Journal of Dermatological Treatment, 2014; 25: 8–14

ISSN: 0954-6634 print / 1471-1753 online
DOI: 10.3109/09546634.2013.769042

informa
healthcare

ORIGINAL ARTICLE

Reliability and validity of the psoriasis symptom inventory in patients with moderate-to-severe psoriasis

Dennis A. Revicki[1], Ying Jin[2], Hilary D. Wilson[1], Dina Chau[3] & Hema N. Viswanathan[3]

[1]*Outcomes Research, United Biosource Corporation, Bethesda, MD, USA,* [2]*Association of American Medical Colleges, Washington DC, USA and* [3]*Amgen, Inc., Thousand Oaks, CA, USA*

Objectives: **The psoriasis symptom inventory (PSI) is a patient-reported outcome measure for assessing symptom severity in patients with moderate-to-severe psoriasis. The primary objective of this study was to evaluate the measurement properties of the PSI.** ***Materials and methods:*** **Analyses of psychometric characteristics (reliability, convergent and known-groups validity,responsiveness, item performance, and dimensionality) were conducted using data from a Phase II trail to evaluate efficacy of brodalumab in subjects with moderate–to-severe psoriasis.** ***Results:*** **The PSI had excellent internal consistency (α = 0.93–0.98) and good test-retest reliability (ICCs = 0.77–0.87). Convergent and discriminant validity was indicated by moderate-to-strong correlations between the PSI and Dermatology Life Quality Index scores, and small correlations between PSI total scores and ShortfFrm-36 Health Survey mental health, role emotional, and role physical scales. Known groups validity was shown as mean PSI total scores varied by Psoriasis Area and Severity Index (PASI) and Static Physician's Global Assessment (sPGA) defined groups ($p < 0.001$). PSI total scores were responsive to changes in clinical status as assessed by PASI ($p < 0.001$) and sPGA ($p < 0.001$). Unidimensionality of the PSI was supported.** ***Conclusions:*** **The PSI is a short and valid unidimensional measure of psoriasis symptom severity that is well suited for use in clinical trials.**

Key words: patient-reported outcome, psoriasis, psoriasis symptom inventory, psychometric, validity, reliability

Introduction

Psoriasis is a common inflammatory skin disease occurring in 2% to 3% of the population worldwide (1,2). Moderate-to-severe plaque psoriasis is, for most patients, a chronic, life-long condition, and the health-related quality of life (HRQOL) of psoriasis patients is severely impaired (3–9). In addition to clinician's considerations regarding disease severity, the inclusion of a patient-reported outcome (PRO) measure to assess patient symptoms is important in evaluating treatment efficacy.

Currently, there are two measures that are frequently utilized in trials evaluating the impact of psoriatic treatments on PROs: the Dermatology Life Quality Index (DLQI) (10) and the Short Form (SF)-36 Health Survey (11,12). The DLQI is disease-specific, and assesses the symptoms of psoriasis and their impact on daily activities, leisure, work and school, and personal relationships. The SF-36 is a general health status measure that may be utilized in any disease population. Evidence exists regarding the utility of these instruments in psoriasis clinical trials (6–8,11); however, neither measure was developed in consistent with the FDA PRO guidance (13) leaving a gap in the availability of appropriate symptom severity instruments for use in psoriasis clinical trials.

To fill this gap, a new eight-item psoriasis symptom instrument, the Psoriasis Symptom Inventory (PSI), was developed to assess the severity of psoriasis symptoms in patients with moderate-to-severe psoriasis consistent with the FDA PRO guidance (13). The content of the PSI was evaluated and saturation of symptom concepts was supported by focus groups and interviews with patients (14). The appropriateness of severity ratings of psoriasis symptoms was supported by both the cognitive interview results and by expert clinician agreement. The preliminary reliability and validity of the PSI were evaluated based on real-world evidence in 139 patients with moderate-to-severe plaque psoriasis (15). Results indicated that the PSI had good distribution of response categories, good evidence of construct validity, and acceptable test-retest reliability.

Currently, no studies have evaluated the responsiveness of the PSI to changes in clinical status. The primary objectives of the current study were to (1) evaluate the reliability, construct, and known-groups validity of the PSI and (2) evaluate the responsiveness of the PSI to change over time following treatment for moderate-to-severe psoriasis.

Materials and methods

Study patients

Data collected from a Phase II clinical trial in patients with moderate-to-severe plaque psoriasis were used in this study. The Phase II clinical trial was a randomized, double-blind, placebo-controlled, multiple-dose study to evaluate the safety, tolerability, and efficacy of brodalumab, an anti-IL17 receptor A antibody, in subjects with moderate-to-severe psoriasis (16). The clinical trial involved a total of 198 subjects with moderate-to-severe psoriasis at 23 study sites worldwide. Each of the subjects was randomized to one of the five treatment arms (four doses of brodalumab and placebo) with ~40 subjects in each arm. Patients were eligible for the study if they were between 18 and 70 years of age, had stable plaque psoriasis for at least 6 months, received or were candidates for phototherapy or systemic psoriasis therapy, moderate-to-severe plaque psoriasis on a minimum of 10% of

Correspondence: Dennis A. Revicki, Outcomes Research, United BioSource Corporation, 7101 Wisconsin Avenue, Suite 600, Bethesda, MD 20814, USA. Tel: +1 3001 664 7261. Fax: +1 301 654 9864. E-mail: dennis.revicki@unitedbiosource.com

(Received 15 November 2012; accepted 28 December 2012)

their body surface area (BSA), and scored ≥12 on the Psoriasis Area and Severity Index (PASI) (17). Patients were excluded if they had nonplaque or drug-induced psoriasis, a recent serious infection or history of recurrent infections, or a serious concurrent medical illness or cancer, other than *in situ* cervical or non-melanoma skin cancers that had been successfully treated.

After randomization, subjects received subcutaneous injections of brodalumab (70, 140, or 210 mg at day 1 and weeks 1, 2, 4, 6, 8, 10, or 280 mg monthly) or placebo. The primary endpoint was the percentage improvement from baseline in the PASI score at week 12, and subjects were followed up for 22 weeks. The study protocol was approved by Institutional Review Boards, and each patient provided written informed consent prior to participating in the study. This analysis is based on pooled data from treatment and placebo groups in the Phase II study.

Study measures

Patient-reported outcomes.
Three PRO measures were included in the study, the PSI, DLQI, and SF-36 Health Survey. Subjects completed the PROs at baseline, and at weeks 2 (PSI and DLQI only), 4, 8, 12, 16, and 22.

Psoriasis symptom inventory. The PSI (14,15) is an eight-item psoriasis-specific, PRO measure used in assessing the severity of psoriasis symptoms. The PSI contains items on itching, redness, scaling, burning, stinging, cracking, flaking, and pain. Subjects rated the severity of each symptom on average over the last 7 days on a five-point Likert-type rating scale ranging from "not at all" to "very severe". A PSI total score was calculated by summing the eight items (range: 0–32), and higher scores indicate greater symptom severity.

Dermatology life quality index. The DLQI (10) is a 10-item, subject completed, dermatology-specific, HRQOL measure. The DLQI is designed to assess the impact of skin disease on symptoms and feelings, daily activities, leisure activities, work and school, personal relationships, and treatment-related distress of subjects. The recall period of the DLQI is the past week. DLQI items are rated on a four-point Likert scale ranging from 0 (not at all) to 3 (very much). The DLQI item scores are summed into a DLQI total score that ranges from 0 to 30, with the lower scores indicating better HRQOL. A score of 0 to 1 is interpreted as having "no effect at all on a patient's life", a score of 2 to 5 as a "small effect on patient's life", 6 to 10 a "moderate effect on patient's life", 11 to 20 a "very large effect on patient's life", and 21 to 31 an "extremely large effect on patient's life". Ranges for the individual subscales are as follows: symptoms and feelings (0–6), daily activities (0–6), leisure (0–6), work and school (0–3), personal relationships (0–6), and treatment-related distress (0–3).

Short form-36 health survey version 2. The SF-36 Health Survey (SF-36v2™ Standard, US version 2.0) (18) is a generic assessment that measures general health status. The SF-36 includes eight domains rated over the prior four weeks: physical function, role limitations – physical, bodily pain, general health perceptions, vitality, social function, and role limitations with regard to emotional and mental health. Two component scores, the Physical Component Summary (PCS) and the Mental Component Summary (MCS) are calculated. The SF-36 domain and summary scores are transformed to a normative scale with a mean of 50 and standard deviation (SD) of 10, with higher scores indicating better physical function or well-being.

Clinician-reported outcomes.
The clinician-rated outcomes included the PASI (17), Static Physician's Global Assessment (sPGA) (19), and BSA (20). The clinician-rated scales were administered at screening, baseline, and at weeks 2, 4, 6, 8, 10, 12, 16, and 22 by certified and trained staff (16).

Psoriasis area and severity index. The PASI (17) is a physician-assessed index that measures psoriasis severity and evaluates erythema, infiltration, and desquamation (scaling) on different body areas including the head, upper extremities, the trunk, and lower extremities. The PASI provides a clinical summary of psoriasis disease activity, as well as a means to assess treatment efficacy in psoriasis.

Static physician's global assessment. The sPGA (19) is a physician-rated assessment of psoriasis disease severity. The sPGA is a single-item question, answered on a six-point scale that measures the degree of overall psoriatic lesion severity. Scores range from 0 (clear or no evidence of plaque elevation, scaling, or erythema) to 5 (severe plaque elevation; severe, very thick scale predominates and; dusty to deep red erythema).

Body surface area. The BSA (20) is a physician-completed disease assessment that is a measure of the BSA affected by psoriasis. The BSA is a numerical score used to measure the physician's assessment of the proportion (or percentage) of the subject's total BSA involved with psoriasis.

Statistical methods

The analysis reported in this study included subjects who had both baseline and week 12 PSI, PASI, and DLQI data (n = 186). Data was pooled across treatment groups, with both baseline and week 12 PSI, PASI, sPGA, and DLQI data from the intent-to-treat population. The psychometric analysis used data collected during the initial 12 weeks of the study. All analyses were conducted with the SAS® system (version 9.1 for Windows), except that Mplus (21) was used for the confirmatory factor analysis (CFA), and MULTILOG (22) was used for the item response theory (IRT) analysis. The statisticians who conducted the analyses were blinded to treatment group. All statistical tests used a significance level of 0.05 (two-sided), unless otherwise noted. Statistical tests were adjusted for multiple comparisons as appropriate using the Sheffe method.

Descriptive statistics.
Sociodemographical and clinical characteristics for the study population were analyzed using descriptive statistics (mean, standard deviation, range for quantitative variables, frequency, and percentage for categorical variables).

Descriptive statistics (n, mean, median, standard deviation, range, percentage at floor, percentage at ceiling, and percentage missing) were also analyzed for the PSI total and item scores at baseline and week 12.

Confirmatory factor analysis.
A CFA was conducted to examine the factor structure of the PSI. The CFA was conducted using the baseline and week 12 PSI data. Based on the results of the exploratory factor analysis in a previous study (15), which found a single factor, we evaluated a model with the eight PSI items loading on a single factor. Overall model fit was assessed using the comparative fit index (CFI, ≥0.90), standardized root mean residual (SRMR) (values of <0.1 are considered acceptable (23)), and the root mean square error of approximation (RMSEA) (values <0.08 are considered acceptable (24)). The 90% CI for the RMSEA was used to give additional confidence in the estimate.

Item response theory analysis.
IRT analysis was used to examine whether each item exhibited psychometric properties with the following criteria: (1) item

Table I. Demographical and clinical characteristics of patients at baseline (n = 186).

Variable	Baseline
Age, years	
Mean (SD) [range]	42.9 (12.18) [21–70]
Gender	
n (%) female	67 (36.0)
Race, n(%)	
American Indian or Alaska native	3 (1.6)
Asian	5 (2.7)
Black or African-American	5 (2.7)
Hispanic or Latino	6 (3.2)
White or Caucasian	166 (89.3)
Other	1 (0.5)
sPGA	
Mean (SD) [range]	3.4 (0.63) [2–5]
PASI score	
Mean (SD) [range]	19.3 (6.79) [12–54]
BSA percentage	
Mean (SD) [range]	24.1 (14.17) [10–88]
Duration of psoriasis, years	
Mean (SD) [range]	19.2 (11.18) [1–52]
Concomitant psoriatic arthritis	
n (%), yes	43 (23.1)

response options are ordered; and (2) items form a unidimensional construct. Items that did not fit the IRT model were to be flagged as candidates for item reduction.

The Samejima's graded response model (25), which allows for analyses of measures with Likert-type scales was used for the IRT analyses. With this model, for each item, the item parameters are the slope parameter, a_i, and the m-1 category threshold parameters, b_{i2}, b_{i3},.b_{im}. These parameters describe the items in relationship with the underlying latent construct, that is, psoriasis symptom severity.

The PSI items were assessed for the model fit. An item characteristic curve plot was used to identify the items that have not demonstrated monotonically increased responses. In addition, the S-X2 fit statistics (26) were computed for each item to assess the model fit. The value of $p < 0.001$ was marked as a misfitting item.

Reliability.

Internal consistency of the PSI total score was assessed using Cronbach's coefficient alpha (27). Values >0.70 are generally considered indicative of a homogenous scale (28). Test-retest reliability was assessed by evaluating the reproducibility of PSI item and total scores among stable patients. Patients who had the same sPGA score at weeks 8 and 12, and those who had the same sPGA score at weeks 12 and 16, were defined as stable. These timepoints were selected based on availability of sPGA score and on the assumption that more patients would be stable following 8 weeks of treatment than prior to treatment. Intraclass correlation coefficients (ICC) were used to evaluate test-retest reliability. Values >0.70 are generally considered acceptable for establishing test-retest reliability (29,30).

Validity.

Validity of an instrument refers to the extent to which an instrument measures the construct it is intended to measure (28,29). To assess the validity of the PSI, construct (convergent, discriminant, and known groups) validity was examined.

Convergent and discriminant validity were evaluated by examining the magnitude of correlations between the PSI item and total scores and the DLQI item and domain scores and SF-36 subscale and component summary scores at baseline and week 12. Convergent validity is supported when the PSI item and total scores are substantially correlated (≥0.40) with items or instruments measuring similar concepts. We hypothesized that we would observe moderate-to-large correlations among the PSI item and total scores and DLQI item 1, SF-36 bodily pain and SF-36 social functioning, SF-36 vitality, DLQI item 2, and DLQI item 5 scores. Conversely, discriminant validity is supported when the PSI scores are not substantially correlated (<0.40) with scales measuring distally related concepts. We hypothesized correlations small in magnitude between the PSI and the SF-36 role emotional, role function, and mental health scales.

Known groups validity is the extent to which scores from an instrument are distinguishable among groups of patients that differ on a known relevant dimension, such as clinical status. To

Table II. Item descriptive characteristics – PSI.

	n	Mean (SD)	Median	Range	Floor (*n*,%)	Ceiling (*n*,%)	Missing (*n*,%)
Baseline							
Itch	186	2.6 (0.94)	3	0–4	3 (1.6%)	33 (17.7%)	0 (0.0%)
Redness	185	2.6 (0.88)	3	0–4	1 (0.5%)	30 (16.2%)	1 (0.5%)
Scaling	185	2.6 (0.91)	3	0–4	3 (1.6%)	32 (17.3%)	1 (0.5%)
Burning	183	1.9 (1.18)	2	0–4	23 (12.6%)	22 (12.0%)	3 (1.6%)
Stinging	186	1.9 (1.21)	2	0–4	27 (14.5%)	21 (11.3%)	0 (0.0%)
Cracking	186	2.2 (1.10)	2	0–4	11 (5.9%)	24 (12.9%)	0 (0.0%)
Flaking	184	2.7 (0.87)	3	1–4	0 (0.0%)	32 (17.4%)	2 (1.1%)
Pain	185	2.0 (1.19)	2	0–4	24 (13.0%)	24 (13.0%)	1 (0.5%)
Total	178	18.3 (6.96)	18	3–32	0 (0.0%)	8 (4.5%)	8 (4.3%)
Week 12							
Itch	185	1.0 (1.14)	1	0–4	79 (42.7%)	7 (3.8%)	1 (0.5%)
Redness	183	1.0 (1.13)	1	0–4	80 (43.7%)	8 (4.4%)	3 (1.6%)
Scaling	184	1.0 (1.16)	1	0–4	88 (47.8%)	8 (4.3%)	2 (1.1%)
Burning	183	0.6 (1.11)	0	0–4	124 (67.8%)	8 (4.4%)	3 (1.6%)
Stinging	184	0.7 (1.10)	0	0–4	120 (65.2%)	7 (3.8%)	2 (1.1%)
Cracking	184	0.7 (1.09)	0	0–4	113 (61.4%)	6 (3.3%)	2 (1.1%)
Flaking	184	0.9 (1.16)	0.5	0–4	92 (50.0%)	8 (4.3%)	2 (1.1%)
Pain	184	0.7 (1.09)	0	0–4	118 (64.1%)	9 (4.9%)	2 (1.1%)
Total	179	6.5 (8.26)	3	0–32	56 (31.3%)	4 (2.2%)	7 (3.8%)

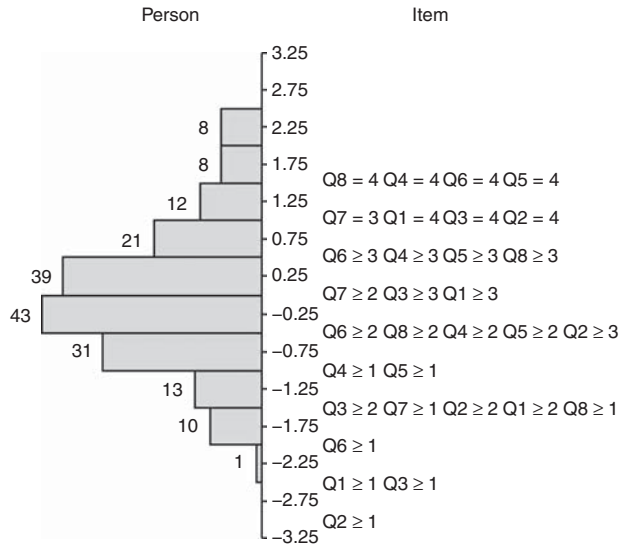

Figure 1. Item map for PSI.

evaluate known groups validity, the PSI total scores were examined using an analysis of variance (ANOVA) model at weeks 8 and 12. These timepoints were selected to ensure adequate variability in patient status, as baseline eligibility criteria resulted in limited variability in patient clinical status at baseline. It was anticipated that by week 8 of treatment, more patients would be categorized as having lower levels of symptom severity allowing for greater variability in clinical status. Groups for these ANOVAs were defined in two ways: based on PASI categories (<12, 12 to 18, >18) and on the sPGA score (0–2, 3, 4–5). *Post hoc* comparisons using Sheffe method were used to test for differences between the levels of each grouping variable.

Responsiveness.

Responsiveness refers to the extent to which the instrument can detect true change in patients known to have changed in clinical status (29).

Responsiveness or the ability of the PSI to detect change was determined using two clinical anchors: PASI improvement status and clinician-rated disease status as rated using the sPGA. PASI improvement categories were ≥75 (clinically meaningful improvement), 50–74 (some improvement), and <50 (none to limited improvement), and sPGA categories were 0–2 (none to mild), 3 (moderate), and 4–5 (marked to severe). An ANOVA model was used to examine the difference in the mean change scores from baseline to week 12 of the PSI total scores as the dependent variable and PASI or sPGA categories at week 12 as the independent variable. *Post hoc* comparisons using Sheffe method were used to test for differences between the levels of each grouping variable.

Results

Descriptive statistics of sociodemographical and clinical characteristics

A total of 186 patients were included in the psychometric evaluation of the PSI based on pooled data. Mean age was 42.9 years (SD = 12.2), with 36% female and 89% Caucasian (Table I). Mean duration of psoriasis was 19 years and 23% also had a diagnosis of psoriatic arthritis. The baseline mean sPGA score was 3.4 (SD = 0.6), baseline mean PASI score was 19.3 (SD = 6.8), and baseline mean BSA was 24.1% (SD = 14.2).

Descriptive statistics of PRO measures

Descriptive characteristics for the item and total scores of the PSI are presented in Table II. The baseline PSI total score was 18.3 (SD = 7.0). The PSI total scores decreased at the follow-up visits from baseline, again indicating reduced symptom severity across the study duration. Descriptive statistics for individual items of the PSI at baseline indicated no evidence of floor or ceiling effects, with the entire range of possible item responses observed. Very little missing data was observed (<1.6%).

The baseline DLQI total score was 11.8 (SD = 6.8), with baseline DLQI domain scores ranging from 0.7 (SD = 0.9; work and school) to 4.0 (SD = 1.5; symptoms and feeling). The baseline SF-36 subscale

Table III. Construct validity: PSI items and total score correlations with DLQI domain and total scores at baseline and week 12.

	DLQI domain						
	Symptoms and feelings	Daily activities	Leisure	Work and school	Personal relationships	Treatment	Total
Baseline							
Itch	0.64*	0.37*	0.33*	0.25†	0.39*	0.27†	0.51*
Redness	0.52*	0.31*	0.31*	0.16‡	0.37*	0.18‡	0.43*
Scaling	0.43*	0.27†	0.32*	0.23‡	0.30*	0.21‡	0.40*
Burning	0.48*	0.29*	0.31*	0.22‡	0.34*	0.22‡	0.43*
Stinging	0.48*	0.30*	0.33*	0.27†	0.36*	0.24‡	0.45*
Cracking	0.50*	0.33*	0.35*	0.18‡	0.39*	0.26†	0.47*
Flaking	0.59*	0.28†	0.28†	0.20‡	0.36*	0.27†	0.44*
Pain	0.55*	0.38*	0.40*	0.32*	0.47*	0.25†	0.54*
PSI total score	0.63*	0.38*	0.39*	0.27†	0.45*	0.28†	0.55*
Week 12							
Itch	0.86*	0.70*	0.59*	0.49*	0.47*	0.50*	0.76*
Redness	0.85*	0.74*	0.63*	0.54*	0.47*	0.52*	0.79*
Scaling	0.85*	0.71*	0.60*	0.55*	0.44*	0.54*	0.77*
Burning	0.79*	0.68*	0.61*	0.52*	0.42*	0.50*	0.74*
Stinging	0.80*	0.71*	0.60*	0.52*	0.41*	0.52*	0.74*
Cracking	0.79*	0.65*	0.61*	0.56*	0.42*	0.51*	0.73*
Flaking	0.82*	0.67*	0.62*	0.52*	0.41*	0.51*	0.74*
Pain	0.75*	0.66*	0.61*	0.57*	0.43*	0.48*	0.73*
PSI total score	0.88*	0.74*	0.68*	0.59*	0.49*	0.55*	0.82*

Pearson's correlation coefficients: *$p < 0.0001$; †$p < 0.001$; ‡$p < 0.05$.

Table IV. Construct validity: PSI total score correlations with SF-36 scale and component scores at baseline and week 12.

	SF-36 scale									
	Physical function	Role physical	Bodily pain	General health	Vitality	Social function	Role emotional	Mental health	SF-36 PCS	SF-36 MCS
Baseline										
PSI total score	−0.35*	−0.30*	−0.49*	−0.28†	−0.20‡	−0.39*	−0.20‡	−0.26†	−0.41*	−0.21‡
Week 12										
PSI total score	−0.36*	−0.43*	−0.49*	−0.27†	−0.33*	−0.50*	−0.38*	−0.32*	−0.41*	−0.37*

Pearson's correlation coefficients: *$p < 0.0001$; †$p < 0.001$; ‡$p < 0.05$.

scores ranged from 45.1 (SD = 12.3; social functioning) to 50.4 (SD = 10.0; vitality).

Confirmatory factor analysis

Examination of the model fit statistics indicated incomplete fit for the single factor model in the initial CFA analysis (CFI = 0.88). After inspecting detailed fit statistics, the source of this incomplete fit was identified, that is, high correlation between the burning and stinging items. After incorporating this correlation into the CFA, the CFI was 0.94, indicating very good fit for the single factor model. The SRMR in the final model was 0.04, and the RMSEA was 0.12, also indicating good fit for the single factor model. The factor loadings for the individual PSI items were all >0.75 (itch = 0.76, redness = 0.80, scaling = 0.80, burning = 0.83, stinging = 0.78, cracking = 0.82, flaking = 0.79, and pain = 0.80).

Item response theory analyses

The slope parameters from the IRT analyses for the PSI were all >2.45 (itch = 2.46, redness = 2.75, scaling = 2.50, burning = 3.97, stinging = 3.37, cracking = 2.56, flaking = 2.53, and pain = 3.02). The threshold values ranged from −2.54 to 1.45, indicating good coverage across the range of the psoriasis-related symptom domain (Figure 1). Based on the measurement error information, there is good assessment of psoriasis symptom severity across the range of symptom severity. The IRT analyses further support the unidimensionality of the PSI and also demonstrate that the item response scales are well ordered. The items within the PSI cover a broad range of the psoriasis symptom severity domain.

Reliability

Internal consistency was 0.93 to 0.98 for the PSI total scores at baseline, and at weeks 4 and 12. At baseline, the correlation between individual items and the PSI total scores was 0.79–0.89 (Pearson correlation).

Test-retest reliability in stable patients ranged from 0.87 (weeks 12–16) to 0.91 (weeks 8–12) for the PSI total scores. For the PSI individual item scores, ICCs ranged from 0.77 to 0.87.

Validity

Concurrent and discriminate validity was tested based on correlations with the DLQI and the SF-36. Table III presents the PSI item and total score correlations with DLQI domain and total scores at baseline and week 12. PSI items were moderately correlated with the DLQI total score (r_p = 0.40–0.54) and most of the domain scores (e.g., symptoms and feelings, personal relationships, daily activities, and leisure) at baseline. Small-to-moderate correlations were found between the PSI items and the DLQI domain scores in work and school and treatment. All item level correlations were found to be higher in magnitude at week 12. The PSI total score moderately correlated with the DLQI domain scores (r_p = 0.27–0.63) and the DLQI total score (r_p = 0.55 at baseline and r_p = 0.82 at week 12; see Table III).

Table IV summarizes the correlations of the PSI total scores with the SF-36 domain and summary scores. For PSI total scores, the highest correlations at baseline were seen for bodily pain (−0.49), social function (−0.39), and physical function scores (−0.35). The correlation between PCS and PSI total score was −0.41. As expected, small-to-moderate correlations were found among the PSI and role emotional (−0.20), role physical (−0.30), and mental health (−0.26) SF-36 scores. Similar patterns were observed at the item level. These correlations maintained this general pattern although the magnitude of the correlations was increased at week 12 (Table IV).

Known groups validity was evaluated by comparing mean PSI item and total scores by PASI groups (<12, 12 to 18, >18), at weeks 8 and 12. Table V presents the mean PSI total scores by PASI severity groups. The PSI total scores were varied significantly by PASI severity group (all $p = < 0.001$), with greater mean PSI scores observed in greater PASI severity groups. Similar results were found at week 8, with the exception of nonsignificant comparisons between the PASI >18 versus PASI 12–18 groups.

Table VI shows the mean PSI item and total scores by sPGA-defined groups (0–2, 3, 4, 5) at weeks 8 and 12. PSI total scores varied significantly by sPGA severity group (all $p = < 0.001$), with greater mean PSI scores observed in "more severe" rated groups.

Responsiveness

The responsiveness of the PSI was evaluated by examining baseline to week 12 changes in PSI scores by specified categories of improvement in the PASI and sPGA. Based on the PASI groups (PASI ≥75, 50–74, and <50), PSI total change scores were significantly different ($p < 0.0001$; Figure 2), with the greatest

Table V. Known groups validity of the PSI total scores at weeks 8 and 12 by PASI groups.

	PASI group							
	<12		12–18		>18			
PSI item and total score	n	Mean (SD)	n	Mean (SD)	n	Mean (SD)	Overall F* (p value)	Pairwise comparisons†
Week 8								
PSI total score	139	3.6 (4.51)	22	15.7 (7.87)	19	18.6 (7.58)	99.30 (<0.0001)	1: <0.0001, 2: <0.0001, 3: 0.2321
Week 12								
PSI total score	138	3.4 (4.85)	19	13.5 (7.16)	22	20.4 (8.40)	102.03 (<0.0001)	1: <0.0001, 2: <0.0001, 3: 0.0006

*ANOVA model includes PASI tertile groups as the independent variable and PSI item and total scores as the dependent variable.
†Pairwise comparisons between means were performed using Scheffe's test adjusting for multiple comparisons.
Comparisons are 1 = < 12 versus 12–18, 2 = <12 versus >18, 3 = 12–18 versus >18.

Table VI. Known groups validity of the PSI total scores at weeks 8 and 12 by sPGA groups.

	sPGA group							
	0–2		3		4–5			
PSI item and total score	n	Mean (SD)	n	Mean (SD)	n	Mean (SD)	Overall F* (p value)	Pairwise comparisons†
Week 8								
PSI total score	127	3.2 (4.61)	38	12.9 (6.37)	15	20.2 (8.30)	99.26 (<0.0001)	1: <0.0001, 2: <0.0001, 3: <0.0001
Week 12								
PSI total score	127	3.2 (5.01)	31	10.8 (6.84)	21	20.8 (8.15)	93.35 (<0.0001)	1: <0.0001, 2: <0.0001, 3: <0.0001

*ANOVA model includes sPGA groups as the independent variable and PSI item and total scores as the dependent variable.
†Pairwise comparisons between means were performed using Scheffe's test adjusting for multiple comparisons.
Comparisons are 1 = 0–2 versus 3, 2 = 0–2 versus 4–5, 3 = 3 versus 4–5.

improvements for the PASI ≥75 and PASI 50–74 groups compared to the PASI <50 ($p = < 0.01$). Based on the sPGA groups (0–2, 3, 4–5), there were significant differences in mean PSI total scores among the sPGA groups ($p < 0.0001$; Figure 3), with the greatest improvements for the patients in the 0–2 category as compared to both the 4–5 category ($p < 0.0001$) and 3 category ($p < 0.0001$).

Discussion

The PSI is a recently developed measure of patient-reported symptom severity for patients with moderate-to-severe psoriasis. Although preliminary evidence supporting the content validity and reliability of the PSI was found in a previous study (14,15), additional evidence was needed to further evaluate the psychometric characteristics and to assess the responsiveness of the PSI. This psychometric analysis is based on data collected during a Phase II clinical trial. Based on the current study, the PSI has good evidence supporting unidimensionality, excellent internal consistency and test-retest reliability, good evidence supporting concurrent and known groups validity, and evidence of responsiveness to changes in clinical status.

Previous exploratory factor analyses of the PSI items indicated a single factor representing psoriasis-related symptoms (15); therefore, we fit a single factor model to the PSI using baseline data. After incorporating a correlation between the burning and stinging items into the model, excellent fit was observed with factor loadings all exceeding 0.7. Based on the CFA, there was strong evidence supporting the unidimensionality of the PSI.

IRT analyses were used to evaluate item characteristics and performance in the PSI. The IRT findings for the PSI demonstrated good coverage across the range of psoriasis symptom severity. The response categories were also well ordered and there was no evidence of misfitting items in the PSI. The IRT analyses further supported good performance of the PSI items and demonstrated good precision (reliability) across a range of the psoriasis symptom severity continuum.

Internal consistency was excellent for the PSI with Cronbach alpha coefficients >0.90 for all assessment visits. Test-retest reliability was also excellent, exceeding 0.85 for each assessment. Individual PSI items also demonstrated good test-retest reliability, with ICCs exceeding 0.75 for each item. ICCs >0.70 indicate good evidence of stability in PSI scores (29). Although high inter-item correlations were found, each item was considered clinically relevant in the assessment of psoriasis symptoms based on clinician input.

Concurrent validity was evaluated based on correlations between the PSI and DLQI and SF-36. Moderate-to-strong associations seen between PSI total scores and DLQI total and domain scores provided good evidence for convergent validity. As hypothesized, constructs related to psoriasis symptoms (SF-36 bodily pain and DLQI symptom item) and their impact on function (SF-36 DLQI social activity impact item) were moderately correlated. In support of discriminant validity, low correlations were found between the PSI total scores and distally related concepts, including the role emotional scale and role physical and mental health scales of the SF-36. The results indicate that increases in symptom severity were associated with greater impairment of HRQOL, particularly in the physical and social functioning areas of HRQOL.

PSI total scores varied significantly by clinical indicators of psoriasis severity. PSI total scores differentiated groups based on the PASI and sPGA scores, with mean scores significantly larger in those patients rated as having more severe psoriasis by clinicians. PSI total scores also demonstrated higher mean scores for patients in groups rated by clinicians as having severe symptoms.

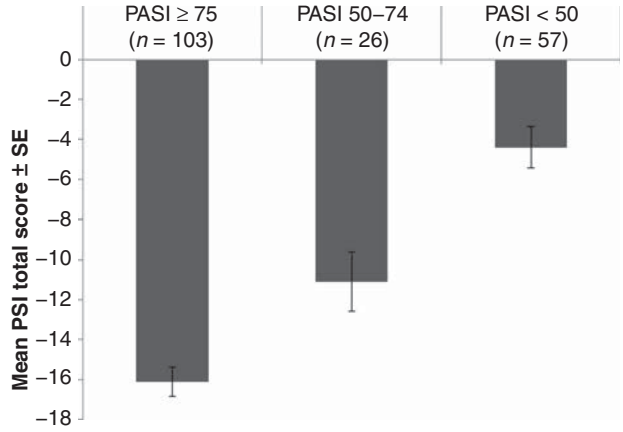

Figure 2. Total mean PSI change scores by PASI improvement status at week 12. Mean PSI total change scores were significantly different among PASI groups ($p < 0.001$) with greater change scores by greater PASI improvement categories (all $p < 0.01$).

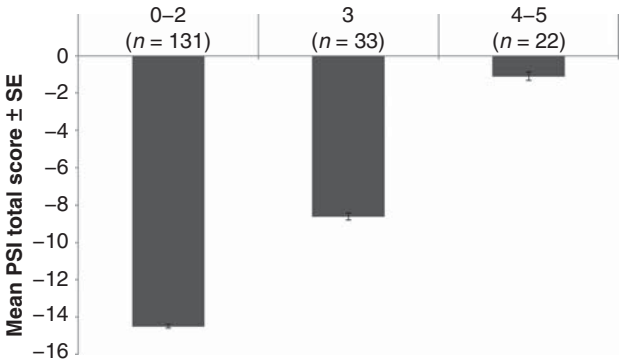

Figure 3. Mean PSI total change scores were significantly different among sPGA groups (0–2, 3, 4–5). PSI improvement scores were significantly greater for patients rated better clinical status by clinician (all $p < 0.01$).

Taken together with the concurrent validity and known groups validity findings, there is good supportive evidence for the construct validity of the PSI in patients with moderate-to-severe psoriasis.

For PRO endpoints in clinical trials to be most useful, it is necessary to demonstrate the ability to detect changes in clinical status. Statistically significant differences were found in mean PSI change scores among groups of subjects with different levels of improvement in the PASI as well as sPGA. These results support the ability of the PSI to detect change in clinical status with treatment. Further research is needed to evaluate the interpretability and clinical meaningfulness of specific responder thresholds based on the PSI total score.

There are certain limitations to this psychometric analysis. First, the data were taken from a Phase II clinical trial, and study inclusion and exclusion criteria and demographic profile of patients willing to participate in a clinical trial may impact the generalizability of the sample. Second, some of the analyses on responsiveness were based on a relatively smaller number of subjects in sPGA and PASI improvement categories. Additional studies can provide further evidence of the ability of the PSI to detect change in psoriasis symptom severity. However, the PSI was correlated with the clinician-reported measures and also demonstrated responsiveness to change over time in the PASI and sPGAs, providing evidence of construct validity.

Based on the current psychometric analyses, the PSI demonstrated unidimensionality, excellent internal consistency and test-retest reliability, concurrent and known groups validity, and evidence of responsiveness to changes in clinical status. Given the evidence supporting the measurement properties of the PSI total scores, especially responsiveness to changes in clinical status, the PSI can be used as a short and valid measure of psoriasis symptom severity in clinical trials evaluating the efficacy of new therapeutic agents for moderate-to-severe psoriasis.

Acknowledgments

The authors would like to acknowledge Elizabeth Merikle for conducting some early analyses. The authors would also like to acknowledge Wen-Hung Chen and Ray Hsieh for their statistical assistance, and Emily Sargent and Joanne Li for their comments on an earlier draft.

Declaration of interest: Dennis Revicki and Hilary Wilson work for United Biosource Corp., who received funding from Amgen for this study. Ying Jin also worked at United Biosource Corp. during the development of the publication. Dina Chau and Hema Viswanathan are employees of Amgen, Inc. and own stocks in Amgen, Inc.

References

1. Koo J. Population-based epidemiologic study of psoriasis with emphasis on quality of life assessment. Dermatol Clin. 1996;14:485–496.
2. Stern RS, Nijsten T, Feldman SR, Margolis DJ, Rolstad T. Psoriasis is common, carries a substantial burden even when not extensive, and is associated with widespread treatment dissatisfaction. J Investig Dermatol Symp Proc. 2004;9:136–139.
3. de Arruda LH, De Moraes AP. The impact of psoriasis on quality of life. Br J Dermatol. 2001;144:33–36.
4. Heydendael VM, de Borgie CA, Spuls PI, Bossuyt PM, Bos JD, de Rie MA. The burden of psoriasis is not determined by disease severity only. J Investig Dermatol Symp Proc. 2004;9:131–135.
5. Rapp SR, Cottrell CA, Leary MR. Social coping strategies associated with quality of life decrements among psoriasis patients. Br J Dermatol. 2001;145:610–616.
6. Revicki DA, Willian MK, Menter A, Saurat JH, Harnam N, Kaul M. Relationship between clinical response to therapy and health--related quality of life outcomes in patients with moderate to severe plaque psoriasis. Dermatology. 2008;216:260–270.
7. Shikiar R, Heffernan M, Langley RG, Willian MK, Okun MM, Revicki DA. Adalimumab treatment is associated with improvement in health-related quality of life in psoriasis: patient-reported outcomes from a phase II randomized controlled trial. J Dermatolog Treat. 2007; 18:25–31.
8. Shikiar R, Willian MK, Okun MM, Thompson CS, Revicki DA. The validity and responsiveness of three quality of life measures in the assessment of psoriasis patients: results of a phase II study. Health Qual Life Outcomes. 2006;4:71.
9. Weiss SC, Kimball AB, Liewehr DJ, Blauvelt A, Turner ML, Emanuel EJ. Quantifying the harmful effect of psoriasis on health-related quality of life. J Am Acad Dermatol. 2002;47:512–518.
10. Finlay AY, Khan GK. Dermatology Life Quality Index (DLQI)–a simple practical measure for routine clinical use. Clin Exp Dermatol. 1994;19:210–216.
11. Kimel M, Bell JA, Sexton C, Revicki DA. HRQL outcomes for assessing effectiveness in clinical trials of biologic agents for psoriasis: a systematic review. Adv Psor Inflam Skin Dis. 2009;1:7–14.
12. Ware JE Jr, Sherbourne CD. The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. Med Care. 1992;30:473–483.
13. Food and Drug Administration. Guidance for industry on patient-reported outcome measures: use in medical product development to support labeling claims. Fed Regist. 2009;74:65132–65133.
14. Martin ML, McCarrier KP, Chiou DF, Gordon K, Kimball AB, Van Voorhees AS, et al. Development of a new patient reported measure for assessing symptoms of psoriasis. Abstract presented at EADV. Lisbon, Portugal. October 20–24, 2011a.
15. Martin ML, Mccarrier KP, Bushnell DM, Gordon K, Chiou CF, Huang X, et al. Validation of the Psoriasis Symptom Inventory (PSI), a patent-reported outcome measure. Abstract presented at EADV. Lisbon, Portugal. October 20–24, 2011b.
16. Papp KA, Leonardi C, Menter A, Ortonne JP, Krueger JG, Kricorian G, et al. Brodalumab, an anti-interleukin-17-receptor antibody for psoriasis. N Engl J Med. 2012;366:1181–1189.
17. Fredriksson T, Pettersson U. Severe psoriasis–oral therapy with a new retinoid. Dermatologica. 1978;157:238–244.
18. Ware JE, Kosinski M, Dewey JE. How to score version two of the SF-36 health survey. Lincoln, RI: QualityMetric, Inc, 2000.
19. Bonifati C, Berardesca E. Clinical outcome measures of psoriasis. Reumatismo. 2007;59:64–67.
20. Ramsay B, Lawrence CM. Measurement of involved surface area in patients with psoriasis. Br J Dermatol. 1991;124:565–570.
21. Muthén LK, Muthén B. Mplus user's guide. Los Angeles: Muthén & Muthén, 2007.
22. Thissen D, Chen WH, Bock D. MULTILOG user's guide. Lincolnwood, IL: Scientific Software International, 2002.
23. Kline R. Principles and pracies of structural equation modeling. 2nd ed. New York: Guilford Press, 2005.
24. Browne MW, Cudeck R. Alternative ways of assessing model fit. In Bollen KA, Long JS, editors. Testing structural equation models. Beverly Hills, CA: Sage, 1993. p 136–162.
25. Samejima F. Estimation of latent trait ability using a response pattern of graded scores. Psychomet Monogr. 1969;34:whole No. 17.
26. Orlando M, Thissen D. Likelihood-based item-fit indices for dichotomous item response theory models. Appl Psychol Meas. 2000;24:50–64.
27. Cronbach LJ. Coefficient alpha and the internal structure of tests. Psychometrika. 1951;16:297–334.
28. Nunnally JC, Bernstein IH. Psychometric theory. 3rd ed. New York: McGraw Hill, 1994.
29. Hays R, Revicki DA. Reliability and validity (including responsiveness). In Fayers P, Hays R, editors. Assessing quality of life in clinical trials. New York: Oxford University Press, 2005. p 25–39.
30. Scientific Advisory Committee of the Medical Outcomes Trust. Assessing health status and quality-of-life instruments: attributes and review criteria. Qual Life Res. 2002;11:193–205.