assessment Questionnaire in Psoriatic Arthritis: Relating Thresholds of Improvement in Functional Ability to Patient-rated Importance and Satisfaction. J Rheumatol. 2011;38(11):2461–2465. doi:10.3899/jrheum.110546 [PubMed: 21885498]

17. Fries JF, Krishnan E, Rose M, Lingala B, Bruce B. Improved responsiveness and reduced sample size requirements of PROMIS physical function scales with item response theory. Arthritis Res Ther. 2011;13(5):R147. doi:10.1186/ar3461 [PubMed: 21914216]
18. Fries JF, Bruce B, Cella D. The promise of PROMIS: using item response theory to improve assessment of patient-reported outcomes. Clin Exp Rheumatol. 2005;23(5 Suppl 39):S53–57
19. Hanmer J, Jensen RE, Rothrock N. A reporting checklist for Health Measures' patient-reported outcomes: ASCQ-Me, Neuro-QoL, NIH Toolbox, and PROMIS. J Patient Rep Outcomes.2020;4:21. 10.1186/s41687-020-0176-4 [PubMed: 32215788]
20. Schoels MM, Aletaha D, Alasti F, Smolen JS. Disease activity in psoriatic arthritis (PsA): defining remission and treatment success using the DAPSA score. Ann Rheum Dis. 2016;75(5):811–818. doi:10.1136/annrheumdis-2015-207507 [PubMed: 26269398]
21. McHugh ML. Interrater reliability: the kappa statistic. Biochem Medica. 2012;22(3):276–282. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3900052/. Accessed November 2, 2019
22. Reichenheim ME. Confidence intervals for the kappa statistic. Stata J. 2004;4:421–8.
23. Fawcett T An introduction to ROC analysis. Pattern Recognition Letters. 2006; 27:861–874. doi:10.1016/j.patrec.2005.10.010
24. Schisterman EF, Perkins NJ, Liu A, Bondell H. Optimal cut-point and its corresponding Youden Index to discriminate individuals using pooled blood samples. Epidemiology. 2005;1:73–81.
25. Mease P, Strand V, Gladman D. Functional impairment measurement in psoriatic arthritis: Importance and challenges. Semin Arthritis Rheum. 2018;48(3):436–448. doi:10.1016/j.semarthrit.2018.05.010 [PubMed: 30029795]
26. Rose M, Bjorner JB, Gandek B, Bruce B, Fries JF, Ware JE. The PROMIS Physical Function Item Bank Was Calibrated to a Standardized Metric and Shown to Improve Measurement Efficiency. J Clin Epidemiol. 2014;67(5):516–526. doi:10.1016/j.jclinepi.2013.10.024 [PubMed: 24698295]
27. Fries JF, Cella D, Rose M, Krishnan E, Bruce B. Progress in Assessing Physical Function in Arthritis: PROMIS Short Forms and Computerized Adaptive Testing. J Rheumatol. 2009;36(9):2061–2066. doi:10.3899/jrheum.090358 [PubMed: 19738214]
28. Gorlier C, Orbai A- M, Puyraimond-Zemmour D, et al. Comparing patient-perceived and physician-perceived remission and low disease activity in psoriatic arthritis: an analysis of 410 patients from 14 countries. Ann Rheum Dis. 2019;78(2):201–208. doi:10.1136/annrheumdis-2018-214140 [PubMed: 30442648]

Significance and Innovations

- This study demonstrates excellent cross-sectional and longitudinal agreement between HAQ-DI and PROMIS-based minimal disease activity (MDA) definitions in PsA.
- PROMIS Physical Function short form 4a or computer adaptive test can adequately replace the HAQ-DI for the purpose of MDA and T2T determination in PsA.
- PROMIS Physical Function has the advantage of assessing physical function on an extended spectrum of ability and can concomitantly be used for MDA calculation in PsA.

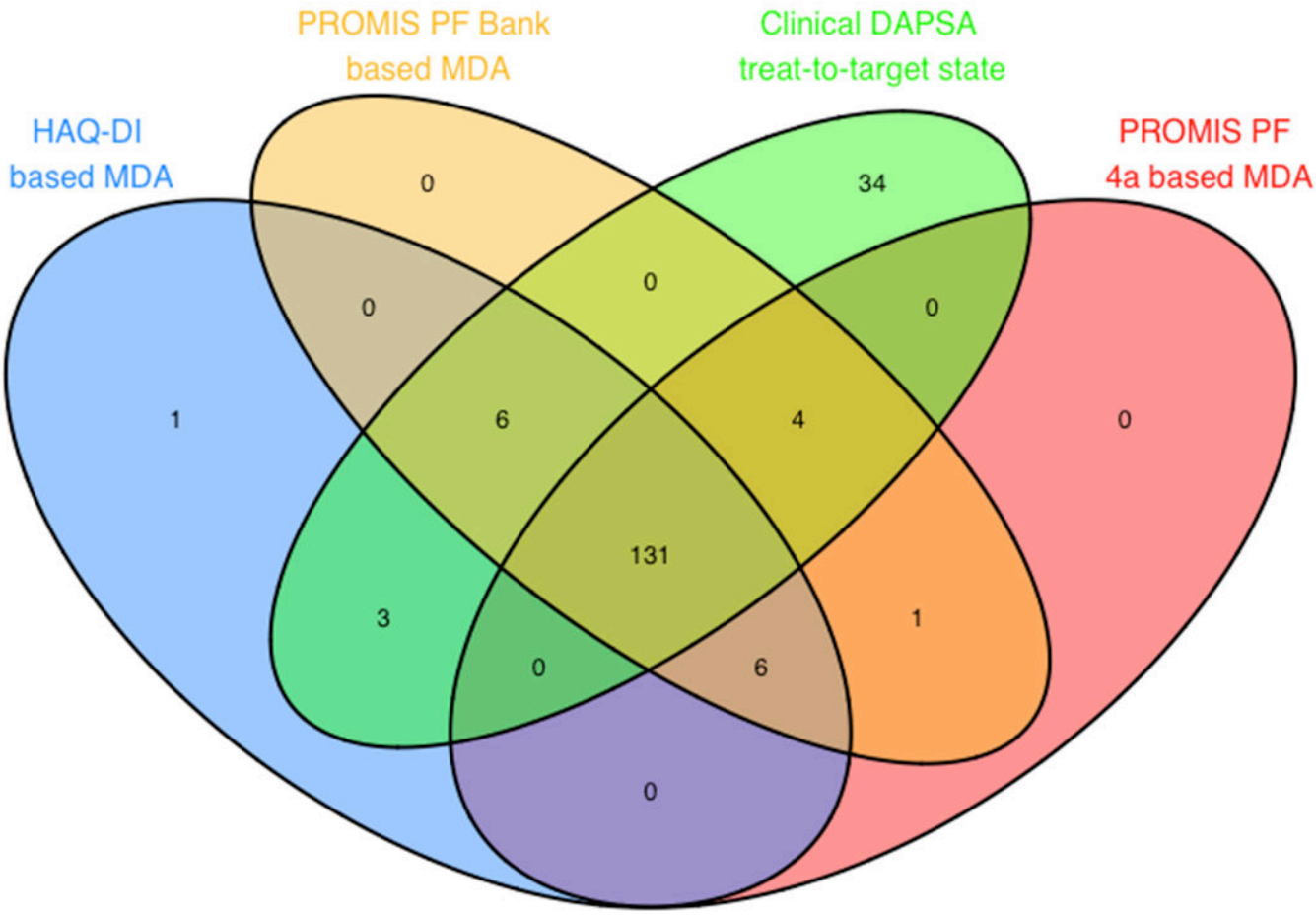

Figure 1 Legend.

Classification overlap between the four proposed definitions of T2T: MDA HAQ-DI, MDA PROMIS-PF 4a, MDA PROMIS-PF CAT, and cDAPSA T2T. Numbers represent available observations across all visits where participants were classified as having met one of the treatment targets. Diagram is not represented to scale.

Table 1.

Summary statistics of psoriatic arthritis cohort at each visit

Measures	Visit 1 N=100	Visit 2 N=93	Visit 3 N=84	Visit 4 N=71
Tender joint count [0-68], mean (SD)	3.23 (4.87)	3.97 (5.92)	3.94 (4.95)	5.66 (7.98)
Swollen joint count [0-66], mean (SD)	3.08 (3.74)	3.09 (3.78)	3.65 (3.69)	4.14 (5.16)
Enthesitis, n (%)	7 (7)	7 (7.53)	8 (9.52)	12 (16.90)
Dactylitis, n (%)	3 (3)	5 (5.38)	4 (4.76)	2 (2.82)
Psoriasis BSA [0-100], mean (SD)	2.23 (4.73)	3.07 (9.35)	4.77 (12.03)	3.03 (6.25)
Percent BSA ≥10, n (%)	5 (5)	5 (5.38)	10 (11.90)	6 (8.45)
Pain NRS [0-10], mean (SD)	3.61 (2.87)	3.45 (2.89)	3.73 (3.13)	3.48 (2.93)
Patient global psoriatic disease NRS [0-10], mean (SD)	3.77 (3.18)	3.29 (2.99)	3.34 (2.99)	3.53 (2.98)
Patient global psoriatic arthritis NRS [0-10], mean (SD)	3.79 (3.02)	3.33 (2.87)	3.31 (2.95)	3.57 (2.98)
PROMIS-PF4a [0-100], T-score mean (SD)	43.03 (9.39)	44.15 (9.74)	43.70 (9.86)	45.15 (9.71)
PROMIS-PF CAT [0-100], T-score mean (SD)	43.76 (10.29)	45.02 (9.97)	44.49 (10.23)	44.85 (9.46)
HAQ-DI [0-3], mean (SD)	0.71 (0.76)	0.61 (0.70)	0.68 (0.73)	0.64 (0.73)
HAQ-DI=0, n (%)	25 (25)	30 (32.26)	25 (29.76)	23 (32.39)
MDA HAQ-DI, n (%)	43 (43)	47 (57.32)	33 (45.21)	28 (48.28)
MDA PROMIS-PF4a, n (%)	43 (43)	43 (51.81)	31 (42.47)	28 (46.67)
MDA PROMIS-PF CAT, n (%)	43 (43)	46 (54.76)	33 (44.59)	31 (51.67)
VLDA HAQ-DI, n (%)	9 (9)	11 (11.83)	10 (11.90)	9 (12.68)
VLDA PROMIS-PF4a, n (%)	9 (9)	11 (11.83)	10 (11.90)	9 (12.68)
VLDA PROMIS-PF CAT, n (%)	10 (10)	12 (12.90)	10 (11.90)	9 (12.68)
cDAPSA T2T state, n (%)	53 (53)	54 (63.53)	45 (55.56)	34 (54.29)
Biologicals alone or in DMARD combination, n (%)	56 (56)	58 (62.37)	58 (69.05)	52 (73.34)
DMARD alone, n (%)	25 (25)	25 (26.88)	17 (20.24)	15 (21.13)

Abbreviations

SD: Standard Deviation; NRS: Numeric Rating Scale; HAQ-DI: Heath Assessment Questionnaire-Disability Index; PROMIS-PF 4a: Patient Reported Outcomes Measurement Information System Short Form v2.0 Physical Function 4a; PROMIS-PF CAT: PROMIS Bank v1.2 – Physical Function Computer Adaptive Test; BSA: body surface area affected by psoriasis; MDA HAQ-DI: Minimal disease activity that includes the HAQ-DI ≤0.5 criterion; MDA PROMIS-PF4a includes the PROMIS-PF4a T-score ≥41.3 criterion; MDA PROMIS-PF CAT includes the PROMIS-PF CAT T-score ≥41.3 criterion; VLDA Very Low Disease Activity, VLDA variables similarly defined to MDA variables; cDAPSA: clinical disease activity psoriatic arthritis; T2T: treat-to-target; DMARD: disease modifying anti-rheumatic drug.

Table 2.

Agreement between HAQ-DI based MDA/VLDA and PROMIS-PF based MDA/VLDA definitions at each visit

Agreement*		Visit 1	Visit 2	Visit 3	Visit 4
MDA HAQ-DI and MDA PROMIS-PF4a	Kappa 95% CI N	0.91 (0.80-0.98) 86	0.93 (0.82-1.00) 81	0.92 (0.80-1.00) 72	0.83 (0.66-0.96) 58
MDA HAQ-DI and MDA PROMIS-CAT	Kappa 95% CI N	0.91 (0.81-0.98) 86	0.98 (0.90-1.00) 82	0.94 (0.84-1.00) 73	0.93 (0.82-1.00) 58
VLDA HAQ-DI and MDA PROMIS-PF4a	Kappa 95% CI N	0.82 (0.72-0.91) 80	0.88 (0.78-0.95) 69	0.85 (0.75-0.93) 68	0.81 (0.68-0.92) 55
VLDA HAQ-DI and VLDA PROMIS-PF CAT	Kappa 95% CI N	0.76 (0.65-0.86) 80	0.88 (0.79-0.95) 70	0.91 (0.83-0.98) 68	0.87 (0.77-0.96) 55

* Bias corrected 95% CI were calculated using bootstrapping with 2000 repetitions of individual patients.

Abbreviations

CI: confidence interval; N: number of observations; HAQ-DI: Heath Assessment Questionnaire-Disability Index; PROMIS-PF 4a: Patient Reported Outcomes Measurement Information System Short Form v2.0 Physical Function 4a; PROMIS-PF CAT: PROMIS Bank v1.2 - Physical Function Computer Adaptive Test; MDA HAQ-DI: Minimal disease activity that includes the HAQ-DI ≤0.5 criterion; MDA PROMIS-PF4a includes the PROMIS-PF4a T score ≥41.3 criterion; MDA PROMIS-PF CAT includes the PROMIS-PF CAT T-score ≥41.3 criterion; VLDA HAQ-DI Very low disease activity that includes the HAQ-DI ≤0.5 criterion; VLDA PROMIS-PF4a includes the PROMIS-PF4a T-score ≥41.3 criterion; VLDA PROMIS-PF CAT includes the PROMIS-PF CAT T-score ≥41.3 criterion.

Table 3.

Agreement between HAQ-DI based MDA and PROMIS-PF based MDA in subgroups defined by Physical Function ability across all visits

Agreement in subgroups*	HAQ-DI MDA & PROMIS-PF4a MDA	HAQ-DI MDA & PROMIS-PF CAT MDA	HAQ-DI VLDA & PROMIS-PF4a VLDA	HAQ-DI VLDA & PROMIS-PF CAT VLDA
HAQ-DI≤0.5, less/no disability				
Kappa	0.86	0.94	0.84	0.93
95% CI	(0.77-0.93)	(0.88-0.99)	(0.78-0.90)	(0.88-0.97)
N	184	186	159	160
HAQ-DI > 0.5, more disability				
Kappa	0.76	0.76	0.80	0.70
95% CI	(0.53-0.95)	(0.53-0.95)	(0.71-0.88)	(0.60-0.79)
N	116	116	116	116
PROMIS-PF 4a T-score ≥41.3, better physical function ability				
Kappa	0.93	0.93	0.86	0.86
95% CI	(0.85-0.98)	(0.85-0.98)	(0.79-0.91)	(0.79-0.91)
N	182	182	157	157
PROMIS-PF 4a T-score < 41.3, worse physical function ability				
Kappa	0.55	0.86	0.78	0.81
95% CI	(0.31-0.78)	(0.70-0.97)	(0.70-0.86)	(0.72-0.88)
N	118	120	118	118
PROMIS-PF CAT T-score ≥41.3, better physical function ability				
Kappa	0.87	0.94	0.80	0.81
95% CI	(0.79-0.94)	(0.88-0.99)	(0.73-0.87)	(0.74-0.87)
N	203	205	177	178
PROMIS-PF CAT T-score < 41.3, less physical function ability				
Kappa	0.65	0.65	0.88	0.88
95% CI	(0.17-0.92)	(0.17-0.92)	(0.80-0.95)	(0.79-0.94)
N	98	98	98	98

*Bias corrected 95% confidence intervals for the kappa statistic were calculated using bootstrapping with 2000 repetitions of individual patients.

Abbreviations

CI: Confidence interval; HAQ-DI: Heath Assessment Questionnaire-Disability Index; PROMIS-PF 4a: Patient Reported Outcomes Measurement Information System Short Form v2.0 Physical Function 4a; PROMIS-PF CAT: PROMIS Bank v1.2 - Physical Function Computer Adaptive Test; MDA: HAQ-DI Minimal disease activity that includes the HAQ-DI ≤0.5 criterion; MDA PROMIS-PF4a includes the PROMIS-PF4a T-score ≥41.3 criterion; MDA PROMIS-PF Bank includes the PROMIS-PF CAT T-score ≥41.3 criterion; VLDA HAQ-DI: Very low disease activity that includes the HAQ-DI ≤0.5 criterion; VLDA PROMIS-PF4a includes the PROMIS-PF4a T-score ≥41.3 criterion; VLDA PROMIS-PF CAT includes the PROMIS-PF CAT T-score ≥41.3 criterion.

Table 4.

Agreement between MDA definitions and Clinical Disease Activity in Psoriatic Arthritis (cDAPSA) T2T state (cDAPSA≤13) across all visits

Agreement*	MDA PROMIS-PF 4a	MDA PROMIS-PF CAT	cDAPSA T2T (≤13)
MDA HAQ-DI			
Kappa	0.90	0.94	0.70
(95% CI)	(0.84-0.95)	(0.90-0.97)	(0.62-0.77)
N	301	303	299
MDA PROMIS-PF4a			
Kappa	-	0.96	0.67
(95% CI)		(0.93-0.99)	(0.59-0.75)
N		307	303
MDA PROMIS-PF CAT			
Kappa	-	-	0.71
(95% CI)			(0.63-0.78)
N			305

*Bias corrected 95% confidence intervals for the kappa statistic were calculated using bootstrapping with 2000 repetitions of individual patients.

Abbreviations

CI confidence interval; HAQ-DI: Heath Assessment Questionnaire-Disability Index; PROMIS-PF 4a: Patient Reported Outcomes Measurement Information System Short Form v2.0 Physical Function 4a; PROMIS-PF CAT: PROMIS Bank v1.2 - Physical Function Computer Adaptive Test; MDA HAQ-DI: Minimal disease activity that includes the HAQ-DI ≤0.5 criterion; MDA PROMIS-PF4a includes the PROMIS-PF4a T-score ≥41.3 criterion; MDA PROMIS-PF CAT includes the PROMIS-PF CAT T-score ≥41.3 criterion; cDAPSA: Clinical disease activity in psoriatic arthritis; T2T: treat-to-target.

Table 5.

Longitudinal agreement between HAQ-DI based MDA and corresponding PROMIS-PF based MDA state changes

Agreement*		V1 - V2	V2 - V3	V3 - V4
MDA HAQ-DI state change with MDA PROMIS-PF4a state change	Kappa 95% CI N	0.75 (0.47-0.95) 71	0.84 (0.58-1.00) 67	0.72 (0.37-0.94) 51
MDA HAQ-DI state change with MDA PROMIS-PF CAT state change	Kappa 95% CI N	0.81 (0.49-1.00) 72	0.94 (0.75-1.00) 68	0.84 (0.48-1.00) 52
VLDA HAQ-DI state change with VLDA PROMIS-PF4a state change	Kappa 95% CI N	0.75 (0.44-0.95) 59	0.84 (0.51-1.00) 57	- - 47
VLDA HAQ-DI state change with VLDA PROMIS-PF CAT state change	Kappa 95% CI N	0.82 (0.55-1.00) 60	0.92 (0.64-1.00) 58	- - 47

*MDA (or VLDA) state changes were defined as transitions in corresponding MDA (or VLDA) state between consecutive visits, for example V1-V2 represents agreement between transitions in MDA (or VLDA) HAQ-DI state from visit 1 to visit 2 with transitions in each MDA (or VLDA) PROMIS-PF state from visit 1 to visit 2. Bias corrected 95% confidence intervals for the kappa statistic were calculated using bootstrapping with 2000 repetitions of individual patients. When sample size was less than 50, kappa and 95% CI were not calculated, designated with "-".

Abbreviations

CI: confidence interval; N: available observations; HAQ-DI: Heath Assessment Questionnaire-Disability Index; PROMIS-PF 4a: Patient Reported Outcomes Measurement Information System Short Form v2.0 Physical Function 4a; PROMIS-PF CAT: PROMIS Bank v1.2 – Physical Function Computer Adaptive Test; MDA HAQ-DI: Minimal disease activity that includes the HAQ-DI ≤0.5 criterion; MDA PROMIS-PF4a includes the PROMIS-PF4a T-score ≥41.3 criterion; MDA PROMIS-PF CAT includes the PROMIS-PF CAT T-score ≥41.3 criterion; VLDA HAQ-DI: Very low disease activity includes the HAQ-DI ≤0.5 criterion; VLDA PROMIS-PF4a includes the PROMIS-PF4a T-score ≥41.3 criterion; VLDA PROMIS-PF CAT includes the PROMIS-PF CAT T-score ≥41.3 criterion.