# Content validity and psychometric evaluation of Functional Assessment of Chronic Illness Therapy-Fatigue in patients with psoriatic arthritis

David Cella[1*], Hilary Wilson[2], Huda Shalhoub[3], Dennis A. Revicki[2], Joseph C. Cappelleri[4], Andrew G. Bushmakin[4], Elizabeth Kudlacz[4] and Ming-Ann Hsu[4]

## Abstract

**Background:** To evaluate the measurement properties (e.g. content validity, reliability and ability to detect change) of the Functional Assessment of Chronic Illness Therapy (FACIT)-Fatigue scale in patients with active psoriatic arthritis (PsA).

**Methods:** One-on-one semi-structured qualitative interviews with adult patients with active PsA evaluated the content validity of FACIT-Fatigue. Quantitative measurement properties were evaluated using data from phase III tofacitinib randomized controlled trials (RCTs) in PsA: OPAL Broaden (NCT01877668) and OPAL Beyond (NCT01882439).

**Results:** Of 12 patients included in the qualitative study, 2 (17%) had mild, 8 (67%) had moderate, and 2 (17%) had severe PsA disease activity; 7 (58%) attributed fatigue to PsA, and 7 (58%) rated fatigue as important or extremely important. Most patients considered the FACIT-Fatigue items relevant to their PsA experience and understood item content and response options as intended. In the psychometric analysis of RCT data, a second-order confirmatory factor model fit the data well (Bentler's Comparative Fit Index ≥0.92). FACIT-Fatigue demonstrated good internal consistency (Cronbach's coefficient $a \geq 0.90$), test-retest reliability (Intraclass Correlation Coefficient ≥ 0.80) and a strong correlation with SF-36 Vitality ($r > 0.80$). A robust relationship between disease activity (based on Patient's Global Assessment of Psoriasis and Arthritis) and FACIT-Fatigue was observed (effect sizes > 1.4), with clinically important difference for the FACIT-Fatigue total score estimated as 3.1 points, and the responder definition estimated as a 4-point improvement for FACIT-Fatigue total score.

**Conclusion:** Fatigue was confirmed to be an important symptom to patients with PsA, and FACIT-Fatigue was found to be a reliable and valid measure in this population.

**Keywords:** FACIT-fatigue, Psoriatic arthritis, Content validity, Tofacitinib, Psychometric properties

## Background

Psoriatic arthritis (PsA) is a chronic inflammatory disease occurring in 6–42% of patients with psoriasis [1]. It is characterized by joint inflammation, enthesitis, dactylitis, and spondylitis, and is often associated with generalized fatigue [1–3].

Fatigue was recently added to the core domain set for PsA randomized controlled trials (RCTs) [4, 5], due to the impact that it has on a patient's quality of life. Patients with PsA have noted statistically significant improvements in fatigue following treatment with newer agents such as certolizumab, secukinumab, and apremilast [6–8], suggesting it is modifiable with treatment. For example, intravenous secukinumab 150 mg led to a least squares mean change from baseline in fatigue of 6.74 ($P < 0.05$ vs. placebo) [8].

Although recognized as a core domain for assessment in RCTs, there is currently no universally accepted measure of fatigue recommended to evaluate this construct in patients with PsA. When measuring a construct

\* Correspondence: d-cella@northwestern.edu
[1]Department of Medical Social Sciences, Northwestern University, 633 N. St. Clair Suite 1900, Chicago, IL 60611, USA
Full list of author information is available at the end of the article

within the RCT setting, it is important to ensure the relevance and comprehension of a questionnaire to the target population, and its reliability, validity, and ability to detect change [9–11].

The Functional Assessment of Chronic Illness Therapy–Fatigue (FACIT-Fatigue) scale [12] (Additional file 1) is a 13-item questionnaire originally designed to assess fatigue/tiredness and its impact on daily functioning in people with cancer; it has now been evaluated in other chronic diseases [12–15]. Each item's response option uses a 5-point scale ranging from "not at all" to "very much." The total FACIT-Fatigue score ranges from 0 to 52, where higher scores represent less fatigue [13, 14]. While commonly applied as a unidimensional measure, previous work has shown that FACIT-Fatigue can also be considered as a multidimensional measure, with the impact and experience of fatigue considered separately [16].

Psychometric data in patients with RA suggest that FACIT-Fatigue (total score; baseline, Week 12, and Week 24 assessments) has good internal consistency ($\alpha = 0.86$ to 0.87) and the ability to differentiate patients according to clinical change using the American College of Rheumatology response criteria, [12]. FACIT-Fatigue also showed a strong association with the longer, 16-item Multidimensional Assessment of Fatigue scale ($r = -0.84$ to $-0.88$), implying a redundancy between these two measures. However, this study also reported that FACIT-Fatigue captured a broader distribution of patients and wider range of self-reported fatigue concepts. A qualitative study of 17 patients with moderate to highly active RA found FACIT-Fatigue to have high content validity; 10 of the 13 items had "high" content validity (determined by the relationship between the intended measurement concept and the methods used [17]), with three having "low to moderate" ("I feel weak all over", "I feel listless [washed out]") or "low" ("I am too tired to eat") content validity [18]. This study also concluded that FACIT-Fatigue captured most fatigue-related, patient-reported concepts. Chandran and colleagues also showed FACIT-Fatigue to have good internal consistency ($\alpha = 0.96$) and significant correlation with actively inflamed joint count ($r = -0.43$) in patients with PsA [14]. However, there is currently no qualitative evidence to support content validity in patients with PsA, and no quantitative evidence supporting other measurement properties specifically in an RCT.

We designed a mixed-methods approach to further evaluate the qualitative and quantitative measurement properties of FACIT-Fatigue in patients with PsA. For the former, a qualitative study was designed to: 1) elicit concepts important to patients with PsA regarding the signs, symptoms, and impact of PsA on daily functioning, focusing on the experience and impact of fatigue; and 2) evaluate the content validity of the FACIT-Fatigue scale. For the latter, a secondary analysis of two phase III RCTs of

tofacitinib assessed FACIT-Fatigue in patients with moderate to severe PsA.

## Patients and methods

### Qualitative FACIT-fatigue study

The qualitative assessment was carried out prior to the quantitative analysis and included one-on-one semi-structured interviews with 12 adult patients (aged ≥18 years) who had a confirmed diagnosis and presence of active PsA (full details in Additional file 2a). Interviews were conducted in-person at two clinical sites in the United States (Florida and Pennsylvania), by two experts (research associates, Evidera) who were trained and experienced in qualitative interviewing methods.

### Patient interview

Prior to the start of each interview, the interviewers fully explained the study to the patient and obtained written, informed consent. Interviewers led the discussion using a standardized, semi-structured interview guide (full guide in Additional file 2b), divided into two parts. Part 1, an open-ended concept elicitation, was designed to assess relevant symptom and impact concepts (e.g. self-reported PsA severity), and understand the relative importance and patients' experience of fatigue. Detailed questions related to fatigue were followed by general questions about patients' overall symptoms and impact on functioning.

In part 2, patients completed the Functional Assessment of Chronic Illness Therapy–Fatigue (FACIT-Fatigue) questionnaire and were asked to provide feedback on overall comprehension, relevance, and content validity. Questions were designed to assess the interpretation of items, thoughts about the relevant recall period, and feedback regarding the content in relation to their most important symptoms and overall symptoms and impacts. Following the interview, patients completed a sociodemographic and clinical questionnaire. Qualitative data were then analyzed using ATLAS.ti qualitative data analysis software version 7.5.15 [19], using a coding dictionary and thematic analysis techniques, as commonly described, to assess content validity [20–25] (further information provided in Additional file 2a).

### Psychometric analysis of FACIT-fatigue in PsA

Subsequently to the qualitative assessment, a series of analyses assessed the quantitative psychometric properties of the FACIT-Fatigue scale, based on data from the phase III RCTs OPAL Broaden (NCT01877668) [26] and OPAL Beyond (NCT01882439) [27]. These analyses were pre-specified in a psychometric statistical analysis plan.

OPAL Broaden was a 12-month RCT in patients with an inadequate response to ≥1 conventional synthetic disease-modifying antirheumatic drug (csDMARD) and

who were tumor necrosis factor inhibitor (TNFi)-naive. Patients ($n = 422$) were randomized 2:2:2:1:1 to tofacitinib 5 mg twice daily (BID; $n = 107$), tofacitinib 10 mg BID ($n = 104$), adalimumab 40 mg once every 2 weeks ($n = 106$), placebo advancing to tofacitinib 5 mg BID at Month 3 ($n = 52$), or placebo advancing to tofacitinib 10 mg BID at Month 3 ($n = 53$) [26]. OPAL Beyond was a 6-month RCT in patients who had an inadequate response to ≥1 TNFi (TNFi-IR). Patients ($n = 395$) were randomized 2:2:1:1 (394 patients received treatment) to tofacitinib 5 mg BID ($n = 131$; one patient randomized but not treated), tofacitinib 10 mg BID ($n = 132$), placebo advancing to tofacitinib 5 mg BID at Month 3 ($n = 66$), or placebo advancing to tofacitinib 10 mg BID at Month 3 ($n = 65$) [27]. In both RCTs, patients received a stable background dose of one csDMARD.

FACIT-Fatigue data from both RCTs were pooled across all treatment groups to provide the largest sample size and response range to the individual items. Two different pooling strategies were used. Strategy 1: Pooled Data 1 (PD1; OPAL Beyond baseline data pooled with OPAL Broaden Month 12 [last study visit]; number of observations = 760, one observation per patient) and Pooled Data 2 (PD2; OPAL Broaden baseline data pooled with OPAL Beyond Month 6 [last study visit]; number of observations = 766, one observation per patient) were used in the cross-sectional analyses (i.e., internal consistency reliability, confirmatory analyses, and correlations). Strategy 2: for longitudinal analyses (i.e., test-retest, clinically important difference [CID], and responder definition [RD]), Pooled Data 3 (PD3) was used, corresponding to all available data from OPAL Broaden pooled longitudinally with all available data from OPAL Beyond.

### Confirmatory factor analysis model
The FACIT-Fatigue measurement model was based on the conceptual framework and was represented by a second-order confirmatory factor analysis. This measurement model was evaluated using PD1 and PD2 and included the two FACIT-Fatigue scale scores and the total score. It was assumed that the latent construct "Experience" (represented by the first-order factor f1) affects items 1, 2, 3, 4, and 7 of FACIT-Fatigue and the latent construct "Impact" (represented by the first-order factor f2) affects all other nine items. The latent aggregated factor (represented by the second-order factor f3) affects "Experience" and "Impact" domains (Additional file 2: Figure S2 and factor loadings shown in Additional file 2: Figure S3).

Bentler's Comparative Fit Index (CFI) was used to measure the fit of the model with the data. An acceptable fit was defined as: 1) CFI > 0.90; 2) unstandardized path coefficients are statistically significant ($P$ value < 0.05); and 3) standardized path coefficients are > 0.40 and are statistically significant.

### Internal consistency reliability
Cronbach's Coefficient $\alpha$ assessed internal consistency reliability of FACIT-Fatigue, with good internal consistency defined as a Cronbach's coefficient $\alpha \geq 0.90$ (Additional file 2d details FACIT-Fatigue conceptual framework).

### Test-retest reliability
Intraclass Correlation Coefficients (ICC) estimated test-retest reliability using baseline and Month 1 data. Because of the treatment intervention, a subgroup of "stable" patients was used in the analysis, with an ICC ≥ 0.70 defined as acceptable [28]. To define a stable subgroup, the Patient's Global Assessment (PtGA; a component of the Patient's Global Joint and Skin Assessment) was used. PtGA was formulated as follows: "In all the ways in which your psoriasis and arthritis, as a whole, affects you, how would you rate the way you felt over the past week?". PtGA is a Visual Analog Scale (VAS) from 0 mm (poor) to 100 mm (excellent). To estimate ICC in this analysis, it was assumed that a less than 10 mm difference at Month 1 from baseline represents a "stable" patient.

### Convergent validity
Evidence of convergent validity (the extent to which two concepts are related to one another [29]) was evaluated by correlation of the FACIT-Fatigue domain scores with other outcomes from the same studies (SF-36 domains, Itch Severity Item [ISI], Dermatology Life Quality Index [DLQI] total score, and Patient's Global Assessment of Psoriasis and Arthritis [PtGA], Patient's Skin Assessment [PtSA], and Patient's Joint Assessment [PtJA], which are components of the Patient's Global Joint and Skin Assessment − Visual Analog Scale [PtGJS-VAS]). Correlations of FACIT-Fatigue with these outcomes were expected to be ≥0.40, previously considered a moderate correlation [30].

### Defining the clinically important difference for FACIT-fatigue domains
Clinically important difference (CID), the difference in scores between two treatment groups that is considered clinically relevant, was estimated using a repeated measures model (RMM), assessing the relationship between the PtGA score and FACIT-Fatigue domains in PD3. A domain (Impact or Experience) of FACIT-Fatigue (including total score) is the outcome, and PtGA is a continuous or categorical anchor (RMM-CID). SF-36 Vitality domain was also used as an anchor in additional sensitivity analyses.

When using PtGA as an anchor, it is important to note that it is a VAS; hence, there are no clear patient-selected categories to use as a basis to define a CID. If it is assumed that 100 mm VAS PtGA (used in OPAL Broaden and OPAL Beyond) can be linearly approximated by a 7-point scale (e.g., Patient Global Impression-Severity), then it can

Cella *et al. Journal of Patient-Reported Outcomes* (2019) 3:5

Page 4 of 11

be assumed that a value of 17 mm could be representative of the one-category difference and could be used to estimate the CID for a FACIT-Fatigue domain (note that 17 mm = 100 mm/6, where 6 is the number of pairwise adjacent categories) (further details in Additional file 2a) [31, 32].

### Defining the responder definition for FACIT-fatigue domains

Responder definition (RD), the amount of change an individual patient would have to report to indicate that a relevant treatment benefit has been experienced, was estimated using a RMM to assess the relationship between a new anchor, the "Subject Global Impression of Change" (SGIC) score with just three categories ("better", "the same", and "worse"), and FACIT-Fatigue domains in PD3 (RMM-RD) (further details in Additional file 2a).

### Known-groups validity

Known-groups validity was evaluated based on a RMM-CID model by comparing FACIT-Fatigue scores between groups known to be different based on PtGA as the criteria. Ability to detect change was based on a RMM-CID model by examining the relationship between FACIT-Fatigue scores and PtGA. Patients were classified as "in remission/low disease" if they reported a score of 0 mm on the PtGA, and patients were classified as "active disease" if they reported a score of 100 mm.

Effect sizes were estimated by dividing the difference in score by standard deviation at baseline, and provide a general set of thresholds or benchmarks through adjectival descriptors on the difference between groups or impact of an intervention, with values of 0.2 generally regarded as "small," 0.5 as "medium," and 0.8 as "large".

### Study oversight

OPAL Broaden (NCT01877668) [26] and OPAL Beyond (NCT01882439) [27] were conducted in accordance with the International Conference on Harmonisation Good Clinical Practice Guidelines and the Declaration of Helsinki. The study protocols and all documentation were approved by the Institutional Review Boards or Independent Ethics Committees at each investigational site. All study procedures complied with current Health Insurance Portability and Accountability Act of 1996 (HIPAA) regulations. All recruitment locations were approved by a central institutional review board (E&I IRB #2 – IRB00007807), and all recruitment procedures adhered to the IRB-approved study protocol. All patients provided written informed consent.

## Results

### Qualitative FACIT-fatigue study

In total, 12 interviews were conducted in February 2017 at two clinical sites (Florida, *n* = 7; Pennsylvania, *n* = 5). The mean age (standard deviation; SD [range]) of patients was 53 (14 [27–80]) years, 6 (50%) patients were male, and 11 (92%) were white. The mean time since diagnosis of PsA (SD [range]) was approximately 10 (9 [1–29]) years. Most patients (*n* = 10, 83%) were currently taking medication/treatment for PsA, including methotrexate (*n* = 5, 42%), adalimumab, etanercept, secukinumab (each *n* = 2, 17%), and others (*n* = 3, 30%).

### PsA symptoms, concept elicitation

As part of the concept elicitation portion of the interview (part 1; Additional file 2b), patients were asked to rate the severity of their PsA and then to rank the importance of their symptoms.

PsA severity was highly variable, described by patients as mild (*n* = 3, 25%), mild to moderate (*n* = 1, 7%), moderate (*n* = 3, 25%), moderate to severe (*n* = 2, 17%), sometimes moderate and sometimes severe (*n* = 1, 7%), severe at first but diminished (*n* = 1, 7%), or severe (*n* = 1, 8%). PsA signs/symptoms experienced over the past 7 days were fatigue (*n* = 12, 100%), pain (in joints, tendons, or entheses; *n* = 11, 92%), skin-related symptoms (itch, dryness, scaling, redness, bleeding, inflammation, or painful skin; *n* = 9, 75%), joint stiffness (any part of body; *n* = 7, 58%), dactylitis (swelling of entire fingers or toes; *n* = 6, 50%), swelling in other parts of body (*n* = 4, 33%), and other symptoms (*n* = 7, 58%). Seven patients (58%) decisively attributed fatigue to PsA.

Patients ranked each symptom relative to their other symptoms from 0 to 4 (0 is "not important at all"; 4 is "extremely important"). Symptoms rated as "important" or "extremely important" are presented in Table 1.

**Table 1** Ranking of PsA symptoms by patients in the qualitative study (*N* = 12)

| Symptom | Rated as "important" or "extremely important", *n* (%) |
| --- | --- |
| Pain (in joints, tendons, or entheses) | 10 (83) |
| Fatigue (tired/listless/lack of energy/washed out/low energy/weak) | 7 (58) |
| Dactylitis (swelling of entire finger[s] or toe[s]) | 6 (50) |
| Swelling (not in finger[s] or toe[s])[a] | 5 (42) |
| Skin-related symptoms (itch, dryness, scaling, redness, bleeding, inflammation, painful skin) | 5 (42) |
| Joint stiffness (any part of the body) | 3 (25) |
| Other symptoms[b] | 6 (50) |

*PsA* psoriatic arthritis

[a]Swelling described as visible inflammation in the feet/ankles (*n* = 4), hips (*n* = 2), jaw (*n* = 1), elbows (*n* = 1), or knees (*n* = 1)

[b]Other symptoms that were ranked as important or extremely important include feeling depressed/down (*n* = 2), burning in knees (*n* = 1), cramping in the knees (*n* = 1), thumb "twisting" (*n* = 1), and weak hands (*n* = 1)

### FACIT-fatigue qualitative interview

Subsequently to the concept elicitation portion of the interview, the cognitive portion of the interview (part 2) asked patients to complete the FACIT-Fatigue questionnaire and to provide feedback.

Mean total FACIT-Fatigue score (SD [range]) was 27.1 (10.8 [13–44]) out of a possible maximum score of 52, with this low value, relative to the total score, indicating higher fatigue. Mean Experience domain (SD [range]) score was 7.4 (4.4 [1–15]; highest possible score 20), and average Impact domain score (SD [range]) was 19.7 (6.8 [12–29]; highest possible score 32). During the FACIT-Fatigue interview, patients with PsA generally provided positive feedback on the instrument. All 12 patients commented that completing the questionnaire was "quick," "easy," "straightforward," and "fine" and found the instructions, item wording, and response options clear and easily understood. Overall impressions of the items were favorable, although one patient indicated that the first four items were repetitive (fatigued, weak all over, listless [washed out], tired).

The recall period (past 7 days) was correctly understood by most patients ($n = 7$, 58%); however, other patients ($n = 5$, 42%) did not use the correct recall period, instead reporting their fatigue experiences over the "past month", "in general", "today", "yesterday", "all the time", and "during the day". Two of these patients reported that they read the instructions but decided to consider a different recall period for their answers. Most patients considered FACIT-Fatigue items 1–9 and 12 (range $n = 10$ [83%] to $n = 12$ [100%]) to be relevant to their experience with PsA. Items 11 "I need help doing my usual activities" and 13 "I have to limit my social activity because I am tired" were considered relevant by nine patients each (75%). Item 10 "I am too tired to eat" was not considered relevant by eight patients (67%).

Most patients ($n = 9$, 75%) reported that there were no important fatigue-related concepts missing from the questionnaire. The remaining three patients (25%) provided suggestions for improvements to existing items, and for additional items/concepts, including making a distinction between physical and mental fatigue ($n = 2$) and asking patients how they relieve their fatigue. One patient suggested incorporating questions that addressed the mental and emotional aspect of PsA.

As most patients reported that no important fatigue-related concepts were missing and did not suggest any additional items to be assessed, no changes to the FACIT-Fatigue items and response options were recommended.

### Psychometric analysis of FACIT-fatigue in PsA
### Confirmatory analysis model

The FACIT-Fatigue measurement model was tested using confirmatory factor analysis, which included two first-order factors (representing Experience and Impact domains) and one aggregated second-order factor (representing total score). CFI indices were 0.92 and 0.93 for PD1 and PD2, respectively, and standardized factor loadings were > 0.4 for all items.

### Internal consistency reliability

Cronbach's Coefficient α was ≥0.90 for the FACIT-Fatigue total score, Impact domain, and Experience domain for both PD1 and PD2 (Table 2). All corrected item-to-total correlations were > 0.40 (range 0.42–0.89).

### Test-retest reliability

An acceptable test-retest reliability was observed for FACIT-Fatigue Experience domain (ICC = 0.80), Impact domain (0.83), and total score (0.83) using pooled data from the OPAL Broaden and OPAL Beyond RCTs. Test-retest reliability assessments for each separate RCT were also acceptable (Additional file 3).

### Convergent validity

The correlation between the FACIT-Fatigue domains and other scales used in phase III RCTs was estimated using PD1 and PD2. With the exception of the Health Transition Item (which has a recall period of 1 year), correlations between FACIT-Fatigue and SF-36 domains generally exceeded 0.60 (all were > 0.50; $P < 0.0001$; Table 3). The correlation between FACIT-Fatigue total score and Experience domain and SF-36 Vitality domain was > 0.80 ($P < 0.0001$). FACIT-Fatigue domain scores also correlated with ISI, DLQI total score, PtGA, PtSA, and PtJA (correlations > 0.4).

**Table 2** Internal consistency reliability of FACIT-Fatigue in patients with PsA

| | Cronbach's Coefficient α | Corrected item-to-total correlations, range |
|---|---|---|
| PD1: OPAL Beyond baseline data pooled with OPAL Broaden Month 12 | | |
| FACIT-Fatigue total score ($N = 760$) | 0.95 | 0.46–0.89 |
| FACIT-Fatigue Experience domain ($N = 763$) | 0.93 | 0.59–0.89 |
| FACIT-Fatigue Impact domain ($N = 762$) | 0.91 | 0.44–0.85 |
| PD2: OPAL Broaden baseline data pooled with OPAL Beyond Month 6 | | |
| FACIT-Fatigue total score ($N = 766$) | 0.94 | 0.44–0.89 |
| FACIT-Fatigue Experience domain ($N = 768$) | 0.91 | 0.53–0.86 |
| FACIT-Fatigue Impact domain ($N = 768$) | 0.90 | 0.42–0.83 |

*FACIT-Fatigue* Functional Assessment of Chronic Illness Therapy–Fatigue, *PD1/2* Pooled Data 1/2, *PsA* psoriatic arthritis

**Table 3** Correlations to assess the convergent validity of FACIT-Fatigue vs. other measures in patients with PsA

| FACIT-Fatigue vs.: | PD1: OPAL Beyond baseline data pooled with OPAL Broaden Month 12 data, Pearson correlation coefficient | | | PD2: OPAL Broaden baseline data pooled with OPAL Beyond Month 6 data, Pearson correlation coefficient | | |
|---|---|---|---|---|---|---|
| | Experience domain | Impact domain | Total score | Experience domain | Impact domain | Total score |
| SF-36 | | | | | | |
|   Bodily pain | 0.71 | 0.69 | 0.72 | 0.66 | 0.64 | 0.68 |
|   General health | 0.57 | 0.55 | 0.58 | 0.59 | 0.55 | 0.59 |
|   Mental health | 0.65 | 0.64 | 0.66 | 0.65 | 0.65 | 0.67 |
|   Physical functioning | 0.64 | 0.69 | 0.69 | 0.56 | 0.62 | 0.62 |
|   Role emotional | 0.63 | 0.70 | 0.70 | 0.56 | 0.64 | 0.63 |
|   Role physical | 0.71 | 0.75 | 0.76 | 0.63 | 0.71 | 0.70 |
|   Social functioning | 0.71 | 0.78 | 0.78 | 0.66 | 0.73 | 0.72 |
|   Vitality | 0.83 | 0.75 | 0.81 | 0.81 | 0.75 | 0.81 |
|   Health transition | −0.33 | −0.29 | −0.32 | −0.35 | −0.33 | −0.35 |
|   Mental component | 0.68 | 0.70 | 0.71 | 0.66 | 0.69 | 0.70 |
|   Physical component | 0.64 | 0.65 | 0.67 | 0.58 | 0.60 | 0.61 |
| ISI | −0.42 | −0.37 | −0.41 | −0.43 | −0.41 | −0.44 |
| DLQI total score | −0.43 | −0.43 | −0.44 | −0.43 | −0.48 | −0.48 |
| PtGA | −0.64 | −0.60 | −0.64 | −0.61 | −0.57 | −0.61 |
| PtJA | −0.65 | −0.62 | −0.65 | −0.61 | −0.57 | −0.61 |
| PtSA | −0.47 | −0.45 | −0.47 | −0.44 | −0.44 | −0.46 |

All results $P < 0.0001$. Correlation coefficients are all in the expected direction

*DLQI* Dermatology Life Quality Index, *FACIT-Fatigue* Functional Assessment of Chronic Illness Therapy–Fatigue, *ISI* Itch Severity Item, *PD1/2* Pooled Data 1/2, *PtGA* Patient's Global Assessment of Psoriasis and Arthritis (a component of the PtGJS-VAS), *PtGJS* Patient's Global Joint and Skin Assessment, *PtJA* Patient's Joint Assessment (a component of the PtGJS-VAS), *PtSA* Patient's Skin Assessment (a component of the PtGJS-VAS), *PsA*: psoriatic arthritis, *SF-36* Short Form Survey-36, *VAS* Visual Analog Scale

### Defining the clinically important difference for FACIT-fatigue domains

CID for FACIT-Fatigue was defined by employing a longitudinal RMM to estimate the relationship between PtGA score and FACIT-Fatigue domains, and linked to a 17 mm change (one category difference on a 7-point scale) on the PtGA. Pooled data showed that PtGA had a substantial correlation with FACIT-Fatigue domains at all time points (with values between 0.5 and 0.7 for post-treatment time points) and with correlations < 0.5 at baseline.

The CID for the FACIT-Fatigue total score was 3.1, and for FACIT-Fatigue Experience and Impact domains was estimated to be 1.5 and 1.7, respectively (Table 4). In the sensitivity analysis, CIDs for each RCT were similar.

### Estimation of the responder definition for FACIT-fatigue domains

A RMM was applied to estimate RD and examine the relationship between FACIT-Fatigue domains and SGIC score as the anchor (see Additional file 2a). SGIC is based on PtGA change from baseline, but has only three categories: "worse" (change from baseline ≥10 mm; value of − 1), "the same" (change from baseline < 10 mm; value of 0), and "better" (change from baseline ≤ − 10 mm; value of + 1).

RD for the FACIT-Fatigue total score was 3.8, and estimated to be 1.7 and 2.1 for FACIT-Fatigue Experience and Impact domains, respectively. In the sensitivity analysis, RDs for the individual RCTs were similar (Table 4). As a whole number would need to be assigned to denote improvement in an individual, this would therefore appear as 4 points for the FACIT-Fatigue total score, and 2 points for each of the domain scores.

### Known-groups validity

The known-groups validity analysis was based on a RMM-CID model and evaluated by analyzing the differences in mean FACIT-Fatigue domain scores between the "remission/low disease activity group" (PtGA score of 0 mm, i.e., "excellent") and the "active disease group" (PtGA score of 100 mm, i.e., "poor"). Differences in the FACIT-Fatigue domain scores and total score between "remission/low disease activity group" and the "active disease group" were statistically different; effect sizes of all differences considered large (all > 1.4), constituting a significant and considerable difference between the groups (Table 5).

### Ability to detect change

The ability to detect change analysis was based on a RMM-CID model. Figure 1 compares changes in FACIT-Fatigue

**Table 4** Clinically important difference (CID) and responder definition (RD) estimations for FACIT-Fatigue in patients with PsA

| Model | Analysis | Anchor | Study data | Experience domain | Impact domain | Total score |
|---|---|---|---|---|---|---|
| | | | | Clinically important difference (SE) [95% CI] | | |
| Repeated measures model-CID | Main analysis | PtGA | Pooled | 1.45 (0.04) | 1.73 (0.05) | 3.12 (0.08) |
| | | | | [1.38, 1.52] | [1.63, 1.82] | [2.97, 3.27] |
| | Sensitivity analysis | PtGA | OPAL Broaden | 1.39 (0.05) | 1.66 (0.06) | 3.00 (0.10) |
| | | | | [1.30, 1.48] | [1.53, 1.78] | [2.80, 3.19] |
| | | | OPAL Beyond | 1.53 (0.06) | 1.82 (0.08) | 3.29 (0.12) |
| | | | | [1.42, 1.64] | [1.67, 1.97] | [3.04, 3.53] |
| | | SF-36 Vitality | Pooled | 1.06 (0.01) | 1.21 (0.02) | 2.25 (0.03) |
| | | | | [1.03, 1.08] | [1.17, 1.25] | [2.18, 2.31] |
| | | | | Effect size (SD) | | |
| Repeated measures model-CID | Main analysis | PtGA | Pooled | 8.73 (4.54) | 18.93 (6.88) | 27.66 (10.95) |
| | Sensitivity analysis | PtGA | OPAL Broaden | 9.13 (4.34) | 19.56 (6.57) | 28.69 (10.47) |
| | | | OPAL Beyond | 8.30 (4.72) | 18.25 (7.14) | 26.54 (11.35) |
| | | | | Responder definition (SE) [95% CI] | | |
| Repeated measures model-RD | Main analysis | SGIC | Pooled | 1.67 (0.09) | 2.12 (0.12) | 3.76 (0.20) |
| | | | | [1.49, 1.86] | [1.88, 2.37] | [3.37, 4.15] |
| | Sensitivity analysis | | OPAL Broaden | 1.54 (0.12) | 2.02 (0.15) | 3.52 (0.25) |
| | | | | [1.31, 1.77] | [1.72, 2.33] | [3.03, 4.01] |
| | | | OPAL Beyond | 1.91 (0.16) | 2.30 (0.20) | 4.19 (0.33) |
| | | | | [1.60, 2.23] | [1.90, 2.70] | [3.54, 4.85] |

Pooled data: OPAL Broaden and OPAL Beyond

*CI* confidence interval, *CID* clinically important difference; *FACIT-Fatigue* Functional Assessment of Chronic Illness Therapy–Fatigue; *PsA* psoriatic arthritis, *PtGA* Patient's Global Assessment of Psoriasis and Arthritis, *RD* responder definition, *RMM* repeated measures model, *SD* standard deviation, *SE* standard error, *SF-36* Short Form Survey-36, *SGIC* Subject Global Impression of Change

total scores with changes in the PtGA scores, and indicates that a patient's state (as measured by FACIT-Fatigue) changes with respect to the PtGA.

## Discussion

Fatigue is recommended as a core domain to measure in RCTs evaluating treatment effects for psoriatic arthritis [4]. This study evaluated the content validity and quantitative measurement properties to assess whether FACIT-Fatigue is fit for purpose as a measure to evaluate this important domain in RCTs in patients with PsA. The US Food and Drug Administration (FDA) patient-reported outcome (PRO) guidance adds that for labeling claims, adequate evidence is required to support the content

validity, construct validity, reliability, and ability of the measure to detect change in the target population of interest [9]. This mixed-methods study evaluated these qualitative and quantitative measurement properties of the FACIT-Fatigue in patients with PsA.
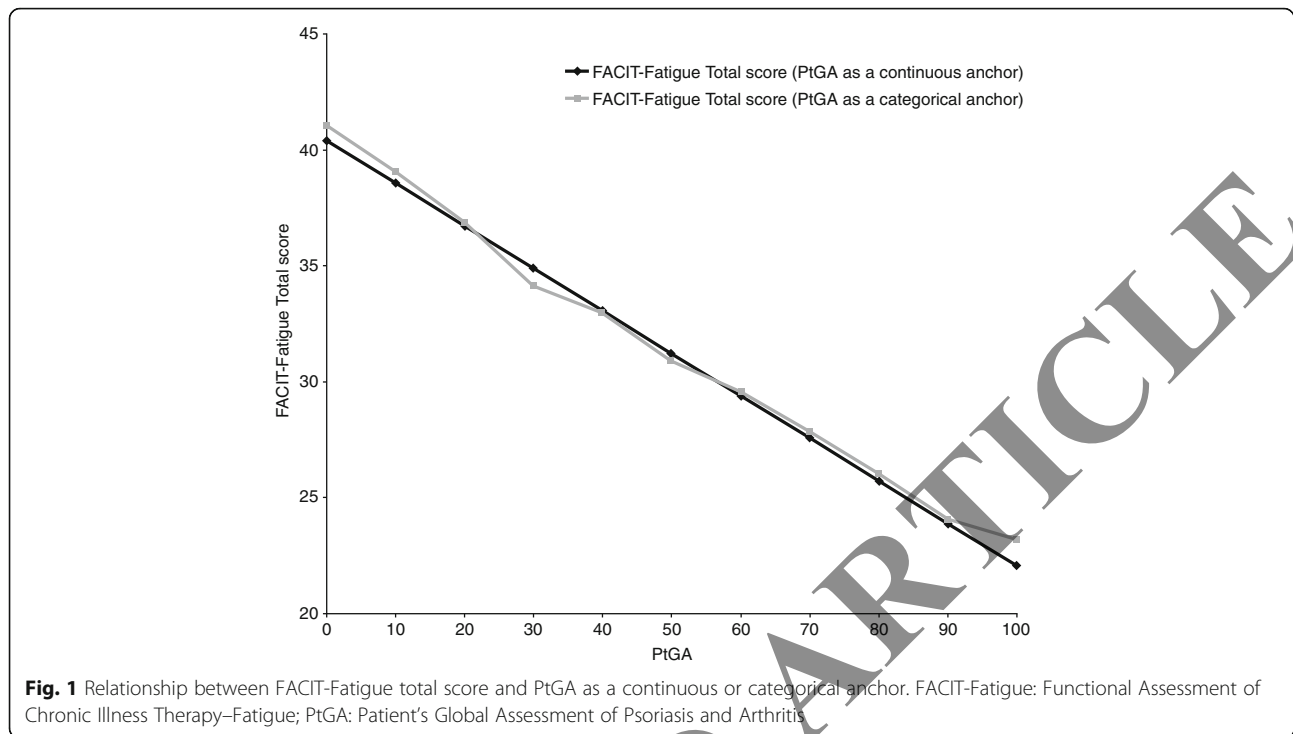
All patients reported experiencing several factors related to their fatigue, which impacted their daily life (e.g., social, psychological, and physical function). This confirms the importance of fatigue symptoms in patients with PsA, consistent with other studies that identify improvements in fatigue as a key outcome signifying improvement in their condition [4, 33, 34]. Furthermore, the reliability of reporting the physical and mental concepts of FACIT-Fatigue (Impact and Experience domains) is also consistent with

**Table 5** Known-groups validity for FACIT-Fatigue in patients with PsA

| Analysis | Differences in the FACIT-Fatigue domain scores between the "remission/low disease activity group" and the "active disease group" | | | | | |
|---|---|---|---|---|---|---|
| | Experience domain | | Impact domain | | Total score | |
| | Difference (95% CI) | Effect size | Difference (95% CI) | Effect size | Difference (95% CI) | Effect size |
| Anchor as a continuous variable | 8.52 (8.11, 8.93) | 1.88 | 10.15 (9.59, 10.71) | 1.48 | 18.35 (17.46, 19.24) | 1.68 |
| Anchor as a categorical variable | 8.22 (7.52, 8.91) | 1.81 | 10.06 (9.12, 10.99) | 1.46 | 17.89 (16.40, 19.38) | 1.63 |

Pooled data: OPAL Broaden and OPAL Beyond

*CI* confidence interval, *FACIT-Fatigue* Functional Assessment of Chronic Illness Therapy–Fatigue, *PsA* psoriatic arthritis

**Fig. 1** Relationship between FACIT-Fatigue total score and PtGA as a continuous or categorical anchor. FACIT-Fatigue: Functional Assessment of Chronic Illness Therapy–Fatigue; PtGA: Patient's Global Assessment of Psoriasis and Arthritis

the reliability of these concepts in other patients with other conditions, such as spinal cord injuries [35].

Overall, patients provided positive feedback on the FACIT-Fatigue questionnaire, believing it was comprehensive and relevant to their experience of fatigue with PsA. Results were similar to a study in patients with RA, where 15 of 17 patients stated that FACIT-Fatigue items were relevant to them. [18] Notably, item 10 "I am too tired to eat" was considered the least relevant item in both this study (8/12 patients, 67%) and the study in RA (9/17 patients, 53%) [18]. In this study, the instructions, item concepts, and response options were well-understood by most patients. Most correctly understood the recall period; however, some did not use the correct recall period. Overall, no changes to the FACIT-Fatigue items and response options were recommended.

In the psychometric analysis of RCT data in patients with PsA, the second-order confirmatory factor analysis model supported the measurement model of the FACIT-Fatigue scale. Good internal consistency reliability was seen in FACIT-Fatigue; Cronbach's Coefficient α's were ≥ 0.90, and all corrected item-to-total correlations were > 0.4. The ability to detect change, while part of instrument validity [36], is of sufficient importance to PRO measurement in longitudinal studies that it may be analyzed separately [28, 37], as done here. These findings demonstrated the sensitivity of FACIT-Fatigue to changes in PtGA scores. Results provided evidence that FACIT-Fatigue is equally sensitive to increases and decreases in PtGA scores, showing that

when a patient's experience of fatigue is predicted to change (i.e., change in severity of illness measured by PtGA), the values for FACIT-Fatigue also change. The test-retest reliability analysis observed an acceptable ICC (≥ 0.80) for all FACIT-Fatigue domains.

FACIT-Fatigue Impact and Experience domains were observed to correlate with almost all measured outcomes, suggesting that the physical and mental impacts of fatigue are closely linked to patient perception of PsA. Furthermore, FACIT-Fatigue total score was observed to correlate strongly ($r > 0.80$) with the SF-36 Vitality domain. As both fatigue and dermatological symptoms improve with PsA therapies (e.g., etanercept or adalimumab) [38, 39], it was expected here that FACIT-Fatigue scores would correlate with dermatological scores. However, ISI, DLQI, and PtSA scores (– 0.37 to – 0.48) were numerically lower than the correlations of FACIT-Fatigue scores with PtJA scores (– 0.57 to – 0.65), potentially indicating that FACIT-Fatigue is more related and sensitive to the effects of arthritis than psoriasis.

Different terms and approaches have been used to characterize and formulate a CID (between-group difference) and RD (within-individual or within-group change) for PROs [40, 41], and some have been used in rheumatology [42, 43]. Here, the CID of FACIT-Fatigue is the clinically relevant difference in scores between two treatment groups, and the RD is the amount of improvement an individual patient would have to report to indicate experience of a relevant treatment benefit. It is therefore

akin to a CID that has been reported in rheumatology [42, 43]. RD was estimated using a RMM, based on the algorithm recommended in the FDA guidance [9].

FACIT-Fatigue domain scores were significantly different between the "remission/low disease activity group" and the "active disease group", corroborating known-groups validity. The CID was defined using PtGA as an anchor and for the FACIT-Fatigue total score was 3.1. This is consistent with the value of 3–4 points reported in patients with other diseases, including cancer and RA [12, 44]. The RD for the FACIT-Fatigue total score was estimated to be a 4-point improvement, based on the average 3.8-point improvement associated with SGIC improvement. Overall, results were highly consistent with previous findings for FACIT-Fatigue [12, 15].

The 13 items of FACIT-Fatigue are also embedded in the Patient-Reported Outcomes Measurement Information System® (PROMIS®) Fatigue item bank, a 95-item fatigue assessment tool. This can be used as either a computerized adaptive test or a fixed-length short form, and was designed to compare differences across a range of chronic conditions, enabling comparative effectiveness research [45]. The use of fatigue short forms from PROMIS has been validated in RA [46], and the current research provides strong evidence supporting the validity of the FACIT-Fatigue scale and its measurement properties in patients with PsA, which opens up the possibility for including PsA data in the unifying PROMIS metric.

Advantages/strengths of this study included the self-reported nature of the PRO measures, and the systematic collection of clinical and PRO data. Moreover, patients' demographic and disease characteristics were well balanced. However, as data were taken from RCTs with specific eligibility criteria, generalizing these data to real-world populations may not be possible. Test-retest reliability, performed separately for OPAL Broaden and OPAL Beyond, confirmed the acceptability of the test-retest reliability from the pooled results.

Limitations of these analyses include that estimated CID (between-group difference) and RD (within-individual or within-group change) may vary due to different methodology and natural sampling variation, along with other considerations, and may not necessarily represent a minimal value [40]. Furthermore, changes in the anchor measures may not fully reflect CID in FACIT-Fatigue. Moreover, it would have been desirable to perform test-retest reliability assessments before treatment (i.e., during the screening [test] visit, and baseline [retest] visit); however, as these assessments were not available, test-retest reliability was performed in a stable group of patients at baseline and Month 1 (based on a < 10 mm difference in PtGA from baseline to Month 1), and provided the largest number of patients within the shortest possible time period.

It should be noted that in the qualitative interviews, the reported range of scores (range 13–44) did not include those for the most severe fatigue; therefore, concepts considered not relevant (e.g., "I'm too tired to eat") may remain relevant in patients with more severe fatigue. It also remains unclear how specific the patient feedback reported in this study is to the FACIT-Fatigue measure, or if this is also applicable to similar measures (e.g., Multidimensional Assessment of Fatigue). Furthermore, use of pooled data from two RCTs with different eligibility criteria, and use of different time points from each study, may confound the results.

## Conclusion

In summary, the findings of this study, including analyses performed for the first time using data from RCTs in PsA, suggest that the content of the FACIT-Fatigue scale is valid for use as an endpoint to measure fatigue in PsA RCTs. Qualitative interviews demonstrated that fatigue was an important symptom to patients with PsA, and the FACIT-Fatigue scale was capable of effectively capturing the relevant and important concepts of fatigue in this patient population. Analysis of FACIT-Fatigue data from two PsA RCTs showed good content validity and reliability, and a strong correlation with other disease measures. These conclusions, in conjunction with confirmations of CID and RD consistent with previous findings, support the use of FACIT-Fatigue in PsA RCTs.

## Additional files

**Additional file 1: Figure S1.** The FACIT-Fatigue scale. (DOCX 267 kb)

**Additional file 2:** FACIT-Fatigue factor loadings. (DOCX 889 kb)

**Additional file 3: Table S1.** Test/retest reliability of FACIT-Fatigue in patients with PsA. (DOCX 26 kb)

## Author details
[1]Department of Medical Social Sciences, Northwestern University, 633 N. St. Clair Suite 1900, Chicago, IL 60611, USA. [2]Evidera, Bethesda, MD, USA. [3]Evidera, Waltham, MA, USA. [4]Pfizer Inc, Groton, CT, USA.

## References

1. Gladman, D. D., Antoni, C., Mease, P., Clegg, D. O., & Nash, P. (2005). Psoriatic arthritis: Epidemiology, clinical features, course, and outcome. *Annals of the Rheumatic Diseases, 64*(Suppl 2), ii14–ii17.
2. Coates, L. C., Kavanaugh, A., Mease, P. J., Soriano, E. R., Laura Acosta-Felquer, M., Armstrong, A. W., et al. (2016). Group for Research and Assessment of psoriasis and psoriatic arthritis 2015 treatment recommendations for psoriatic arthritis. *Arthritis & Rhematology, 68*(5), 1060–1071.
3. Gudu, T., Etcheto, A., de Wit, M., Heiberg, T., Maccarone, M., Balanescu, A., et al. (2016). Fatigue in psoriatic arthritis - a cross-sectional study of 246 patients from 13 countries. *Joint, Bone, Spine, 83*(4), 439–443. https://doi.org/10.1016/j.jbspin.2015.07.017.
4. Orbai, A. M., de Wit, M., Mease, P., Shea, J. A., Gossec, L., Leung, Y. Y., et al. (2017). International patient and physician consensus on a psoriatic arthritis core outcome set for clinical trials. *Annals of the Rheumatic Diseases, 76*(4), 673–680.
5. Orbai, A. M., de Wit, M., Mease, P. J., Callis Duffin, K., Elmamoun, M., Tillett, W., et al. (2017). Updating the psoriatic arthritis (PsA) Core domain set: A report from the PsA workshop at OMERACT 2016. *The Journal of Rheumatology, 44*(10), 1522–1528.
6. Gladman, D., Fleischmann, R., Coteur, G., Woltering, F., & Mease, P. J. (2014). Effect of certolizumab pegol on multiple facets of psoriatic arthritis as reported by patients: 24-week patient-reported outcome results of a phase III, multicenter study. *Arthritis Care and Research, 66*(7), 1085–1092.
7. Strand, V., Schett, G., Hu, C., & Stevens, R. M. (2013). Patient-reported health-related quality of life with apremilast for psoriatic arthritis: A phase II, randomized, controlled study. *The Journal of Rheumatology, 40*(7), 1158–1165.
8. Strand, V., Mease, P., Gossec, L., Elkayam, O., van den Bosch, F., Zuazo, J., et al. (2017). Secukinumab improves patient-reported outcomes in subjects with active psoriatic arthritis: Results from a randomised phase III trial (FUTURE 1). *Annals of the Rheumatic Diseases, 76*(1), 203–207.
9. FDA. Guidance for industry patient-reported outcome measures: Use in medical product development to support labeling claims. 2009. https://www.fda.gov/downloads/drugs/guidances/ucm193282.pdf. Accessed 18 Jan 2019.
10. Revicki, D., Cella, D., Hays, R., Sloan, J., Lenderking, W., Aaronson, N. (2006). Responsiveness and minimal important differences for patient reported outcomes. *Health and Quality of Life Outcomes, 4*(70). http://www.hqlo.com/content/4/1/70
11. Revicki, D. A., Osoba, D., Fairclough, D., Barofsky, I., Berzon, R., Leidy, N. K., et al. (2000). Recommendations on health-related quality of life research to support labeling and promotional claims in the United States. *Quality of Life Research, 9*(8), 887–900.
12. Cella, D., Yount, S., Sorensen, M., Chartash, E., Sengupta, N., & Grober, J. (2005). Validation of the functional assessment of chronic illness therapy fatigue scale relative to other instrumentation in patients with rheumatoid arthritis. *The Journal of Rheumatology, 32*(5), 811–819.
13. Pouchot, J., Kherani, R. B., Brant, R., Lacaille, D., Lehman, A. J., Ensworth, S., et al. (2008). Determination of the minimal clinically important difference for seven fatigue measures in rheumatoid arthritis. *Journal of Clinical Epidemiology, 61*(7), 705–713.
14. Chandran, V., Bhella, S., Schentag, C., & Gladman, D. D. (2007). Functional Assessment of Chronic Illness Therapy-Fatigue scale is valid in patients with psoriatic arthritis. *Annals of the Rheumatic Diseases, 66*(7), 936–939.
15. Yellen, S. B., Cella, D. F., Webster, K., Blendowski, C., & Kaplan, E. (1997). Measuring fatigue and other anemia-related symptoms with the functional assessment of Cancer therapy (FACT) measurement system. *Journal of Pain and Symptom Management, 13*(2), 63–74.
16. Cella, D., Lai, J. S., & Stone, A. (2011). Self-reported fatigue: One dimension or more? Lessons from the functional assessment of chronic illness therapy--fatigue (FACIT-F) questionnaire. *Supportive Care in Cancer: Official Journal of the Multinational Association of Supportive Care in Cancer, 19*(9), 1441–1450. https://doi.org/10.1007/s00520-010-0971-1.
17. Rothman, M., Burke, L., Erickson, P., Leidy, N. K., Patrick, D. L., & Petrie, C. D. (2009). Use of existing patient-reported outcome (PRO) instruments and their modification: The ISPOR good research practices for evaluating and documenting content validity for the use of existing instruments and their modification PRO task force report. *Value in Health, 12*(8), 1075–1083. https://doi.org/10.1111/j.1524-4733.2009.00603.x.
18. Kaiser, K., Shaunfield, S., Clayman, M. L., Ruderman, E., & Cella, A. (2016). Content validation of the functional assessment of chronic illness therapy (FACIT)-fatigue scale in moderately to highly active rheumatoid arthritis. *Rheumatology Current Research, 6*(2). https://doi.org/10.4172/2161-1149.1000193
19. Friese, S., & Ringmayr, T. (2013). *ATLAS.ti 7 user guide and reference*. Berlin: ATLAS.ti Scientific Software Development GmbH.
20. Willis, G. (2015). *Analysis of the cognitive interview in questionnaire design. Understanding qualitative research*. New York: Oxford University Press.

21. Welch, L. C., Trudeau, J. J., Silverstein, S. M., Sand, M., Henderson, D. C., & Rosen, R. C. (2017). Initial development of a patient-reported outcome measure of experience with cognitive impairment associated with schizophrenia. *Patient Related Outcome Measures, 8*, 71–81. https://doi.org/10.2147/PROM.S123266.

22. Patrick, D. L., Burke, L. B., Gwaltney, C. J., Leidy, N. K., Martin, M. L., Molsen, E., et al. (2011). Content validity—Establishing and reporting the evidence in newly developed patient-reported outcomes (PRO) instruments for medical product evaluation: ISPOR PRO good research practices task force report: Part 1—Eliciting concepts for a new PRO instrument. *Value in Health, 14*(8), 967–977. https://doi.org/10.1016/j.jval.2011.06.014.

23. Patrick, D. L., Burke, L. B., Gwaltney, C. J., Leidy, N. K., Martin, M. L., Molsen, E., et al. (2011). Content validity—Establishing and reporting the evidence in newly developed patient-reported outcomes (PRO) instruments for medical product evaluation: ISPOR PRO good research practices task force report: Part 2—Assessing respondent understanding. *Value in Health, 14*(8), 978–988. https://doi.org/10.1016/j.jval.2011.06.013.

24. Boeije, H. (2002). A purposeful approach to the constant comparative method in the analysis of qualitative interviews. *Quality and Quantity, 36*(4), 391–409. https://doi.org/10.1023/A:1020909529486.

25. Leidy, N. K., & Vernon, M. (2008). Perspectives on patient-reported outcomes : Content validity and qualitative research in a changing clinical trial environment. *Pharmacoeconomics, 26*(5), 363–370.

26. Mease, P., Hall, S., Fitzgerald, O., van der Heijde, D., Merola, J. F., Avila-Zapata, F., et al. (2017). Tofacitinib or adalimumab versus placebo for psoriatic arthritis. *The New England Journal of Medicine, 377*(16), 1537–1550.

27. Gladman, D., Rigby, W., Azevedo, V. F., Behrens, F., Blanco, R., Kaszuba, A., et al. (2017). Tofacitinib for psoriatic arthritis in patients with an inadequate response to TNF inhibitors. *The New England Journal of Medicine, 377*(16), 1525–1536.

28. Reeve, B. B., Wyrwich, K. W., Wu, A. W., Velikova, G., Terwee, C. B., Snyder, C. F., et al. (2013). ISOQOL recommends minimum standards for patient-reported outcome measures used in patient-centered outcomes and comparative effectiveness research. *Quality of Life Research, 22*(8), 1889–1905. https://doi.org/10.1007/s11136-012-0344-y.

29. McDowell, I. (2006). *Measuring health: A guide to rating scales and questionnaires.* USA: Oxford University Press.

30. Tveter, A. T., Dagfinrud, H., Moseng, T., & Holm, I. (2014). Measuring health-related physical fitness in physiotherapy practice: Reliability, validity, and feasibility of clinical field tests and a patient-reported measure. *The Journal of Orthopaedic and Sports Physical Therapy, 44*(3), 206–216. https://doi.org/10.2519/jospt.2014.5042.

31. Norman, G. R., Sloan, J. A., & Wyrwich, K. W. (2003). Interpretation of changes in health-related quality of life: The remarkable universality of half a standard deviation. *Medical Care, 41*(5), 582–592. https://doi.org/10.1097/01.Mlr.0000062554.74615.4c.

32. Norman, G. R., Sloan, J. A., & Wyrwich, K. W. (2004). The truly remarkable universality of half a standard deviation: Confirmation through another look. *Expert Review of Pharmacoeconomics & Outcomes Research, 4*(5), 581–585. https://doi.org/10.1586/14737167.4.5.581.

33. Overman, C. L., Kool, M. B., da Silva, J. A., & Geenen, R. (2016). The prevalence of severe fatigue in rheumatic diseases: An international study. *Clinical Rheumatology, 35*(2), 409–415.

34. Gossec, L., de Wit, M., Kiltz, U., Braun, J., Kalyoncu, U., Scrivo, R., et al. (2014). A patient-derived and patient-reported outcome measure for assessing psoriatic arthritis: Elaboration and preliminary validation of the psoriatic arthritis impact of disease (PsAID) questionnaire, a 13-country EULAR initiative. *Annals of the Rheumatic Diseases, 73*(6), 1012–1019.

35. Palimaru, A. I., Cunningham, W. E., Dillistone, M., Vargas-Bustamante, A., Liu, H., & Hays, R. D. (2018). Development and psychometric evaluation of a fatigability index for full-time wheelchair users with spinal cord injury. *Archives of Physical Medicine and Rehabilitation.* https://doi.org/10.1016/j.apmr.2018.04.003.

36. Hays, R. D., & Hadorn, D. (1992). Responsiveness to change: An aspect of validity, not a separate dimension. *Quality of Life Research, 1*(1), 73–75.

37. Mokkink, L. B., Terwee, C. B., Patrick, D. L., Alonso, J., Stratford, P. W., Knol, D. L., et al. (2010). The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. *Journal of Clinical Epidemiology, 63*(7), 737–745. https://doi.org/10.1016/j.jclinepi.2010.02.006.

38. Gladman, D. D., Bombardier, C., Thorne, C., Haraoui, B., Khraishi, M., Rahman, P., et al. (2011). Effectiveness and safety of etanercept in patients with psoriatic arthritis in a Canadian clinical practice setting: The REPArE trial. *The Journal of Rheumatology, 38*(7), 1355–1362. https://doi.org/10.3899/jrheum.100698.

39. Paul, C., van de Kerkhof, P., Puig, L., Unnebrink, K., Goldblum, O., & Thaci, D. (2012). Influence of psoriatic arthritis on the efficacy of adalimumab and on the treatment response of other markers of psoriasis burden: Subanalysis of the BELIEVE study. *European Journal of Dermatology, 22*(6), 762–769. https://doi.org/10.1684/ejd.2012.1863.

40. Sloan, J. A., Cella, D., & Hays, R. D. (2005). Clinical significance of patient-reported questionnaire data: Another step toward consensus. *Journal of Clinical Epidemiology, 58*(12), 1217–1219.

41. Revicki, D., Hays, R. D., Cella, D., & Sloan, J. (2008). Recommended methods for determining responsiveness and minimally important differences for patient-reported outcomes. *Journal of Clinical Epidemiology, 61*(2), 102–109.

42. Beaton, D. E., Bombardier, C., Katz, J. N., Wright, J. G., Wells, G., Boers, M., et al. (2001). Looking for important change/differences in studies of responsiveness. OMERACT MCID working group. Outcome measures in rheumatology. Minimal clinically important difference. *The Journal of Rheumatology, 28*(2), 400–405.

43. Wells, G., Beaton, D., Shea, B., Boers, M., Simon, L., Strand, V., et al. (2001). Minimal clinically important differences: Review of methods. *The Journal of Rheumatology, 28*(2), 406–412.

44. Cella, D., Eton, D. T., Lai, J. S., Peterman, A. H., & Merkel, D. E. (2002). Combining anchor and distribution-based methods to derive minimal clinically important differences on the functional assessment of Cancer therapy (FACT) anemia and fatigue scales. *Journal of Pain and Symptom Management, 24*(6), 547–561.

45. Cella, D., Lai, J. S., Jensen, S. E., Christodoulou, C., Junghaenel, D. U., Reeve, B. B., et al. (2016). PROMIS fatigue item Bank had clinical validity across diverse chronic conditions. *Journal of Clinical Epidemiology, 73*, 128–134.

46. Bartlett, S. J., Gutierrez, A. K., Butanis, A., Bykerk, V. P., Curtis, J. R., Ginsberg, S., et al. (2018). Combining online and in-person methods to evaluate the content validity of PROMIS fatigue short forms in rheumatoid arthritis. *Quality of Life Research.* https://doi.org/10.1007/s11136-018-1880-x.