Received: 30 November 2023 | Accepted: 19 January 2024

DOI: 10.1111/jdv.19830

JEADV JOURNAL OF THE EUROPEAN ACADEMY OF DERMATOLOGY & VENEREOLOGY

ORIGINAL ARTICLE

# Psychometric validation of the Psoriasis Symptom Scale, Functional Assessment of Chronic Illness Therapy–Fatigue and pain-Visual Analogue Scale in patients with generalized pustular psoriasis

A. David Burden[1] | Robert Bissonnette[2] | Milena Anatchkova[3] | Ismail Budhiarso[3] | Anne M. Skalicky[3] | Ana C. S. Liberato[4] | Na Hu[5] | Christian Thoma[6] | Tristan Gloede[6] | Thomas Kohlmann[7] | Mark G. Lebwohl[8]

[1]School of Infection and Immunity, University of Glasgow, Glasgow, UK

[2]Innovaderm Research Inc, Montréal, Quebec, Canada

[3]Evidera, Bethesda, Maryland, USA

[4]Evidera PPD, London, UK

[5]Boehringer Ingelheim (China) Investment Co. Ltd, Shanghai, China

[6]Boehringer Ingelheim International GmbH, Ingelheim, Germany

[7]Institute for Community Medicine, Medical University Greifswald, Greifswald, Germany

[8]Icahn School of Medicine at Mount Sinai, New York, New York, USA

**Correspondence**
Mark G. Lebwohl, Icahn School of Medicine at Mount Sinai, New York, New York, USA.
Email: lebwohl@aol.com

**Funding information**
Boehringer Ingelheim

## Abstract

**Background:** Generalized pustular psoriasis (GPP) is a rare, chronic, inflammatory skin disease associated with considerable patient burden. The Psoriasis Symptom Scale (PSS), Functional Assessment of Chronic Illness Therapy–Fatigue (FACIT-Fatigue) and pain-Visual Analogue Scale (pain-VAS) are patient-reported outcomes (PROs) that have not yet been validated in patients with GPP.

**Objectives:** To evaluate the psychometric properties of the PSS, FACIT-Fatigue and pain-VAS using data from Effisayil 1, a randomised trial of spesolimab in patients with moderate-to-severe GPP.

**Methods:** Inter-item correlations and confirmatory factor analysis (CFA) were performed using Week 1 data. Internal consistency was assessed with Cronbach's α coefficient using baseline and Week 1 data. Test–retest reliability was assessed using intraclass correlation coefficients (ICCs); change data for the GPP Physician Global Assessment total score and pustulation subscore were used to define a stable population. Convergent validity was assessed at baseline and Week 1 using Spearman's rank-order correlations. Known-groups validity was measured by analysis of variance using Week 1 data. Ability to detect change from baseline to Week 1 was evaluated by analysis of covariance.

**Results:** Inter-item and item-to-total correlations were moderate or strong for most PSS and FACIT-Fatigue items. CFA demonstrated the unidimensionality of the PSS and FACIT-Fatigue, with high factor loadings for most items (PSS range, 0.75–0.94; FACIT-Fatigue range, 0.11–0.93) and acceptable fit statistics. Both scores demonstrated internal consistency (Cronbach's α, 0.71 and 0.95, respectively). The PSS, FACIT-Fatigue and pain-VAS demonstrated test–retest reliability (ICCs ≥0.70) and good evidence of convergent validity. Furthermore, the PROs could differentiate between known groups of varying symptom severity (range, $p < 0.0001$–$0.0225$) and detect changes in symptom severity from baseline to Week 1 (range, $p < 0.0001$–$0.0002$).

**Conclusions:** Overall, these results support the reliability, validity and ability to detect change of the PSS, FACIT-Fatigue and pain-VAS as PROs in patients with GPP.

# INTRODUCTION

Generalized pustular psoriasis (GPP) is a rare, chronic inflammatory skin disease characterised by the widespread eruption of painful skin pustules, erythema and scaling.[1,2] Skin lesions may or may not be accompanied by symptoms of systemic inflammation, such as fever and fatigue. These can, in turn, develop into life-threatening complications, including sepsis and multisystem organ failure.[1,3,4] Moreover, patients with GPP often experience debilitating comorbidities, such as hypertension, arthritis, cardiovascular disease, liver dysfunction, obesity and chronic obstructive pulmonary disease.[3,5,6] The signs and symptoms associated with GPP flares, such as burning, itching, pain and fatigue, have an adverse impact on patient health-related quality of life (QoL).[7]

Recently, the GPP Physician Global Assessment (GPPGA) and GPP Area and Severity Index (GPPASI) have been developed and validated as suitable clinician-reported outcomes (ClinROs) for assessing GPP severity.[8] However, the use of patient-reported outcomes (PROs) in clinical studies is also needed to fully capture symptoms and impact on QoL from the patient perspective.[9] PROs used in previous trials in GPP include the Psoriasis Symptom Scale (PSS), Functional Assessment of Chronic Illness Therapy–Fatigue (FACIT-Fatigue) and pain-Visual Analogue Scale (pain-VAS), which assess symptom severity, as well as the Dermatology Life Quality Index (DLQI) and 36-Item Short Form Health Survey (SF-36), which assess impact on QoL.[10–12] There is existing evidence supporting the content validity of the PSS in GPP[13]; however, the psychometric properties of PROs have not been evaluated in a GPP patient population, as is recommended by the US Food and Drug Administration (FDA).[14,15]

Effisayil 1 (NCT03782792) was a randomised, placebo-controlled trial of spesolimab, an anti-interleukin-36 receptor monoclonal antibody, in patients with moderate-to-severe GPP.[12] PROs collected throughout the study period included the PSS, FACIT-Fatigue and pain-VAS scores, among others. Here, we assess the psychometric properties of the PSS, FACIT-Fatigue and pain-VAS using data from Effisayil 1, including their reliability, validity and ability to detect change in patients with GPP.

# MATERIALS AND METHODS

## Study design

In Effisayil 1, 53 patients presenting with a GPP flare were randomised (2:1) to receive a single intravenous dose of spesolimab 900 mg ($n = 35$) or matching placebo ($n = 18$). Eligible patients were aged between 18 and 75 years, had a history of GPP as per the European Rare and Severe Psoriasis Expert Network (ERASPEN) diagnostic criteria and had a GPP flare of moderate-to-severe intensity at baseline. The primary trial endpoint was a GPPGA pustulation subscore of 0 at Week 1, and the key secondary endpoint was a GPPGA total score of 0 or 1 at Week 1. With regard to PROs, secondary trial endpoints included the changes from baseline in PSS score, FACIT-Fatigue score and pain-VAS score, all assessed at Week 4. Full details on the study design have been published previously.[12,16]

## Patient-reported outcomes

All PRO assessments were completed in a designated quiet area prior to any other assessments or treatments. The PSS is a four-item instrument originally developed to assess disease severity in patients with plaque psoriasis.[17] Patients report their symptoms of pain, redness, itching and burning over the past 24 hours on a five-point scale (0 = none; 1 = mild; 2 = moderate; 3 = severe; 4 = very severe). Individual item scores are tallied to give an unweighted total PSS score, with higher scores indicating more severe disease.

The FACIT-Fatigue is a 13-item questionnaire that assesses a patient's fatigue and the consequent impact on their daily activities.[18] Patients rate a list of 13 statements on a five-point scale (0 = not at all; 1 = a little; 2 = somewhat; 3 = quite a bit; 4 = very much) to indicate how strongly the statements reflect their experience over the previous week. Items are reverse-scored and aggregated to give a total score (0–52), with lower scores indicating worse fatigue.

The pain-VAS is a single-item measure of pain intensity for which patients place a vertical mark on a horizontal line (usually 100 mm in length) to indicate their pain severity; one end of the continuous scale is labelled as 'no pain' and the other as 'very severe pain'.[19] The distance between the 'no pain' anchor and the patient's mark is measured to give a score ranging from 0 to 100, with higher scores indicating greater pain intensity. In Effisayil 1, patients used the pain-VAS to report how much pain they had experienced due to GPP over the previous week.

## Psychometric analyses

The psychometric analyses performed in this study are in line with those recommended in FDA guidance on PRO validation.[14] Baseline patient demographics and characteristics are presented descriptively. Only psychometric analyses from the first week of the study are presented due to limited variance in the PRO data for some analyses at Week 4.

Psychometric properties of the PSS, FACIT-Fatigue and pain-VAS were evaluated using data from Effisayil 1. ClinROs used for validation analyses were the GPPGA and GPPASI. The GPPGA and GPPASI are novel, validated ClinROs developed for the assessment of GPP-specific disease severity.[8,9] The GPPGA has three subscores assessing pustulation, erythema and scaling, which are key features of GPP. The GPPGA total score is the average of the subscores, rounded to the nearest integer, with higher scores indicating higher disease severity. The GPPASI

includes a severity score assessing pustulation, erythema and scaling across several body regions, as well as a score assessing the area of involvement for each body region on a seven-point scale. The GPPASI total score is the sum of the individual scores for each body region, with higher scores reflecting more severe disease. PROs used for validation analyses were the DLQI, EuroQol 5-Dimension 5-Level (EQ-5D-5L) and EuroQol VAS (EQ-VAS). The DLQI is a 10-item questionnaire that asks patients to rate how much their skin disease affects their QoL using a four-point ordinal scale (0 = not at all; 1 = a little; 2 = a lot; 3 = very much).[20] The questionnaire is widely used in clinical trials in dermatology and evaluates the impact of the patient's disease on their daily activities, leisure activities, work/school life, personal relationships, symptoms/feelings and treatment-related distress over the previous week. The overall DLQI score is the total of individual item scores (0–30), with higher scores indicating greater impairment of QoL. The EQ-5D-5L and the EQ-VAS are patient-reported measures of general health developed by the EuroQol group.[21] For the EQ-5D-5L, patients are asked to report their health on the current day across five dimensions (mobility, self-care, usual activities, pain/discomfort and anxiety/depression) on a five-point scale (0 = no problems; 1 = slight problems; 2 = moderate problems; 3 = severe problems; 4 = extreme problems). Individual scores are combined, and a weighted overall score is calculated using coefficients from a validated value set. Of note, only the EQ-5D pain/discomfort item was used in this study. Similar to the pain-VAS, the EQ-VAS is a vertical continuous scale from 0 to 100 on which patients mark a judgement of their general health that day; a score of 0 represents the worst health imaginable, and 100 represents the best health imaginable.

### Inter-item correlations and confirmatory factor analysis

Item-to-item and item-to-total Pearson correlations were measured for the PSS and FACIT-Fatigue scores, with moderate correlations (>0.40) expected for optimal items.[22] Confirmatory factor analysis (CFA) was performed for the PSS and FACIT-Fatigue using an unconstrained model, with factor loadings >0.40 considered as acceptable evidence that an item contributes to the construct measurement.[23] A comparative fit index (CFI) ≥0.9, a standardised root mean residual (SRMR) <0.1 and a root mean square error of approximation (RMSEA) <0.08 are considered as acceptable measures of fit.[24–26]

### Reliability: Internal consistency and test–retest reliability

Internal consistency assessed the degree of agreement between items in the PSS and FACIT-Fatigue instruments using Cronbach's α coefficient (performed with data from baseline and Week 1); coefficients >0.70 indicate good internal consistency.[27] Inter-item correlations, CFA and internal consistency analysis were not performed for the pain-VAS as it is a single-item instrument. To assess the test–retest reliability of the PROs, intraclass correlation coefficients (ICCs) were calculated using change data for the GPPGA total score and GPPGA pustulation subscore (Day 3–4 and Week 3–4) to define a stable population. ICCs >0.70 are considered acceptable for establishing test–retest reliability.[22,28]

### Validity: Convergent validity and known-groups validity

Convergent validity between PROs and related instruments (DLQI total score, DLQI item 1 [itchy, sore, painful and stinging skin], EQ-5D pain/discomfort item, EQ-VAS score) was assessed at baseline and Week 1 using Spearman's rank-order correlation coefficients. The PSS, FACIT-Fatigue and pain-VAS scores were hypothesised to correlate with greater symptom burden as per the DLQI, EQ-5D pain/discomfort item and the EQ-VAS score. Therefore, known-groups validity was performed using these PROs as anchor categories to group patients by different levels of disease severity. Analysis of variance was used to determine group differences, and *F*-statistics were examined.

### Ability to detect change

To evaluate the ability of the PSS, FACIT-Fatigue and pain-VAS to detect change between baseline and Week 1, responder groups were determined using DLQI item 1, EQ-5D pain/discomfort item and EQ-VAS score as anchor categories. Differences in mean PSS, FACIT-Fatigue and pain-VAS scores across groups were tested by analysis of covariance adjusted for baseline scores, and *F*-statistics were examined.

## RESULTS

### Baseline patient demographics and characteristics

Patients in this study had a mean (standard deviation [SD]) age of 43.0 (10.9) years, were mostly female (67.9%) and were either Asian (54.7%) or White (45.3%; Table 1). Mean (SD) body mass index was 27.0 (8.3) kg/m$^2$, and most patients had never smoked (71.7%). At baseline, 18.9% of patients had an *IL36RN* mutation, although mutation status was unknown for most patients (62.3%). A total of 46 out of 53 patients underwent genetic testing as part of the trial, 14 of whom (30.4%) had a confirmed *IL36RN* mutation.

## Inter-item correlations and confirmatory factor analysis

For the PSS, item-to-total correlations were moderate to strong, with the strongest correlations observed between the total score and the 'burning' item (0.94; $p < 0.0001$; Table S1). Between items, the correlations were strongest between 'pain' and 'burning' (0.84; $p < 0.0001$) and 'pain' and 'redness' (0.79; $p < 0.0001$). For FACIT-Fatigue, item-to-total correlations were moderate to strong for all items, with the strongest correlations observed between the total score and the 'I have trouble finishing things because I am tired' (−0.94; $p < 0.0001$) and 'I have trouble starting things because I am tired' (−0.93; $p < 0.0001$; Table S2) items. Inter-item correlations were moderate to strong across almost all items, except the 'I have energy' (range, −0.13 to −0.41) and the 'I am able to do usual activities' (range, −0.17 to 0.51) items, which had weak-to-moderate correlations with other items. CFA demonstrated the unidimensionality of the PSS (Table 2) and FACIT-Fatigue (Table 3) total scores. For the PSS, there were high factor loadings for each item (range, 0.75–0.94), and fit statistics were within the thresholds for acceptable fit (CFI ≥0.9; RMSEA <0.08; SRMR <0.1). Similarly, factor loadings were high for most FACIT-Fatigue items (range, 0.105–0.934); fit statistics were within the thresholds for acceptable fit for the CFI and SRMR but not for the RMSEA.

**TABLE 1** Baseline patient demographics and characteristics.

| Demographic | *N* = 53 |
|---|---|
| Age, mean (SD) | 43.0 (10.9) |
| Sex, *n* (%) | |
| Male | 17 (32.1) |
| Female | 36 (67.9) |
| Ethnicity, *n* (%) | |
| White | 24 (45.3) |
| Asian | 29 (54.7) |
| BMI, kg/m$^2$, mean (SD) | 27.0 (8.3) |
| Smoking status, *n* (%) | |
| Never | 38 (71.7) |
| Former | 4 (7.5) |
| Current | 11 (20.8) |
| *IL36RN* mutation (historical data)[a], *n* (%) | |
| No | 10 (18.9) |
| Yes | 10 (18.9) |
| Unknown | 33 (62.3) |

Abbreviations: BMI, body mass index; SD, standard deviation.

[a]Of the 53 enrolled patients, 46 underwent genetic testing as part of the study, 14 of whom (30.4%) had an *IL36RN* mutation.

**TABLE 2** Confirmatory factor analysis at Week 1: PSS.

| PSS item | Factor loading |
|---|---|
| Pain | 0.90 |
| Redness | 0.88 |
| Itching | 0.75 |
| Burning | 0.94 |
| **Fit statistic** | **Value** |
| CFI | 1.00 |
| RMSEA (90% CI) | 0 (0–0.179) |
| SRMR | 0.007 |

Abbreviations: CFI, comparative fit index; CI, confidence interval; PSS, Psoriasis Symptom Scale; RMSEA, root mean square error of approximation; SRMR, standardised root mean residual.

## Internal consistency and test–retest reliability

Both the PSS and FACIT-Fatigue scores demonstrated internal consistency, with Cronbach's α coefficients within the acceptable range (≥0.70) at baseline (0.71 and 0.95, respectively) and Week 1 (0.92 and 0.95, respectively; Table 4). Overall, PSS, FACIT-Fatigue and pain-VAS scores all demonstrated satisfactory test–retest reliability when using the GPPGA total score and GPPGA pustulation subscore to define a stable population. When using the changes in GPPGA total score and GPPGA pustulation subscore from Week 3 to 4 to define a stable population, ICCs were ≥0.70 for all three scores (Table 4). ICCs were marginally lower than the threshold for acceptability for the pain-VAS score (0.67) and the FACIT-Fatigue score

**TABLE 3** Confirmatory factor analysis at Week 1: FACIT-Fatigue.

| FACIT-Fatigue item | Factor loading |
|---|---|
| I feel fatigue | 0.888 |
| I feel weak | 0.922 |
| I feel listless | 0.887 |
| I feel tired | 0.892 |
| I have trouble starting things because I am tired | 0.934 |
| I have trouble finishing things because I am tired | 0.914 |
| I have energy | 0.304 |
| I am able to do usual activities | 0.263 |
| I need to sleep | 0.615 |
| I am too tired to eat | 0.594 |
| I need help doing usual activities | 0.657 |
| I am frustrated by being tired | 0.862 |
| I have to limit social activity | 0.105 |
| **Fit statistic** | **Value** |
| CFI | 0.921 |
| RMSEA (90% CI) | 0.126 (0.087–0.163) |
| SRMR | 0.056 |

Abbreviations: CFI, comparative fit index; CI, confidence interval; FACIT-Fatigue, Functional Assessment of Chronic Illness Therapy–Fatigue; RMSEA, root mean square error of approximation; SRMR, standardised root mean residual.

*Note*: Five modifications were made for each model to account for items with correlated model residuals.

**TABLE 4** Internal consistency and test–retest reliability analyses.

| | PSS | | FACIT-Fatigue | | Pain-VAS | |
|---|---|---|---|---|---|---|
| **Internal consistency reliability** | ***N*** | **Cronbach's α** | ***N*** | **Cronbach's α** | ***N*** | **Cronbach's α** |
| Baseline | 53 | 0.71 | 53 | 0.95 | – | – |
| Week 1 | 53 | 0.92 | 53 | 0.95 | – | – |
| **Test–retest reliability** | ***N*** | **ICC** | ***N*** | **ICC** | ***N*** | **ICC** |
| GPPGA total score stable definition from Day 3 to 4 | 34 | 0.75 | 16 | 0.73 | 16 | 0.67 |
| GPPGA pustulation subscore stable definition from Day 3 to 4 | 30 | 0.85 | 12 | 0.63 | 12 | 0.96 |
| GPPGA total score stable definition from Week 3 to 4 | 35 | 0.84 | 34 | 0.77 | 35 | 0.81 |
| GPPGA pustulation subscore stable definition from Week 3 to 4 | 42 | 0.74 | 41 | 0.70 | 42 | 0.80 |

Abbreviations: FACIT-Fatigue, Functional Assessment of Chronic Illness Therapy–Fatigue; GPPGA, Generalized Pustular Psoriasis Physician Global Assessment; ICC, intraclass correlation coefficient; pain-VAS, pain-Visual Analogue Scale; PSS, Psoriasis Symptom Scale.

*Note*: A Cronbach's α coefficient >0.70 indicates good internal consistency reliability. An ICC ≥0.70 is considered acceptable for establishing test–retest reliability.

**TABLE 5** Convergent validity analyses.

| | Correlations[a] | | |
|---|---|---|---|
| **PRO instrument** | **PSS** | **FACIT-fatigue** | **Pain-VAS** |
| Baseline | | | |
| DLQI total score | 0.46** | −0.55*** | 0.46** |
| DLQI item 1: How itchy, sore, painful or stinging has your skin been? | 0.58*** | −0.44* | 0.32* |
| EQ-5D pain/discomfort | 0.64*** | −0.43* | 0.58*** |
| EQ-VAS score | −0.43* | 0.66*** | −0.49** |
| Week 1 | | | |
| DLQI total score | 0.50** | −0.64*** | 0.58*** |
| DLQI item 1: How itchy, sore, painful or stinging has your skin been? | 0.61*** | −0.41* | 0.43* |
| EQ-5D pain/discomfort | 0.80*** | −0.52*** | 0.51*** |
| EQ-VAS score | −0.62*** | 0.44* | −0.41* |

Abbreviations: DLQI, Dermatology Life Quality Index; EQ-5D, EuroQol 5-Dimension; EQ-VAS, EuroQol Visual Analogue Scale; FACIT-Fatigue, Functional Assessment of Chronic Illness Therapy–Fatigue; pain-VAS, pain-Visual Analogue Scale; PRO, patient-reported outcome; PSS, Psoriasis Symptom Scale.

*Note*: Significance levels for correlation *p*-values are: *$p<0.05$; **$p<0.001$; ***$p<0.0001$.

[a]Spearman's rank-order correlation; correlation interpretation: <0.3 = weak; 0.3–0.7 = moderate; 0.7–0.9 = strong; >0.9 = very strong.

(0.63) when a stable population was defined by the change from Day 3 to 4 in GPPGA total score and GPPGA pustulation subscore, respectively (Table 4).

## Convergent validity and known-groups validity

There was good evidence of convergent validity for the PSS, FACIT-Fatigue and pain-VAS scores, with moderate, statistically significant correlations observed for the DLQI total score (range, 0.46 to −0.64), DLQI item 1 (range, 0.32–0.61), EQ-5D pain/discomfort item (range, −0.43 to 0.80) and the EQ-VAS score (range, −0.41 to 0.66) at baseline and Week 1 (Table 5). There was a strong correlation between the PSS score and EQ-5D pain/discomfort item at Week 1 (0.80; $p<0.0001$).

The PSS, FACIT-Fatigue and pain-VAS scores were all able to differentiate between known groups of varying symptom severity, with significantly higher mean scores in patients with severe self-reported symptoms, as per DLQI item 1 (range, $p<0.0001$–$0.0090$), EQ-5D pain/discomfort item (range, $p<0.0001$–$0.0007$) and EQ-VAS score (range, $p=0.0010$–$0.0225$) (Table 6).

## Ability to detect change

The PSS, FACIT-Fatigue and pain-VAS all demonstrated the ability to detect change in patient-reported symptoms from baseline to Week 1, with significantly different least squares (LS) mean change scores between 'worsened', 'no change' and 'improved' patient groups as per DLQI item 1, EQ-5D pain/discomfort item and EQ-VAS score (Table 7). For example, patients who improved from baseline to Week 1, as per the EQ-5D pain/discomfort item, had LS mean (standard error) change scores of −5.13 (0.62), 12.39 (1.60) and −32.70 (4.36) for the PSS, FACIT-Fatigue and pain-VAS, respectively (all $p<0.0001$; Table 7).

**TABLE 6** Known-groups validity at Week 1 using PRO grouping categories.

| | PSS | | | FACIT-Fatigue | | | Pain-VAS | | |
|---|---|---|---|---|---|---|---|---|---|
| **Grouping category** | ***N*** | **Mean (SD)** | **Overall *F*-test (*p*-value)** | ***N*** | **Mean (SD)** | **Overall *F*-test (*p*-value)** | ***N*** | **Mean (SD)** | **Overall *F*-test (*p*-value)** |
| DLQI item 1: How itchy, sore, painful or stinging has your skin been? (three-category) | | | | | | | | | |
| 0–1 (not at all/a little) | 22 | 4.1 (3.0) | <0.0001**** | 22 | 32.9 (11.1) | 0.0053**. | 22 | 32.9 (27.6) | 0.0090** |
| 2 (a lot) | 17 | 7.0 (3.7) | | 17 | 29.6 (11.7) | | 17 | 54.0 (17.9) | |
| 3 (very much) | 13 | 10.8 (4.0) | | 13 | 19.3 (11.9) | | 13 | 57.5 (29.0) | |
| EQ-5D pain/discomfort (four-category) | | | | | | | | | |
| No problems | 10 | 2.5 (2.3) | <0.0001**** | 10 | 36.0 (7.1) | 0.0007*** | 10 | 22.8 (27.2) | <0.0001**** |
| Slight problems | 20 | 5.0 (1.9) | | 20 | 31.7 (10.8) | | 20 | 45.8 (20.7) | |
| Moderate problems | 9 | 7.0 (3.1) | | 9 | 28.8 (14.5) | | 9 | 37.1 (26.0) | |
| Severe/extreme problems | 13 | 12.5 (2.6) | | 13 | 17.4 (10.5) | | 13 | 70.0 (18.1) | |
| EQ-VAS score (three-category) | | | | | | | | | |
| Bad to very bad (0–65) | 24 | 8.5 (4.3) | 0.0010** | 24 | 23.5 (13.4) | 0.0210* | 24 | 54.7 (26.3) | 0.0225* |
| Moderate (66–85) | 17 | 6.7 (4.0) | | 17 | 31.5 (9.2) | | 17 | 45.2 (22.1) | |
| Good to very good (86–100) | 11 | 2.9 (1.9) | | 11 | 34.6 (11.8) | | 11 | 27.9 (29.4) | |

Abbreviations: DLQI, Dermatology Life Quality Index; EQ-5D, EuroQol 5-Dimension; EQ-VAS, EuroQol Visual Analogue Scale; FACIT-Fatigue, Functional Assessment of Chronic Illness Therapy–Fatigue; pain-VAS, pain-Visual Analogue Scale; PRO, patient-reported outcome; PSS, Psoriasis Symptom Scale; SD, standard deviation.
*$p<0.05$; **$p<0.01$; ***$p<0.001$; ****$p<0.0001$.

**TABLE 7** PSS, FACIT-Fatigue and pain-VAS ability to detect change by DLQI item 1, EQ-5D pain/discomfort and EQ-VAS.

| | Worsened | | No change | | Improved | | Overall *F*-test | | |
|---|---|---|---|---|---|---|---|---|---|
| **Change score from baseline to Week 1** | ***N*** | **LS mean (SE)** | ***N*** | **LS mean (SE)** | ***N*** | **LS mean (SE)** | ***F*-test** | ***p*-value[a]** | **Effect size[b]** |
| DLQI item 1 change score (three-category) | | | | | | | | | |
| PSS | 7 | −0.60 (1.55) | 21 | −2.37 (0.87) | 23 | −5.39 (0.85) | 8.28 | 0.0002 | −1.0575 |
| FACIT-Fatigue | 7 | 8.69 (3.81) | 21 | 7.04 (2.18) | 23 | 12.79 (2.10) | 9.39 | <0.0001 | 0.6890 |
| Pain-VAS | 7 | −25.66 (10.87) | 21 | −22.03 (6.10) | 23 | −30.23 (5.86) | 9.00 | <0.0001 | −1.2026 |
| EQ-5D pain/discomfort change score (three-category) | | | | | | | | | |
| PSS | 7 | 0.33 (1.45) | 9 | −0.29 (1.22) | 36 | −5.13 (0.62) | 13.02 | <0.0001 | −1.0749 |
| FACIT-Fatigue | 7 | 4.35 (3.62) | 9 | 3.38 (3.21) | 36 | 12.39 (1.60) | 11.49 | <0.0001 | 0.6833 |
| Pain-VAS | 7 | −22.57 (9.88) | 9 | −2.84 (8.78) | 36 | −32.70 (4.36) | 13.23 | <0.0001 | −1.2094 |
| EQ-VAS change score (three-category) | | | | | | | | | |
| PSS | 11 | 0.69 (0.98) | 5 | 0.86 (1.42) | 36 | −5.47 (0.53) | 22.00 | <0.0001 | −1.0749 |
| FACIT-Fatigue | 11 | 1.34 (2.69) | 5 | 2.26 (4.07) | 36 | 13.36 (1.50) | 15.99 | <0.0001 | 0.6833 |
| Pain-VAS | 11 | −6.73 (7.81) | 5 | −5.53 (11.08) | 36 | −34.97 (4.20) | 15.46 | <0.0001 | −1.2094 |

Abbreviations: ANCOVA, analysis of covariance; DLQI, Dermatology Life Quality Index; EQ-5D, EuroQol 5-Dimension; EQ-VAS, EuroQol Visual Analogue Scale; ES, effect size; FACIT-Fatigue, Functional Assessment of Chronic Illness Therapy–Fatigue; LS, least squares; pain-VAS, pain-Visual Analogue Scale; PSS, Psoriasis Symptom Scale; SE, standard error.

[a]ANCOVA adjusted by baseline/Day 1 score and anchor change score.

[b]ES calculated using Cohen's d, a calculation of the difference of the means divided by the standard deviation at baseline/Day 1. ES: small ES = 0.20; moderate ES = 0.50; and large ES = 0.80. Pairwise comparisons were only calculated if there were five or more patients in each group.

## DISCUSSION

The use of PROs in clinical trials is important for the evaluation of aspects of the disease that may not be fully reflected in clinician assessments. To date, several PROs have been included as secondary endpoints in studies of GPP; however, the psychometric properties of these have not been evaluated in a GPP patient population. The results of our study support the reliability, validity and ability to detect change of the PSS, FACIT-Fatigue and pain-VAS scores in

measuring GPP severity and QoL impact from the patient perspective.

Scores for items in the PSS and FACIT-Fatigue correlated well with each other and with the respective total scores, and CFA confirmed that both scores are unidimensional (i.e. the items in the score exclusively capture the construct they were designed to capture). Overall, all three scores showed good test–retest reliability. Notably, ICCs were higher when analyses used a stable population defined using change scores at later time points (e.g. Week 3–4 rather than Day 3–4); this effect is likely because more patients were post-flare, and therefore, more stable, at later time points. Indeed, patient numbers were higher in analyses using later time points, particularly for FACIT-Fatigue and pain-VAS. The three scores also showed good evidence of convergent validity and were able to distinguish between patient groups with different self-reported symptom severity. Moreover, the PSS, FACIT-Fatigue and pain-VAS were also sensitive to change in disease status. Our findings complement recently published evidence validating the use of the GPPGA and GPPASI ClinROs in patients in GPP.[8] Taken together, these studies highlight these ClinROs and PROs as suitable outcomes that may be considered for use in future clinical trials in GPP.

It is important to acknowledge that, as this study used outcome data from a randomised clinical trial, it was not specifically designed to evaluate the psychometric properties of the PROs. In the trial, patients receiving treatment showed rapid improvements in skin symptoms. This may have affected some of the psychometric analyses. For example, patient numbers were low for the 'worsened' and 'no change' response categories in the ability to detect change analyses. Large effect sizes in these analyses should, therefore, be interpreted with caution. Furthermore, patient numbers in this study were considerably below the recommended minimum sample size of 130–150 for CFA. Nevertheless, sample size is a considerable challenge in rare diseases such as GPP, and a stand-alone psychometric study with a larger sample size ($N > 200$) may not be feasible.[29] In this context, the use of the Effisayil 1 data set, which is one of the largest and most diverse to date, is a key strength of this study. The study limitations highlight common challenges associated with outcome development and validation in rare diseases. Future psychometric studies in larger patient cohorts should replicate our analyses and investigate meaningful change estimates for these PROs in patients with GPP.

Overall, the results presented here are the first to demonstrate the reliability, validity and ability to detect change of the PSS, FACIT-Fatigue and pain-VAS PROs in a GPP patient population. Our findings suggest that these PROs are suitable for assessing patient-perceived disease severity and may represent important tools for the clinical management of GPP.

## ACKNOWLEDGEMENTS

The authors meet criteria for authorship as recommended by the International Committee of Medical Journal Editors (ICMJE). The authors did not receive payment related to the development of this manuscript. Boehringer Ingelheim was given the opportunity to review the manuscript for medical and scientific accuracy, as well as intellectual property considerations. Isabella Goldsbrough Alves, PhD, of Hyperion, OPEN Health Communications (London, UK), provided medical writing, editorial and formatting support, which were contracted and funded by Boehringer Ingelheim.

## FUNDING INFORMATION

The study was supported and funded by Boehringer Ingelheim.

## CONFLICT OF INTEREST STATEMENT

ADB reports consulting fees from AbbVie, Boehringer Ingelheim, Bristol Myers Squibb and UCB; honoraria from Almirall and Boehringer Ingelheim; and grant support from Boehringer Ingelheim, Bristol Myers Squibb and Novartis. RB is an advisory board member, consultant, speaker and/or investigator for and received honoraria and/or grants from AbbVie, Almirall, Alumis, Amgen, AnaptysBio, Arcutis, Aristea, Bausch Health, Boehringer Ingelheim, Boston, Bristol Myers Squibb, Dermavant Sciences, Eli Lilly, Escalier, Janssen, Kyowa Kirin, LEO Pharma, Nimbus, Novartis, Pfizer, Regeneron, Sienna, UCB, Ventyx Biosciences and Xencor; and an employee and shareholder of Innovaderm Research. MA, IB, AMS and ACSL are employees of Evidera, which was contracted by Boehringer Ingelheim for the purposes of this study. NH, CT and TG are employees of Boehringer Ingelheim. TK has received research funds from Grünenthal, Hexal AG, Pfizer and Sanofi-Aventis. MGL is an employee of the Icahn School of Medicine at Mount Sinai and has received research funds from AbbVie, Amgen, Arcutis, Avotres, Boehringer Ingelheim, Cara Therapeutics, Dermavant Sciences, Eli Lilly, Incyte, Janssen Research & Development, LEO Pharma, Ortho Dermatologics, Pfizer, Regeneron and UCB; and is a consultant for Aditum Bio, Allergan, Almirall, AltruBio, AnaptysBio, Arcutis, Aristea, Avotres, BirchBioMed, BMD Beauty, Boehringer Ingelheim, Brickell Biotech, Bristol Myers Squibb, Cara Therapeutics, Castle Biosciences, Celltrion, CorEvitas, Dermavant Sciences, Dr. Reddy's Laboratories, EMD Serono, EPI Health, Evelo Biosciences, Evommune, Facilitation of International Dermatology Education, Forte Biosciences, Foundation for Research and Education in Dermatology, Galderma, Genentech, Helsinn, Incyte, Inozyme Pharma, Kyowa Kirin, LEO Pharma, Meiji Seika Pharma, Menlo, Mindera, Mitsubishi, Neuroderm, Pfizer, Seanergy, Strata, Theravance Biopharma, Trevi and Verrica Pharmaceuticals.

## DATA AVAILABILITY STATEMENT

To ensure independent interpretation of clinical study results and enable authors to fulfil their role and obligations under the International Committee of Medical Journal Editors (ICMJE) criteria, Boehringer Ingelheim grants all external authors access to clinical study data pertinent to the development of the publication. In adherence with

the Boehringer Ingelheim Policy on Transparency and Publication of Clinical Study Data, scientific and medical researchers can request access to clinical study data when it become available on https://vivli.org/, and earliest after publication of the primary manuscript in a peer-reviewed journal, regulatory activities are complete and other criteria are met. Please visit Medical & Clinical Trials | Clinical Research | MyStudyWindow for further information.

## ORCID

*Robert Bissonnette* https://orcid.org/0000-0001-5927-6587

## REFERENCES

1. Fujita H, Gooderham M, Romiti R. Diagnosis of generalized pustular psoriasis. Am J Clin Dermatol. 2022;23:31–8.
2. Navarini AA, Burden AD, Capon F, Mrowietz U, Puig L, Köks S, et al. European consensus statement on phenotypes of pustular psoriasis. J Eur Acad Dermatol Venereol. 2017;31:1792–9.
3. Choon SE, Lai NM, Mohammad NA, Nanu NM, Tey KE, Chew SF. Clinical profile, morbidity, and outcome of adult-onset generalized pustular psoriasis: analysis of 102 cases seen in a tertiary hospital in Johor, Malaysia. Int J Dermatol. 2014;53:676–84.
4. Puig L, Choon SE, Gottlieb AB, Marrakchi S, Prinz JC, Romiti R, et al. Generalized pustular psoriasis: a global Delphi consensus on clinical course, diagnosis, treatment goals and disease management. J Eur Acad Dermatol Venereol. 2023;37:737–52.
5. Bachelez H. Pustular psoriasis: the dawn of a new era. Acta Derm Venereol. 2020;100:adv00034.
6. Okubo Y, Kotowsky N, Gao R, Saito K, Morita A. Clinical characteristics and health-care resource utilization in patients with generalized pustular psoriasis using real-world evidence from the Japanese Medical Data Center database. J Dermatol. 2021;48:1675–87.
7. Lebwohl M, Medeiros RA, Mackey RH, Harrold LR, Valdecantos WC, Flack M, et al. The disease burden of generalized pustular psoriasis: real-world evidence from CorEvitas' Psoriasis Registry. J Psoriasis Psoriatic Arthritis. 2022;7:71–8.
8. Burden AD, Bissonnette R, Lebwohl MG, Gloede T, Anatchkova M, Budhiarso I, et al. Psychometric validation of the generalized pustular psoriasis physician global assessment (GPPGA) and generalized pustular psoriasis area and severity index (GPPASI). J Eur Acad Dermatol Venereol. 2023;37:1327–35.
9. Burden AD, Choon SE, Gottlieb AB, Navarini AA, Warren RB. Clinical disease measures in generalized pustular psoriasis. Am J Clin Dermatol. 2022;23:39–50.
10. Morita A, Yamazaki F, Matsuyama T, Takahashi K, Arai S, Asahina A, et al. Adalimumab treatment in Japanese patients with generalized pustular psoriasis: results of an open-label phase 3 study. J Dermatol. 2018;45:1371–80.
11. Saeki H, Nakagawa H, Nakajo K, Ishii T, Morisaki Y, Aoki T, et al. Efficacy and safety of ixekizumab treatment for Japanese patients with moderate to severe plaque psoriasis, erythrodermic psoriasis and generalized pustular psoriasis: results from a 52-week, open-label, phase 3 study (UNCOVER-J). J Dermatol. 2017;44:355–62.
12. Bachelez H, Choon SE, Marrakchi S, Burden AD, Tsai TF, Morita A, et al. Trial of spesolimab for generalized pustular psoriasis. N Engl J Med. 2021;385:2431–40.
13. Burden AD, Mrowietz U, Skalicky AM, Rentz AM, Esser D, Gloede T, et al. Symptom experience and content validity of the Psoriasis Symptom Scale (PSS) in patients with generalized pustular psoriasis (GPP). Dermatol Ther (Heidelb). 2022;12:1367–81.
14. US Food and Drug Administration. Guidance for industry. Patient-reported outcome measures: use in medical product development to support labeling claims. 2009. Available from: https://www.fda.gov/media/77832/download [Accessed 11 July 2023].
15. US Food and Drug Administration. Patient-focused drug development: incorporating clinical outcome assessments into endpoints for regulatory decision-making. 2023. Available at: https://www.fda.gov/media/166830/download [Accessed 20 September 2023].
16. Choon SE, Lebwohl MG, Marrakchi S, Burden AD, Tsai TF, Morita A, et al. Study protocol of the global Effisayil 1 phase II, multicentre, randomised, double-blind, placebo-controlled trial of spesolimab in patients with generalized pustular psoriasis presenting with an acute flare. BMJ Open. 2021;11:e043666.
17. Rentz AM, Skalicky AM, Esser D, Zema C, Becker K, Bodhani A, et al. Reliability, validity, and the ability to detect change of the Psoriasis Symptom Scale (PSS) in patients with plaque psoriasis. J Dermatolog Treat. 2020;31:460–9.
18. Webster K, Cella D, Yost K. The Functional Assessment of Chronic Illness Therapy (FACIT) measurement system: properties, applications, and interpretation. Health Qual Life Outcomes. 2003;1:79.
19. Hawker GA, Mian S, Kendzerska T, French M. Measures of adult pain: Visual Analog Scale for pain (VAS pain), Numeric Rating Scale for Pain (NRS pain), McGill Pain Questionnaire (MPQ), Short-Form McGill Pain Questionnaire (SF-MPQ), Chronic Pain Grade Scale (CPGS), Short Form-36 Bodily Pain Scale (SF-36 BPS), and measure of Intermittent and Constant Osteoarthritis Pain (ICOAP). Arthritis Care Res (Hoboken). 2011;63:S240–52.
20. Finlay AY, Khan GK. Dermatology Life Quality Index (DLQI)-a simple practical measure for routine clinical use. Clin Exp Dermatol. 1994;19:210–6.
21. Herdman M, Gudex C, Lloyd A, Janssen MF, Kind P, Parkin D, et al. Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L). Qual Life Res. 2011;20:1727–36.
22. Nunnally JC. Psychometric theory. New York, NY: McGraw-Hill; 1967. p. xiii–640.
23. Ford JK, MacCallum RC, Tait M. The application of exploratory factor analysis in applied psychology: a critical review and analysis. Pers Psychol. 1986;39:291–314.
24. Byrne BM. Structural equation modeling with AMOS: basic concepts, applications, and programming. 2nd ed. New York, NY: Routledge/Taylor & Francis Group; 2010. p. xix–396.
25. Kline RB. Principles and practice of structural equation modeling. 4th ed. New York, NY: Guilford Press; 2016. p. xvii–534.
26. Browne MW, Cudeck R. Alternative ways of assessing model fit. Sociol Methods Res. 1992;21:230–58.
27. Cronbach LJ. Coefficient alpha and the internal structure of tests. Psychometrika. 1951;16:297–334.
28. Aaronson N, Alonso J, Burnam A, Lohr KN, Patrick DL, Perrin E, et al. Assessing health status and quality-of-life instruments: attributes and review criteria. Qual Life Res. 2002;11:193–205.
29. Kyriazos TA. Applied psychometrics: sample size and sample power considerations in factor analysis (EFA, CFA) and SEM in general. Psychology. 2018;9:2207–30.

## SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

**How to cite this article:** Burden AD, Bissonnette R, Anatchkova M, Budhiarso I, Skalicky AM, Liberato ACS, et al. Psychometric validation of the Psoriasis Symptom Scale, Functional Assessment of Chronic Illness Therapy–Fatigue and pain-Visual Analogue Scale in patients with generalized pustular psoriasis. J Eur Acad Dermatol Venereol. 2024;00:1–8. https://doi.org/10.1111/jdv.19830