



Psychometric Validation of the Psoriasis Symptoms and Impacts Measure (P-SIM), a Novel Patient-Reported Outcome Instrument for Patients with Plaque Psoriasis, Using Data from the BE VIVID and BE READY Phase 3 Trials

Richard B. Warren · Alice B. Gottlieb · Joseph F. Merola · Llenalia Garcia ·
Christopher Cioffi · Luke Peterson · Christopher Pelligra · Valerie Ciaravino

Received: May 27, 2021 / Accepted: June 16, 2021 / Published online: July 14, 2021
© The Author(s) 2021

ABSTRACT

Introduction: Plaque psoriasis can significantly impact patients' quality of life. We assessed psychometric properties of the Psoriasis Symptoms and Impacts Measure (P-SIM), developed

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s13555-021-00570-4>.

R. B. Warren (✉)
Dermatology Centre, Manchester NIHR Biomedical Research Centre, Salford Royal NHS Foundation Trust, The University of Manchester, Stott Lane, Salford M6 8HD, Greater Manchester, UK
e-mail: Richard.Warren@manchester.ac.uk

A. B. Gottlieb
Icahn School of Medicine at Mount Sinai, New York, NY, USA

J. F. Merola
Harvard Medical School, Brigham and Women's Hospital, Boston, MA, USA

L. Garcia
UCB Pharma, Slough, UK

C. Cioffi · L. Peterson
UCB Pharma, Raleigh, NC, USA

C. Pelligra
Evidera, Medellín, Colombia

V. Ciaravino
UCB Pharma, Colombes, France

to capture patients' experiences of signs, symptoms and impacts of psoriasis.

Methods: Pooled, blinded, 16-week data from 1002 patients in the BE VIVID and BE READY bimekizumab phase 3 trials were analysed. The suitability of the P-SIM missing score rule (weekly scores considered missing if ≥ 4 daily scores were missing) was assessed. Test–retest reliability was evaluated using intraclass correlation coefficients (ICCs). Convergent validity was assessed between P-SIM and relevant patient-reported outcome (PRO) (Dermatology Life Quality Index [DLQI], DLQI item 1 [skin symptoms], Patient Global Assessment of Psoriasis) and clinician-reported outcome (ClinRO) scores (Psoriasis Area and Severity Index [PASI], Investigator's Global Assessment [IGA]) at baseline and week 16. Known-groups validity was assessed, comparing P-SIM scores between patient subgroups predefined using PASI/IGA scores. Sensitivity to change over 16 weeks was evaluated; responder definition (RD) thresholds were explored.

Results: The missing score rule used did not impact P-SIM scores. Test–retest reliability analyses demonstrated excellent score reproducibility (ICC 0.91–0.98). Inter-item correlations at baseline and week 16 were strong (> 0.5), apart from “choice of clothing” with “skin pain” and “burning” at baseline (both 0.49). All P-SIM scores were moderately to

strongly correlated with other outcomes, demonstrating convergent validity, apart from ClinROs (PASI, IGA) at baseline that had low variability. P-SIM scores discriminated known groups at week 16, confirming known-groups validity. Changes from baseline to week 16 in P-SIM and other clinically relevant outcomes were strongly correlated (> 0.5 ; weaker with ClinROs), establishing sensitivity to change. Anchor-based RD analyses determined a four-point P-SIM item score decrease as indicative of marked clinically meaningful improvement.

Conclusion: P-SIM scores demonstrated good reliability, validity and sensitivity to change. A four-point RD threshold could be used to assess 16-week treatment effects.

Trial Registration: BE VIVID: NCT03370133; BE READY: NCT03410992.

Keywords: Bimekizumab; Patient-reported outcome; Plaque psoriasis; Psychometric validation; Responder definition

Key Summary Points

Why carry out this study?

The clinical manifestations of plaque psoriasis have been shown to severely impact patients' quality of life and emotional well-being; therefore, it is important to measure patient experience alongside clinical parameters in the evaluation of treatments for plaque psoriasis.

The Psoriasis Symptoms and Impacts Measure (P-SIM) is a novel, patient-reported outcome tool developed to specifically capture patients' experiences of the key signs, symptoms and impacts of plaque psoriasis; here, it was completed daily by patients at home on a handheld device on a 0–10 numeric rating scale, and item scores were derived as averages of daily scores over a week.

In these analyses, the psychometric properties of the P-SIM were evaluated using data pooled from the first 16 weeks of the BE VIVID and BE READY phase 3 trials investigating bimekizumab in the treatment of moderate to severe plaque psoriasis.

What was learned from this study?

In these analyses, the 14 item scores of the P-SIM demonstrated good reliability, validity and sensitivity to change in the assessment of patients' experiences of key psoriasis signs, symptoms and impacts when completed daily on a handheld device over a period of 16 weeks.

Anchor-based analyses determined a 4-point decrease in item scores of the P-SIM as indicative of a marked clinically meaningful improvement when the P-SIM was completed daily on a handheld device; this threshold can be used to assess treatment effects over 16 weeks in patients with moderate to severe plaque psoriasis.

INTRODUCTION

Psoriasis is an immune-mediated, chronic, inflammatory skin disease that can take a variety of forms [1]. The symptoms most frequently experienced by patients with psoriasis include itching and scaling [2], and the clinical manifestations of psoriasis can have wide-ranging impacts on quality of life and emotional well-being [3–5]; therefore, it is important to measure patient experience alongside clinical parameters. Understanding patients' experiences of psoriasis and needs due to their disease is key to supporting their treatment.

Patient-reported outcome measures have been developed to assess quality of life in patients with skin conditions. For example, the Dermatology Life Quality Index (DLQI) is widely used to assess quality of life changes in dermatology. However, it is not specific to

psoriasis, so may not fully capture the experiences of patients with this disease [6, 7]. Psoriasis-specific tools have also been developed for use in specific clinical trials, such as the 16-item Psoriasis Symptom Diary [8, 9], the Psoriasis Symptoms Scale, which captures severity of pain, redness, itching and burning [10], and the single-item Itch Numeric Rating Scale [11], among others [12–14]. These instruments assess patient-reported signs and symptoms of psoriasis. However, they are not available for use in the clinical evaluation of new treatments, and some of these instruments may omit relevant signs, symptoms and impacts.

The Psoriasis Symptoms and Impacts Measure (P-SIM) was developed to capture patients' experiences of the signs, symptoms and impacts of psoriasis in the bimekizumab in plaque psoriasis clinical programme. It was designed as a 14-item daily electronic diary that can be completed at home by the patient for ease of use. The development and validity of the P-SIM content have been described previously, including a literature review investigating commonly reported signs, symptoms and impacts of plaque psoriasis, clinical expert interviews and patient interviews for the purposes of concept elicitation, cognitive debriefing and usability testing [15]. Quantitative analyses following these steps led to the finalisation of the 14 items of the P-SIM [15]. In addition to content validity, the US Food and Drug Administration (FDA) also recommends assessment of other psychometric properties of newly developed patient-reported outcome measures (reliability, construct validity, and ability to detect change) [16], and determination of thresholds for interpreting the clinical meaningfulness of within-patient score changes over a predetermined time period (a responder definition [RD]) [16]. To date, no RD thresholds have been reported for the P-SIM.

Here, the psychometric properties of the P-SIM are assessed using data from the BE VIVID (NCT03370133) and BE READY (NCT03410992) phase 3 trials, as well as the suitability of the weekly missing data rule that is currently applied to the instrument. RD thresholds for the P-SIM items are also estimated.

METHODS

Study Design

Blinded data were pooled from the first 16 weeks of the BE VIVID and BE READY phase 3 trials investigating bimekizumab in the treatment of moderate to severe plaque psoriasis.

BE VIVID was a randomised, double-blinded, placebo- and active comparator-controlled phase 3 study (Supplementary Fig. 1a). Patients in BE VIVID were randomised 4:2:1 to bimekizumab 320 mg every 4 weeks (Q4W), ustekinumab 45/90 mg (by weight) administered at baseline and then every 12 weeks (Q12W) from week 4, or placebo for initial treatment (weeks 0–16). BE READY was a double-blinded, placebo-controlled, randomised withdrawal phase 3 study (Supplementary Fig. 1b). Patients in BE READY were randomised 4:1 to bimekizumab 320 mg Q4W or placebo for initial treatment (weeks 0–16). Both trials enrolled adult patients with a diagnosis of moderate to severe plaque psoriasis ≥ 6 months prior to screening, with baseline Psoriasis Area and Severity Index (PASI) ≥ 12 (on a scale from 0–72), $\geq 10\%$ body surface area (BSA) affected and an Investigator's Global Assessment (IGA) score ≥ 3 on a 5-point scale. The co-primary endpoints for both studies were 90% improvement in PASI (PASI 90) and IGA 0/1 (0 [clear] or 1 [almost clear] with ≥ 2 -category improvement from baseline) responses at week 16. Full study designs and efficacy and safety outcomes have been reported previously [17, 18].

The study protocols, amendments and patient informed consent were reviewed by a national, regional or independent ethics committee or institutional review board. BE VIVID (NCT03370133) and BE READY (NCT03410992) were conducted in accordance with the current version of the applicable regulatory and International Conference on Harmonisation (ICH)-Good Clinical Practice (GCP) requirements, the ethical principles that have their origin in the principles of the Declaration of Helsinki, and the local laws of the countries involved.

Table 1 Items of the P-SIM

Item	Response scale
1. During the past 24 h, how severe was your worst skin <i>itching</i> ?	0 (No itching)–10 (Very severe itching)
2. During the past 24 h, how severe was your worst skin <i>redness</i> ?	0 (No skin redness)–10 (Very severe skin redness)
3. During the past 24 h, how severe was your worst skin <i>pain</i> ?	0 (No skin pain)–10 (Very severe skin pain)
4. During the past 24 h, how severe was your worst skin <i>burning</i> ?	0 (No skin burning)–10 (Very severe skin burning)
5. During the past 24 h, how severe was your worst skin <i>scaling</i> ?	0 (No skin scaling)–10 (Very severe skin scaling)
6. During the past 24 h, how severe was your worst skin <i>cracking</i> ?	0 (No skin cracking)–10 (Very severe skin cracking)
7. During the past 24 h, how severe was your worst skin <i>dryness</i> ?	0 (No skin dryness)–10 (Very severe skin dryness)
8. During the past 24 h, how severe was your worst skin <i>irritation</i> ?	0 (No skin irritation)–10 (Very severe skin irritation)
9. During the past 24 h, how severe was your worst skin <i>sensitivity</i> ?	0 (No skin sensitivity)–10 (Very severe skin sensitivity)
10. During the past 24 h, how severe was your worst skin <i>lesions</i> (red marks, spots, or inflammation)?	0 (No skin lesions)–10 (Very severe skin lesions)
11. During the past 24 h, how severe was your worst skin <i>thickening</i> (hardening or roughening)?	0 (No skin thickening)–10 (Very severe skin thickening)
12. During the past 24 h, at its worst, how severe was your psoriasis-related <i>fatigue</i> (weariness, tiredness)?	0 (No fatigue)–10 (Worst possible fatigue)
13. During the past 24 h, at its worst, how much <i>embarrassment</i> did you feel as a result of your psoriasis?	0 (No feelings of embarrassment)–10 (Worst possible feelings of embarrassment)
14. During the past 24 h, at its worst, how much did your psoriasis impact your <i>choice of clothing</i> ?	0 (No impact on my choice of clothing)–10 (Completely impacted my choice of clothing)

Measurements and Outcomes

The P-SIM e-diary was to be completed daily in the evening by patients at home via a handheld electronic device from baseline through week 16. It consisted of 14 items, listed in Table 1. Each item was scored for worst severity or level of impact in the previous 24 h on a scale from 0 to 10 (0 meaning “no sign, symptom or impact”, 10 meaning “very severe sign, symptom or impact”). Average weekly scores were derived for each of the 14 P-SIM items. Weekly scores were considered missing if ≥ 4 daily scores were missing for that week.

Other outcomes reported in BE VIVID and BE READY relevant to this validation analysis

include PASI, IGA, DLQI, and Patient Global Assessment of Psoriasis (PGAP). The DLQI consists of ten items, each scored from 0 to 3, with 3 representing the highest impact (overall score range 0–30) [7]. DLQI item 1 (Question: “over the last week, how itchy, sore, painful, or stinging has your skin been?”; answer: “not at all”, “a little”, “a lot”, or “very much”) is specific to skin symptoms commonly seen in psoriasis and is considered individually in these analyses. The PGAP consists of a multiple-choice question (“How severe are your psoriasis-related symptoms right now?”), with answers scored from 1 to 5 (1, “no symptoms”; 2, “mild symptoms”; 3, “moderate symptoms”; 4, “severe symptoms”; 5, “very severe symptoms”). Both

the DLQI item 1 and PGAP represent verbal rating scales. Details of PASI and IGA score measurements can be found in the Supplementary Methods [19, 20]. The PGAP was completed on the same electronic device as the P-SIM; all other outcome data were collected on a tablet during on-site study visits (week 0, 1, 2, 4, 8, 12 and 16).

Confirmation of Weekly Scoring Rule

Analyses were conducted to establish whether alternatives to the currently applied weekly scoring rule for the P-SIM, in which weekly item scores are considered missing for patients in weeks where they had ≥ 4 missing daily scores in that item, impacted the variability of weekly item scores.

Patients with no missing data were included for analysis of each item. Scenarios for the number of days missing (1, 2, 3, 4, 5 or 6 days missing) were simulated using a bootstrapping method with replacement by randomly sampling the appropriate number of daily scores for that missing day scenario and calculating average weekly scores. For each missing day scenario for a given item, each patient's simulated weekly score was the mean of 100 replications of a random selection of daily scores. Means and standard deviations (SDs) were calculated for each item for each patient for the non-missing case and each missing day scenario. Overall means and SDs were also calculated for each item by pooling the weekly scores of patients. The SDs for missing and non-missing scenarios for each item were compared by visual inspection and the Brown–Forsythe test.

Psychometric Validation

Construct Validity

Convergent validity was assessed for P-SIM items by calculating Spearman's rank correlation coefficients at baseline and week 16 for inter-item correlations and correlations between P-SIM item scores and clinician-reported outcome (PASI total score and IGA score) and patient-reported outcome (DLQI total score, DLQI item 1 score and PGAP score)

scores. Only pairs assessed on the same date were included in these analyses.

A correlation coefficient > 0.3 and ≤ 0.5 indicated moderate convergent validity, and > 0.5 indicated strong convergent validity [21]. It was hypothesised that P-SIM items would have strong correlations with each other, and moderate to strong correlations with PASI, IGA, DLQI and PGAP total scores. Items 1, 3, 4 and 8 (itching, skin pain, burning and irritation, respectively) of the P-SIM were expected to have strong correlations with DLQI item 1.

The ability of P-SIM items to discriminate between clinically different groups of patients, predefined according to relevant clinician-reported outcomes, was also assessed. Mean P-SIM item scores were assessed at week 16 in known subgroups of patients defined on the basis of absolute PASI total score thresholds (≤ 1 , > 1 to ≤ 3 , > 3 to < 5 , ≥ 5 to < 12 , and ≥ 12) and IGA scores (0, "clear"; 1, "almost clear"; 2, "mild"; 3, "moderate"; and 4, "severe") at week 16. PASI and IGA are well-accepted clinical measures of psoriasis disease severity and were used to define the primary efficacy endpoints in the bimekizumab phase 3 studies, BE VIVID and BE READY. As expected, as a result of the inclusion criteria of the BE VIVID and BE READY trials (PASI ≥ 12 , IGA ≥ 3), there was low variability in scores for both measures at baseline; assessment of known-groups validity therefore focused on week 16. The absolute PASI values 1, 3 and 5 used to define subgroups have been shown to provide reliable estimates of disease activity that can be used to define treatment goals for psoriasis treatment and facilitate clinical decisions [22], while an absolute PASI value of 12 was used to define moderate to severe plaque psoriasis in the BE VIVID and BE READY inclusion criteria. The IGA values used to define subgroups with clinically different psoriasis severity levels are the response options of the 5-point IGA scale, which has been shown to be a valid and reliable measure of psoriasis severity [20].

It was hypothesised that P-SIM item score means and SDs would be higher in patients with higher PASI and IGA scores. *P* values from the Kruskal–Wallis test were calculated to compare distributions among the known groups.

Test–Retest Reliability

Intraclass correlation coefficients (ICCs; calculated as the ratio of the between-patient variance to the total variance) were calculated for each item's score between baseline and week 2 in stable patients during this period. Stable patients were defined as those whose IGA score did not change in this interval. ICCs ≥ 0.70 were considered acceptable for test–retest reliability [23].

Sensitivity to Change Over Time

Spearman's correlation coefficients and *p* values were calculated between changes from baseline to week 16 in P-SIM item scores and in other clinically relevant outcomes (PASI, IGA, PGAP, DLQI item 1 and DLQI). Correlations using the change from baseline to week 12 for PGAP were assessed as a sensitivity analysis; this was because a large amount of PGAP data were missing at week 16 as a result of technical limitations of the electronic device. A Spearman correlation ≥ 0.30 was considered to demonstrate acceptable responsiveness [24].

Determination of Responder Definition

Anchor-based analyses were performed in line with the recommendations for defining response thresholds for within-patient meaningful changes provided in the FDA patient-reported outcome guidance. Distribution-based analyses were conducted to provide supportive information around the variability of the scores, as per FDA guidance [16]. Triangulation was performed, examining the estimates from the anchor-based analyses and considering the distribution-based analyses as supportive information, to estimate values for clinically meaningful changes in each P-SIM item score.

For the anchor-based analyses, only anchors with Spearman correlation ≥ 0.30 between changes from baseline to week 16 in the anchor (week 12 for PGAP) and P-SIM items were considered. The anchors used, based on PASI, IGA, DLQI, DLQI item 1 and PGAP response, are shown in Supplementary Table 1. As a result of small sample sizes ($n < 15$) in some anchor

response categories, some of the adjacent original categories were collapsed.

Actual mean change scores from baseline to week 16 were calculated for each level of improvement in each anchor. The empirical cumulative distribution function (eCDF) and probability density function (PDF) curves of observed changes at week 16 for each P-SIM item were plotted separately for each responder group for each anchor. These curves were examined to see if separation between levels of response on anchors could be observed. If so, actual change scores that optimally differentiated responders from non-responders were identified to support determination of the RD for each item. Effect sizes for each level of change in the anchors were estimated, and 95% confidence intervals for the observed mean change were calculated using bootstrapping methods.

Distribution-based analyses used the standard error of measurement for P-SIM items, and half an SD of item scores at baseline and study visits [25, 26]. If the SD changed significantly over time, the baseline SD was used; otherwise the mean was used.

Items 1 (itching), 3 (skin pain) and 5 (scaling) of the P-SIM were prioritised in the RD analyses owing to their clinical relevance, importance to patient experience, and use as efficacy endpoints in the BE VIVID and BE READY trials.

RESULTS

Patient Disposition and Baseline Characteristics

In BE VIVID, 567 patients were randomised at baseline: 321 to bimekizumab, 163 to ustekinumab and 83 to placebo. In BE READY, 435 patients were randomised at baseline: 349 to bimekizumab and 86 to placebo. Blinded 16-week data were pooled for the 1002 patients randomised across both studies.

Demographics and baseline disease characteristics are shown in Table 2. No major differences were noted between studies, except for race/region; BE VIVID included patients from

Table 2 Demographics and baseline characteristics

Characteristic	BE VIVID (<i>n</i> = 567)	BE READY (<i>n</i> = 435)	Pooled (<i>n</i> = 1002)
Age, years, mean (SD)	46.1 (13.9)	44.3 (12.9)	45.3 (13.5)
Gender, <i>n</i> (%)			
Male	406 (71.6%)	313 (72.0%)	719 (71.8%)
Female	161 (28.4%)	122 (28.0%)	283 (28.2%)
Race, <i>n</i> (%)			
White	420 (74.1%)	403 (92.6%)	823 (82.1%)
Black or African American	12 (2.1%)	6 (1.4%)	18 (1.8%)
Asian	127 (22.4%)	18 (4.1%)	145 (14.5%)
American Indian or Alaska native	2 (0.4%)	0 (0.0%)	2 (0.2%)
Native Hawaiian or other Pacific islander	0 (0.0%)	2 (0.5%)	2 (0.2%)
Other or mixed	6 (1.1%)	6 (1.4%)	12 (1.2%)
Country, <i>n</i> (%)			
USA	116 (20.5%)	85 (19.5%)	201 (20.1%)
Canada	61 (10.8%)	89 (20.5%)	150 (15.0%)
Belgium	6 (1.1%)	0 (0.0%)	6 (0.6%)
Germany	62 (10.9%)	38 (8.7%)	100 (10.0%)
Italy	4 (0.7%)	0 (0.0%)	4 (0.4%)
UK	7 (1.2%)	6 (1.4%)	13 (1.3%)
Hungary	27 (4.8%)	31 (7.1%)	58 (5.8%)
Poland	143 (25.2%)	150 (34.5%)	293 (29.2%)
Russian Federation	19 (3.4%)	19 (4.4%)	38 (3.8%)
Australia	14 (2.5%)	10 (2.3%)	24 (2.4%)
Japan	108 (19.0%)	0 (0.0%)	108 (10.8%)
Korea	0 (0.0%)	7 (1.6%)	7 (0.7%)
Duration of disease, years, mean (SD)	17.1 (12.0)	19.5 (13.1)	18.1 (12.6)
Baseline PASI, <i>n</i> (%)			
PASI < 20	326 (57.5%)	274 (63.0%)	600 (59.9%)
PASI ≥ 20	240 (42.3%)	161 (37.0%)	401 (40.0%)
Missing	1 (0.2%)	0 (0.0%)	1 (0.1%)
BSA affected by psoriasis, %, mean (SD)	28.2 (16.9)	24.5 (15.4)	26.6 (16.3)
mNAPSI total score, mean (SD)	12.8 (18.8)	12.3 (19.4)	12.6 (19.0)

Table 2 continued

Characteristic	BE VIVID (<i>n</i> = 567)	BE READY (<i>n</i> = 435)	Pooled (<i>n</i> = 1002)
PGADA for arthritis VAS Score, mean (SD)	20.1 (27.5)	24.7 (27.7)	22.1 (27.6)
Nail involvement, <i>n</i> (%)			
Yes	354 (62.4%)	260 (59.8%)	614 (61.3%)
No	213 (37.6%)	175 (40.2%)	388 (38.7%)
mNAPSI total score for patients with nail involvement			
<i>n</i>	354	260	614
Mean (SD)	20.3 (20.2)	20.6 (21.4)	20.4 (20.7)
Scalp involvement, <i>n</i> (%)			
Yes	530 (93.5%)	397 (91.3%)	927 (92.5%)
No	37 (6.5%)	38 (8.7%)	75 (7.5%)
Palmoplantar involvement, <i>n</i> (%)			
Yes	227 (40.0%)	161 (37.0%)	388 (38.7%)
No	340 (60.0%)	274 (63.0%)	614 (61.3%)

BSA body surface area affected, *mNAPSI* modified Nail Psoriasis Severity Index, *PASI* Psoriasis Area and Severity Index, *PGADA* Patient Global Assessment of Disease Activity, *SD* standard deviation, *VAS* visual analogue scale

Japan (19.0%), while BE READY did not. The majority of patients were white (82.1%) and male (71.8%), and 40.0% of patients had a baseline PASI \geq 20, with the mean BSA affected by psoriasis being 26.6%.

P-SIM Completion Rates

Completion rate (number of patients with a non-missing weekly item score divided by the number of patients randomised at baseline) for the P-SIM ranged from 89.0% at week 1 to 77.8% at week 16. There were no item-level missing P-SIM data; patients either completed all items or no items.

Confirmation of Weekly Scoring Rule

There were 821 patients (81.9%) who had at least one week with zero daily scores missing; these patients were included in these analyses. For all items and for all missing day scenarios, the SDs found through simulations were not

statistically different to those from the ‘zero days missing’ scenario, and the mean item scores were similar (Supplementary Table 2). This suggests that P-SIM item score mean and variability remain similar regardless of the missing score rule applied, confirming that the weekly scoring rule used here (in which weekly scores were considered missing if \geq 4 daily scores were missing for that week) is acceptable.

Psychometric Validation

Construct Validity

As hypothesised, P-SIM items demonstrated strong inter-item convergent validity (Supplementary Table 3): almost all P-SIM items were strongly correlated (Spearman’s correlation coefficient, $r > 0.5$) with each other at both baseline and week 16, with the exception of item 14 (choice of clothing) with items 3 (skin pain) and 4 (burning) at baseline ($r = 0.49$).

As expected, P-SIM items 1, 3, 4 and 8 (itching, skin pain, burning and irritation,

respectively) were strongly correlated with DLQI item 1 score at both time points (Supplementary Table 4). All P-SIM items were moderately to strongly correlated with DLQI total score, DLQI item 1 score and PGAP score at both baseline and week 16 (week 12 scores were used for PGAP because of substantial amounts of missing data at week 16). Lower correlations were observed between P-SIM outcomes and the clinician-reported outcomes, PASI total score and IGA score, at baseline. This can be explained by the low variability in PASI total score and IGA at baseline, and was expected owing to the inclusion criteria of both studies ($\text{PASI} \geq 12$ and $\text{IGA} \geq 3$). Strong correlations with clinician-reported outcomes were, however, observed at week 16 across nearly all items (with the exception of items 12 [fatigue] and 14 [choice of clothing], for which correlations were moderate, ranging from 0.46 to 0.50 at week 16).

The P-SIM behaved as expected across all items when discriminating between known groups of patients defined on the basis of clinically relevant outcomes, with higher P-SIM item scores seen in patient known groups with higher PASI total score and IGA score (Fig. 1 and Supplementary Fig. 2) at week 16.

Sensitivity to Change Over Time

Changes from baseline to week 16 in all P-SIM item scores had strong Spearman correlation coefficients ($r > 0.50$) with those in the other patient-reported outcomes (DLQI total score, DLQI item 1 and PGAP [change from baseline to week 12 was used for PGAP as a result of substantial missing data at week 16]), with the exception of P-SIM item 14 with DLQI item 1 ($r = 0.50$) (Table 3). Moderate correlations were seen between changes from baseline to week 16 in P-SIM item scores and the clinician-reported outcomes, PASI and IGA ($r \geq 0.38$ for all items). Generally, changes from baseline in P-SIM items 12, 13 and 14 (fatigue, embarrassment and choice of clothing) had lower correlations with those in other outcome scores; however, changes from baseline in DLQI total score had the highest correlations with changes from baseline in P-SIM items 13 and 14).

Test–Retest Reliability

ICCs for all P-SIM items were well above the acceptability threshold of 0.70, ranging from 0.91 (item 5—scaling) to 0.98 (item 12—fatigue, item 13—embarrassment and item 14—choice of clothing) (Table 4). This indicates excellent test–retest reliability of P-SIM items.

Determination of Responder Definition

Supplementary Table 1 shows the response categorisations used for each anchor. The mean P-SIM item score changes obtained for one level of improvement in the anchors were considered as appropriate RD thresholds. The eCDF curves for one level of improvement consistently separated from those in the no meaningful change groups. Collapsing of improvement categories for the PASI percentage change from baseline, PASI total score and IGA change from baseline anchors, as a result of small sample sizes in certain categories, mechanically increased the observed mean changes from baseline to week 16 in P-SIM item scores, making the thresholds obtained from these anchors more conservative.

Values obtained for one level of improvement in DLQI item 1 were prioritised over other anchors. This was because its change from baseline to week 16 generally had the highest correlation with those of P-SIM items, it is patient-reported, it is measured on an ordinal scale with directly interpretable verbal descriptors, and eCDF and PDF findings supported the thresholds obtained using it. Results obtained using anchors based on PASI total score, IGA change from baseline and DLQI total score were also considered in the triangulation process, though were deprioritised as they did not meet all of the above criteria. IGA absolute score and PGAP change from baseline-derived thresholds were considered as supportive only (IGA had only one improvement level, and PGAP had a substantial amount of missing data at week 16 leading to week 12 being used).

The P-SIM item RD thresholds found using mean score changes ranged from -2.72 (item 12—fatigue) to -4.39 (item 5—scaling) (Supplementary Table 5). Distribution-based

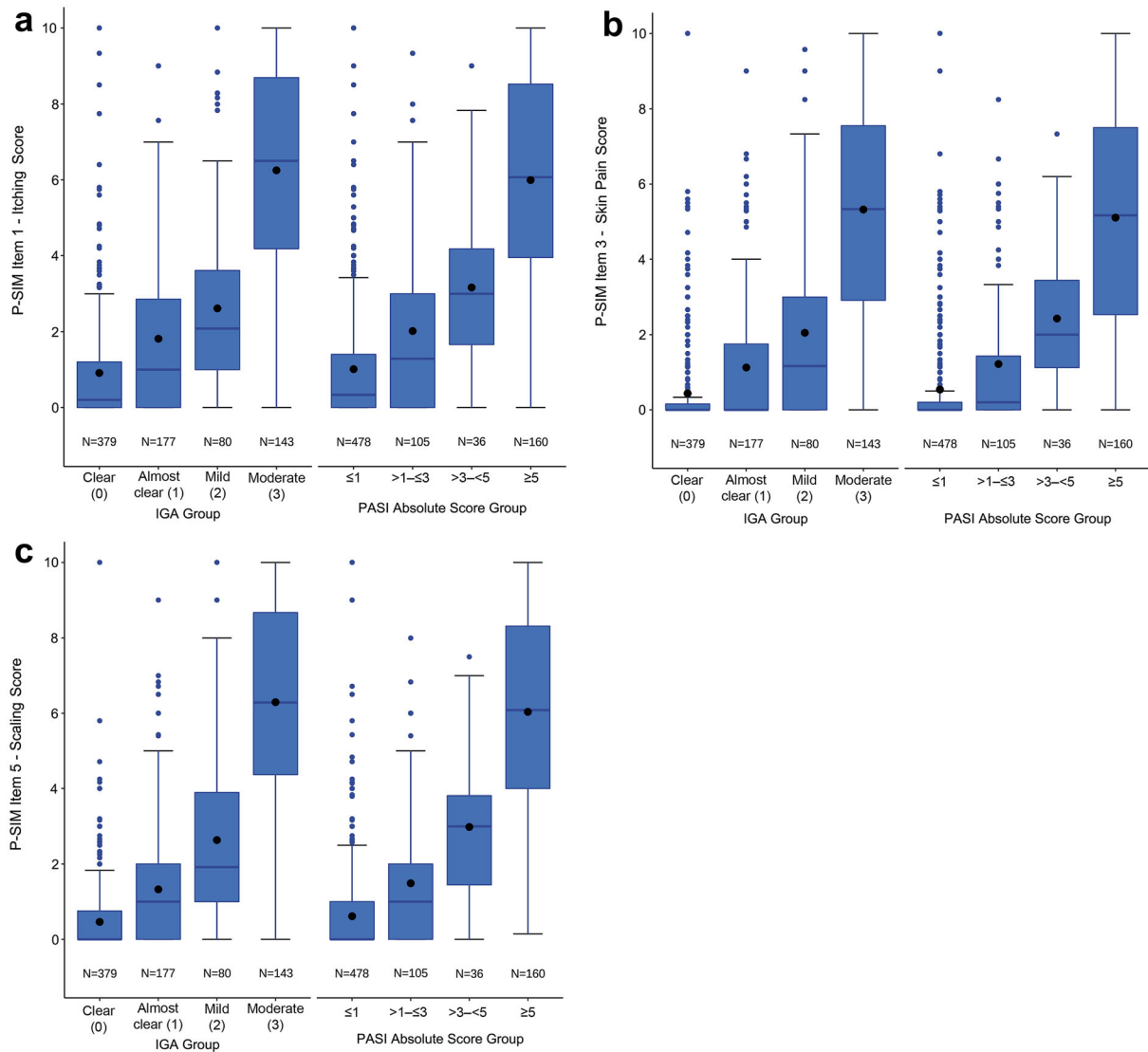


Fig. 1 P-SIM items 1, 3 and 5 scores at week 16 by IGA score and PASI total score subgroups. **a** P-SIM item 1 (itching). **b** P-SIM item 3 (skin pain). **c** P-SIM item 5 (scaling). Data were pooled from BE VIVID and BE READY. Black circular markers indicate mean scores

and blue circular markers indicate outliers. IGA Investigator's Global Assessment, PASI Psoriasis Area and Severity Index, P-SIM Psoriasis Symptoms and Impacts Measure

values were much smaller than anchor-based values, as expected; these were considered as supportive only. The mean change in P-SIM item scores obtained with one-point improvement on DLQI item 1, the prioritised anchor, were close to the four-point FDA-recommended threshold. A four-point threshold also discriminated between known groups based on other anchors and was thus considered appropriate

for the RD across all items. Inspection of eCDF and PDF curves supported this finding for all items (P-SIM items 1, 3 and 5 were of primary consideration; their CDF curves for DLQI item 1 and PASI are presented in Figs. 2 and 3). On the basis of these results, a four-point reduction on each P-SIM item is considered to represent a marked clinical improvement and can be used to define responders in patients with moderate

Table 3 Spearman correlations between changes from baseline to week 16 in P-SIM item scores and other clinician- and patient-reported outcomes

P-SIM item	Spearman correlation coefficient for changes from baseline to week 16				
	PASI total score (<i>n</i> = 698)	DLQI total score (<i>n</i> = 698)	DLQI item 1 score (<i>n</i> = 696)	IGA score (<i>n</i> = 698)	PGAP score ^a (<i>n</i> = 314)
Item 1—itching	0.42	0.61*	0.74*	0.48	0.72*
Item 2—redness	0.45	0.61*	0.68*	0.50	0.74*
Item 3—skin pain	0.44	0.60*	0.69*	0.46	0.68*
Item 4—burning	0.44	0.61*	0.70*	0.48	0.69*
Item 5—scaling	0.49	0.61*	0.66*	0.53*	0.73*
Item 6—cracking	0.43	0.60*	0.65*	0.47	0.68*
Item 7—dryness	0.44	0.60*	0.68*	0.49	0.75*
Item 8—irritation	0.44	0.63*	0.70*	0.49	0.73*
Item 9—sensitivity	0.44	0.61*	0.68*	0.48	0.70*
Item 10—lesions	0.45	0.62*	0.65*	0.49	0.72*
Item 11—thickening	0.44	0.61*	0.64*	0.49	0.71*
Item 12—fatigue	0.38	0.61*	0.54*	0.44	0.60*
Item 13—embarrassment	0.38	0.69*	0.56*	0.46	0.61*
Item 14—choice of clothing	0.38	0.68*	0.50	0.46	0.59*

Data were pooled from BE VIVID and BE READY

DLQI Dermatology Life Quality Index, *IGA* Investigator's Global Assessment, *PASI* Psoriasis Area and Severity Index, *PGAP* Patient Global Assessment of Psoriasis, *P-SIM* Psoriasis Symptoms and Impacts Measure

*Correlation coefficients that are strong (> 0.50)

^a Changes from baseline to week 12 were used for PGAP as a result of substantial amounts of missing data at week 16

to severe plaque psoriasis treated over a 16-week period.

DISCUSSION

This study aimed to evaluate the psychometric properties of the P-SIM to complement the previously published evidence for its content validity [15]. Specifically, comprehensive psychometric assessment of the P-SIM item scores, including convergent and known-groups validity, test-retest reliability and sensitivity to change over time, was complemented by assessment of the currently applied missing data rule and estimation of the RD threshold for P-SIM items, as recommended by FDA

guidelines [16]. The 14 items are not combined into an overall summary score; while this means that the P-SIM does not produce one number providing an overview of all patient symptoms, it does allow individual and specific capture of various independent signs, symptoms and impacts of plaque psoriasis. Other instruments that use fewer items are simpler to analyse, but capture fewer concepts [11, 13].

Blinded data were used for the psychometric validation and RD analyses, removing sources of potential bias. The applied weekly missing score rule, in which weekly item scores were set to 'missing' if ≥ 4 daily scores were missing for that item, was found to be adequate. The P-SIM item scores were shown to have excellent test-retest reliability, with all ICCs ≥ 0.91 , well

Table 4 Intraclass correlation coefficients for P-SIM item scores

P-SIM item	ICC ($n = 1002$)
Item 1—itching	0.94
Item 2—redness	0.93
Item 3—skin pain	0.95
Item 4—burning	0.94
Item 5—scaling	0.91
Item 6—cracking	0.93
Item 7—dryness	0.94
Item 8—irritation	0.94
Item 9—sensitivity	0.95
Item 10—lesions	0.94
Item 11—thickening	0.94
Item 12—fatigue	0.98
Item 13—embarrassment	0.98
Item 14—choice of clothing	0.98

ICC intraclass correlation coefficient, P-SIM Psoriasis Symptoms and Impacts Measure

above the acceptability threshold of 0.70; this demonstrates that the P-SIM is consistent in its measurement of signs, symptoms and impacts of plaque psoriasis in patients whose disease state has not changed, and that results are repeatable.

The P-SIM demonstrated strong inter-item correlations at both baseline and week 16, apart from item 14 (choice of clothing), which had a moderate correlation with items 3 (skin pain) and 4 (burning) at baseline. P-SIM item scores also showed moderate to strong correlations with PASI, IGA, DLQI, DLQI item 1 and PGAP scores at week 16, and with DLQI, DLQI item 1 and PGAP scores at baseline. The lower correlations seen with PASI and IGA scores at baseline may have been due to low variability in these measures upon trial entry, owing to the eligibility criteria for BE VIVID and BE READY (PASI ≥ 12 and IGA ≥ 3). Together, these results demonstrated good convergent validity of P-SIM items.

The P-SIM item scores also demonstrated the ability to discriminate between patient known groups defined on the basis of absolute PASI thresholds and IGA scores; statistically significant between-subgroup score differences were seen across all P-SIM items for nearly all known-group comparisons.

Changes from baseline to week 16 in P-SIM items were strongly correlated with those in other patient-reported outcomes and moderately correlated with those in clinician-reported outcomes. Previously it has been reported that patient-reported outcomes do not always correlate well with clinician-reported outcomes, highlighting the importance of including patient-reported outcomes in clinical studies to capture patient experiences [27, 28]. Changes from baseline in items 12 (fatigue), 13 (embarrassment) and 14 (choice of clothing) all generally had slightly weaker correlations with other outcomes, hypothesised to be due to the concepts measured being more distal and more difficult to assess compared with the more proximal skin symptoms assessed by items 1–11, which are also directly evaluated by clinician-reported outcomes and DLQI item 1. These results indicate that the P-SIM items are sensitive to changes in the concepts they are intended to measure.

RD thresholds were estimated, as recommended by the FDA guidance for industry patient-reported outcome measures [16]. This was of particular importance as responder proportions in P-SIM items 1 (itching), 3 (skin pain) and 5 (scaling) have been identified as key secondary endpoints in bimekizumab phase 3 studies. Anchor-based analyses were prioritised, as recommended by the FDA [16]; the RD thresholds obtained ranged from 2.72 to 4.39 (Supplementary Table 5) and supported the use of a four-point RD threshold to represent marked clinical improvement and to define responders in patients with moderate to severe plaque psoriasis in clinical trials. Though variations around this four-point RD threshold were seen, these were of limited amplitude. This is in line with FDA feedback and US prescribing information for compounds approved in psoriasis, which suggest that a four-point threshold

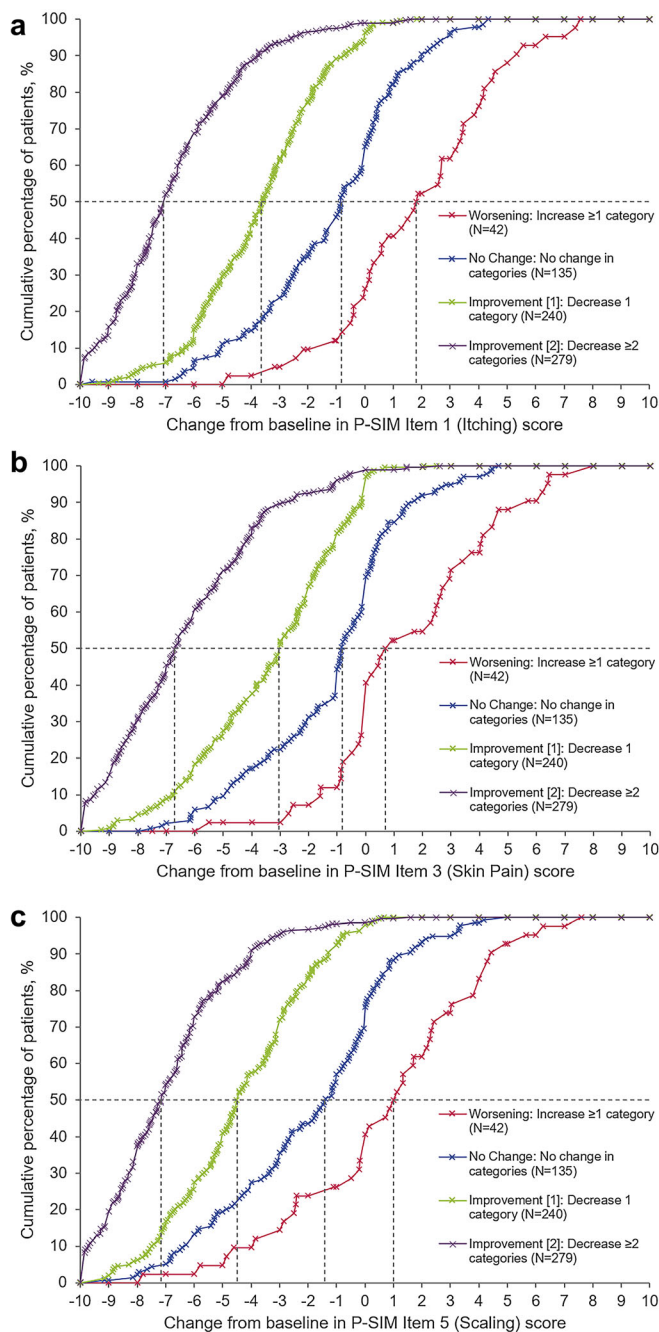
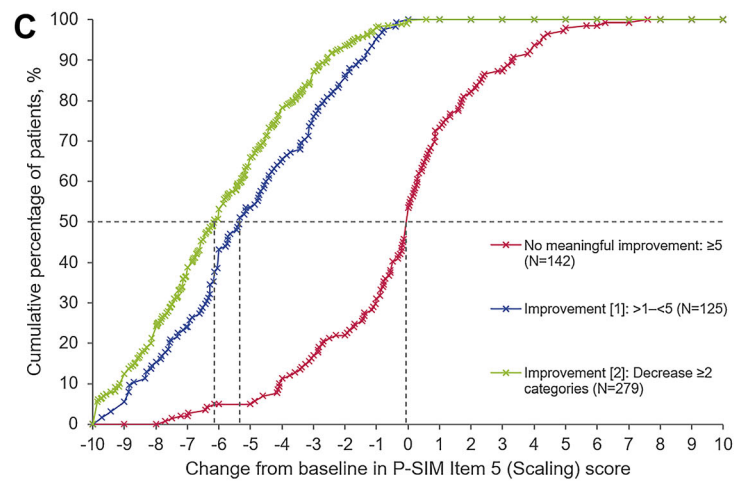
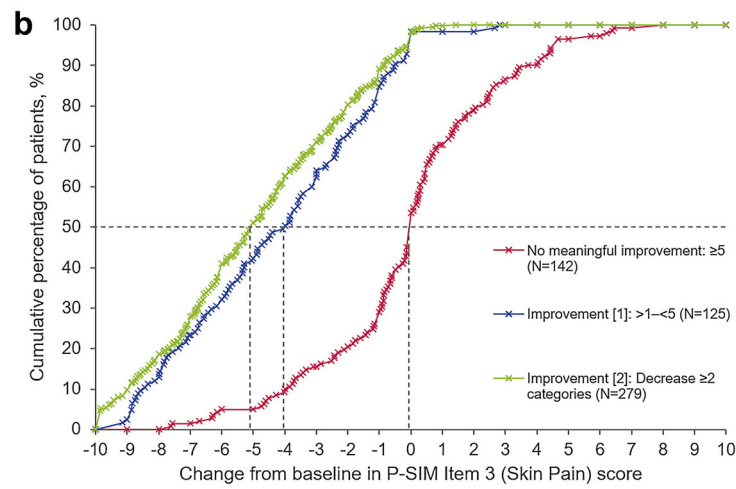
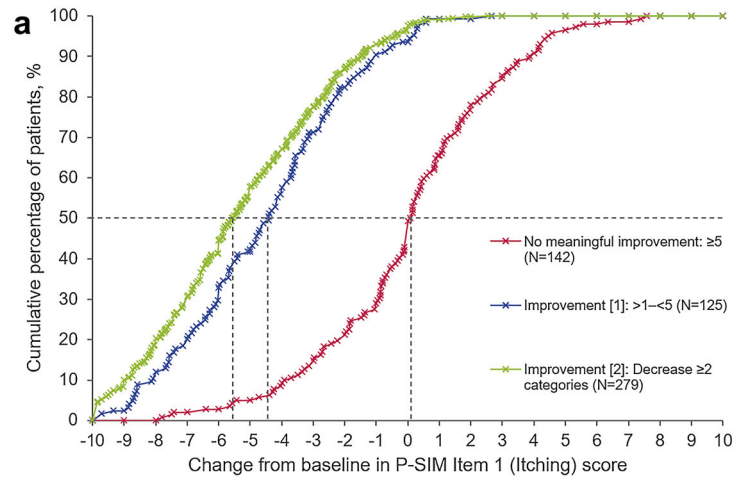


Fig. 2 eCDF curves of observed changes from baseline to week 16 in P-SIM item scores by DLQI item 1 change score category. **a** P-SIM item 1 (itching). **b** P-SIM item 3 (skin pain). **c** P-SIM item 5 (scaling). Data were pooled from BE VIVID and BE READY. DLQI item 1 was prioritised for the RD anchor-based analysis as it is patient-reported, measured on a directly interpretable ordinal scale, and its change from baseline to week 16 had the highest correlation with those in P-SIM

items. Negative changes from baseline in P-SIM item scores indicate improvement. Dotted lines show median values. eCDFs for itching, skin pain and scaling items are shown as these items were used as efficacy endpoints in the BE VIVID and BE READY trials; findings for other items were similar. eCDF empirical cumulative distribution function, DLQI Dermatology Life Quality Index, P-SIM Psoriasis Symptoms and Impacts Measure, RD responder definition



◀ **Fig. 3** eCDF curves of observed changes from baseline to week 16 in P-SIM item scores by PASI absolute score category. **a** P-SIM item 1 (itching). **b** P-SIM item 3 (skin pain). **c** P-SIM item 5 (scaling). Data were pooled from BE VIVID and BE READY. Negative changes from baseline in P-SIM item scores indicate improvement. Dotted lines show median values. eCDFs for itching, skin pain and scaling items are shown as these items were used as efficacy endpoints in the BE VIVID and BE READY trials; findings for other items were similar. eCDF empirical cumulative distribution function, PASI Psoriasis Area and Severity Index, P-SIM Psoriasis Symptoms and Impacts Measure

is a relevant RD for a single-item score assessed on an 11-point numeric rating scale [29, 30].

These results support the use of the P-SIM to define efficacy endpoints measuring patient experiences of signs, symptoms and impacts of psoriasis. The use of a four-point RD threshold over an 11-point numeric rating scale throughout all 14 items represents a simple and stringent criterion for defining marked clinical improvement; however, responder analyses using this threshold to assess treatment effect will be restricted to patients reporting a score of > 4 at baseline. Complementary analyses or endpoints, such as attaining a P-SIM score of 0 or 1, could be considered, as has been done for similar instruments in trials of other compounds approved in plaque psoriasis [31, 32]. This would allow all symptomatic patients to be taken into account.

Lower RD thresholds for items 1, 3 and 5 (itch, skin pain and scaling, respectively) were determined previously in preliminary phase 2b data analyses (1.98–2.86-point reduction), and were considered to represent meaningful improvement in bimekizumab in psoriasis phase 3 study protocols. The more stringent four-point RD thresholds identified here instead indicate marked clinical improvement.

Limitations of this study include the proportion of missing P-SIM data; 304/1002 of patients were excluded from the RD analysis because of missing data due to technical issues at baseline, week 16 or both. Although 30.3% of randomised patients were not included in the P-SIM RD estimation, the sample size included (698) is on par with or greater than sample sizes

of psychometric validation studies of other instruments in psoriasis [8, 33, 34]. It is also important to note that those patients who were excluded from the RD analysis had similar baseline characteristics to those who were included (with the exception of region, where a larger proportion excluded were from North America [49.3% of those excluded versus 28.8% of those included]; Supplementary Table 6).

Furthermore, while the daily nature of P-SIM completion helps to capture how patients' experiences may change from day to day, the regularity of assessment may affect overall completion rates; however, the simulation to confirm the weekly scoring rule demonstrated that the results do not vary substantially when the number of days included in the weekly score change, thus showing that the weekly scores are robust with respect to variation between days. Additionally, completion rates were high in this study even when using the 4-day missing weekly score rule.

Additionally, there were substantial amounts of missing data for the PGAP at week 16, limiting its use in analyses to a 12-week period instead. However, the RD analysis utilised several robust anchors (both patient- and clinician-reported), each yielding relatively similar results (predominantly between 3 and 5 points). CDF curves for each anchor showed separation between those who experienced an improvement and those who experienced no change at a four-point improvement in P-SIM item scores. Thus, it is likely that conclusions would be similar if there had not been missing data on the PGAP at week 16.

A large proportion of the 1002 patients enrolled in the phase 3 BE VIVID and BE READY trials were male (71.8%) and white (82.1%). It has been suggested that male patients generally have greater severity of psoriasis [35], leading to a greater proportion being included in the trials; however, another study has indicated that female patients experience greater impact on their quality of life [36], and the vast majority of clinical trials exclude pregnant women, meaning that they do not appear in true proportion [37]. Additionally, while the majority of patients in the general moderate to severe psoriasis disease population are white, this study

may not fully represent the experiences of patients with skin of colour, who may have increased impact of psoriasis on quality of life [38, 39]; for example, post-inflammatory hyperpigmentation has a greater impact on patients with skin of colour [40].

CONCLUSIONS

Here, the P-SIM weekly item scores demonstrated excellent test–retest reliability alongside good convergent validity, known-groups validity, and sensitivity to change in the concepts they are intended to measure. The weekly scoring rule applied was shown to be appropriate, and a four-point RD threshold was established for P-SIM item scores. This RD could be used to assess treatment effects on the signs, symptoms and impacts of psoriasis. Alongside previous studies confirming content validity of the P-SIM, this supports use of the instrument as an outcome measure in clinical trials in patients with psoriasis.

ACKNOWLEDGEMENTS

The authors thank the patients, the investigators and their teams who took part in these studies.

Funding. This study was sponsored by UCB Pharma, Brussels, Belgium. This article was based on the original studies BE VIVID (NCT03370133) and BE READY (NCT03410992), sponsored by UCB Pharma. The Rapid Service Fee was funded by UCB Pharma. RBW is supported by the Manchester NIHR Biomedical Research Centre.

Authorship. All named authors meet the International Committee of Medical Journal Editors (ICMJE) criteria for authorship of this article, take responsibility for the integrity of the work as a whole, and have given their approval for this version to be published.

Authorship Contributions. Substantial contributions to study conception and design:

RBW, ABG, JFM, LG, CC, LP, CP, VC; substantial contributions to analysis and interpretation of the data RBW, ABG, JFM, LG, CC, LP, CP, VC; drafting the article or revising it critically for important intellectual content: RBW, ABG, JFM, LG, CC, LP, CP, VC; final approval of the version of the article to be published: RBW, ABG, JFM, LG, CC, LP, CP, VC.

Medical Writing, Editorial and Other Assistance. The authors also acknowledge Susanne Wiegratz, UCB Pharma, Monheim, Germany for publication coordination, Daniel Smith, Costello Medical, Cambridge, UK, for medical writing and editorial assistance based on the authors' input and direction, and Shien Guo, Weiqin Liao and Xiaomei Ye, Evidera, Waltham, MA, USA, for their support in data analysis. Support for third-party writing assistance for this article was funded by UCB Pharma in accordance with Good Publication Practice (GPP3) guidelines (<http://www.ismpp.org/gpp3>).

Disclosures. Richard B. Warren: Consulting fees from AbbVie, Almirall, Amgen, Arena, Astellas, Avillion, Bristol Myers Squibb, Boehringer Ingelheim, Celgene, Eli Lilly, GSK, Janssen, LEO Pharma, Novartis, Pfizer, Sanofi, and UCB Pharma; research grants to his institution from AbbVie, Almirall, Janssen, LEO Pharma, Novartis, and UCB Pharma; honoraria from Astellas, DiCE, GSK, and Union. Alice B. Gottlieb: Received honoraria as an advisory board member and consultant for Anaptys Bio, Avotres Therapeutics, Beiersdorf, Boehringer Ingelheim, Bristol Myers Squibb, Eli Lilly, Incyte, Janssen, LEO Pharma, Novartis, Pfizer, Sun Pharma, UCB Pharma, and XBiotech (only stock options); research/educational grants (paid to Mount Sinai Medical School) from Boehringer Ingelheim, Incyte, Janssen, Novartis, Sun Pharma, UCB Pharma, and XBiotech. Joseph F. Merola: Has been a consultant for AbbVie, Amgen, Bayer, Biogen, Bristol Myers Squibb, Celgene, Eli Lilly, Janssen, Novartis, Pfizer, Sanofi-Regeneron, UCB Pharma; principal investigator for Dermavant, LEO Pharma, UCB Pharma. Llenalia Garcia, Valerie Ciaravino: Employees of UCB Pharma. Christopher Cioffi,

Luke Peterson: Employees and shareholders of UCB Pharma. Christopher Pelligra: Salaried employee of Evidera, which received funding from UCB Pharma to conduct the study.

Compliance with Ethics Guidelines. The study protocols, amendments, and patient informed consent were reviewed by a national, regional or independent ethics committee or institutional review board. BE VIVID (NCT03370133) and BE READY (NCT03410992) were conducted in accordance with the current version of the applicable regulatory and International Conference on Harmonisation (ICH)-Good Clinical Practice (GCP) requirements, the ethical principles that have their origin in the principles of the Declaration of Helsinki, and the local laws of the countries involved.

Data Availability. Data sharing from non-clinical studies is outside of UCB's data sharing policy. However, underlying data from the BE VIVID (NCT03370133) and BE READY (NCT03410992) phase 3 clinical trials reported in this manuscript may be requested by qualified researchers six months after product approval in the US and/or Europe, or global development is discontinued, and 18 months after trial completion. Investigators may request access to anonymised individual patient-level data and redacted trial documents which may include: analysis-ready datasets, study protocol, annotated case report form, statistical analysis plan, dataset specifications and clinical study report. Prior to use of the data, proposals need to be approved by an independent review panel at www.Vivli.org and a signed data sharing agreement will need to be executed. All documents are available in English only, for a pre-specified time, typically 12 months, on a password protected portal.

Open Access. This article is licensed under a Creative Commons Attribution-Non-Commercial 4.0 International License, which permits any non-commercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative

Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc/4.0/>.

REFERENCES

1. Boehncke WH, Schön MP. Psoriasis. *Lancet*. 2015;386:983–94.
2. Krueger G, Koo J, Lebwohl M, et al. The impact of psoriasis on quality of life: results of a 1998 National Psoriasis Foundation patient-membership survey. *Arch Dermatol*. 2001;137:280–4.
3. Armstrong AW, Schupp C, Wu J, et al. Quality of life and work productivity impairment among psoriasis patients: findings from the National Psoriasis Foundation survey data 2003–2011. *PLoS ONE*. 2012;7: e52935.
4. Kimball AB, Jacobson C, Weiss S, et al. The psychosocial burden of psoriasis. *Am J Clin Dermatol*. 2005;6:383–92.
5. Augustin M, Radtke MA. Quality of life in psoriasis patients. *Expert Rev Pharmacoecon Outcomes Res*. 2014;14:559–68.
6. Basra MK, Fenech R, Gatt RM, et al. The dermatology life quality index 1994–2007: a comprehensive review of validation data and clinical results. *Br J Dermatol*. 2008;159:997–1035.
7. Finlay AY, Khan GK. Dermatology life quality index (DLQI)—a simple practical measure for routine clinical use. *Clin Exp Dermatol*. 1994;19:210–6.
8. Strober B, Zhao Y, Tran MH, et al. Psychometric validation of the Psoriasis Symptom Diary using phase III study data from patients with chronic plaque psoriasis. *Int J Dermatol*. 2016;55:e147–55.
9. Strober BE, Nyirady J, Mallya UG, et al. Item-level psychometric properties for a new patient-reported psoriasis symptom diary. *Value in Health*. 2013;16: 1014–22.

10. Armstrong AW, Edson-Heredia E, Naegeli AN, et al. Development of the Psoriasis Symptoms Scale (PSS) in patients with moderate-to-severe psoriasis: qualitative and quantitative analyses. *J Dermatolog Treat.* 2020;31:452–9.
11. Kimball AB, Naegeli AN, Edson-Heredia E, et al. Psychometric properties of the Itch Numeric Rating Scale in patients with moderate-to-severe plaque psoriasis. *Br J Dermatol.* 2016;175:157–62.
12. Mathias SD, Feldman SR, Crosby RD, et al. Measurement properties of a patient-reported outcome measure assessing psoriasis severity: the psoriasis symptoms and signs diary. *J Dermatol Treat.* 2016;27:322–7.
13. Rentz AM, Skalicky AM, Burslem K, et al. The content validity of the PSS in patients with plaque psoriasis. *J Patient Rep Outcomes.* 2017;1:4.
14. Strober B, van de Kerkhof PCM, Callis Duffin K, et al. Feasibility and utility of the psoriasis symptom inventory (PSI) in clinical care settings: a study from the international psoriasis council. *Am J Clin Dermatol.* 2019;20:699–709.
15. Gottlieb AB, Ciaravino V, Cioffi C, et al. Development and content validation of the psoriasis symptoms and impacts measure (P-SIM) for assessment of plaque psoriasis. *Dermatol Ther (Heidelb).* 2020;10(6):1255–72.
16. United States Department of Health and Human Services, Food and Drug Administration, Center for Drug Evaluation and Research, Center for Biologics Evaluation and Research, Center for Devices and Radiological Health. Guidance for industry patient-reported outcome measures: use in medical product development to support labeling claims 2009. Rockville: FDA; 2009.
17. Reich K, Papp KA, Blauvelt A, et al. Bimekizumab versus ustekinumab for the treatment of moderate to severe plaque psoriasis (BE VIVID): efficacy and safety from a 52-week, multicentre, double-blind, active comparator and placebo controlled phase 3 trial. *Lancet.* 2021;397:487–98.
18. Gordon KB, Foley P, Krueger JG, et al. Bimekizumab efficacy and safety in moderate to severe plaque psoriasis (BE READY): a multicentre, double-blind, placebo-controlled, randomised withdrawal phase 3 trial. *Lancet.* 2021;397:475–86.
19. Feldman SR. A quantitative definition of severe psoriasis for use in clinical trials. *J Dermatol Treat.* 2004;15:27–9.
20. Langley RG, Feldman SR, Nyirady J, et al. The 5-point Investigator's Global Assessment (IGA) scale: a modified tool for evaluating plaque psoriasis severity in clinical trials. *J Dermatol Treat.* 2015;26:23–31.
21. Cohen J. *Statistical power analysis for the behavioral sciences.* 2nd ed. Hillsdale: Lawrence Erlbaum Associates; 1988. p. 13.
22. Gordon KB, Reich K, Crowley JJ, et al. Disease activity and treatment efficacy using patient-level Psoriasis Area and Severity Index scores from til-drakizumab phase 3 clinical trials. *J Dermatol Treat.* 2020. <https://doi.org/10.1080/09546634.2020.1747590>.
23. Cappelleri JC, Zou KH, Bushmakin AG, et al. *Patient-reported outcomes: measurement, implementation and interpretation.* Boca Raton: CRC; 2013.
24. Revicki D, Hays RD, Cella D, et al. Recommended methods for determining responsiveness and minimally important differences for patient-reported outcomes. *J Clin Epidemiol.* 2008;61:102–9.
25. Wyrwich, K. Session 5: What's the Score? Moving from Items to Scores – Methods, Considerations, and Case Examples. At the Eighth Annual Patient-Reported Outcome Consortium Workshop, April 26–27, 2017. Hyatt Regency Bethesda, 1 Bethesda Metro Center, Bethesda, MD 20814. <https://c-path.org/eighth-annual-patient-reported-outcome-consortium-workshop-2/>.
26. Norman GR, Streiner DL. *PDQ statistics.* PMPH-USA; 2003.
27. Lebwohl MG, Bachelez H, Barker J, et al. Patient perspectives in the management of psoriasis: results from the population-based Multinational Assessment of Psoriasis and Psoriatic Arthritis Survey. *J Am Acad Dermatol.* 2014;70:871–81 (e1–30).
28. Krueger GG, Feldman SR, Camisa C, et al. Two considerations for patients with psoriasis and their clinicians: what defines mild, moderate, and severe psoriasis? What constitutes a clinically significant improvement when treating psoriasis? *J Am Acad Dermatol.* 2000;43:281–5.
29. US FDA prescribing information for TALTZ (ixekizumab). Can be accessed at: https://www.accessdata.fda.gov/drugsatfda_docs/label/2020/125521s020lbl.pdf
30. US FDA MultiDisciplinary Review and Evaluation for TREMFYA (guselkumab). Can be accessed at: https://www.accessdata.fda.gov/drugsatfda_docs/nda/2017/761061Orig1s000MultidisciplineR.pdf
31. US FDA prescribing information for TREMFYA (guselkumab). Can be accessed at: https://www.accessdata.fda.gov/drugsatfda_docs/label/2017/761061Orig1s000MultidisciplineR.pdf

- accessdata.fda.gov/drugsatfda_docs/label/2020/761061s0071bl.pdf
32. US FDA prescribing information for SKYRIZI (risankizumab). Can be accessed at: https://www.accessdata.fda.gov/drugsatfda_docs/label/2019/761105s0001bl.pdf
 33. Viswanathan HN, Mutebi A, Milmont CE, et al. Measurement properties of the psoriasis symptom inventory electronic daily diary in patients with moderate to severe plaque psoriasis. *Value Health*. 2017;20:1174–9.
 34. Gilet H, Roborel de Climens A, Arnould B, et al. Development and psychometric validation of the reflective evaluation of psoriasis Efficacy of Treatment and Severity (REFLETS) questionnaire: a common measure of plaque-type psoriasis severity and treatment efficacy for patients and clinicians. *J Eur Acad Dermatol Venereol*. 2015;29:498–506.
 35. Hägg D, Sundström A, Eriksson M, et al. Severity of psoriasis differs between men and women: a study of the clinical outcome measure psoriasis area and severity index (PASI) in 5438 Swedish register patients. *Am J Clin Dermatol*. 2017;18:583–90.
 36. Napolitano M, Mastroeni S, Fania L, et al. Sex- and gender-associated clinical and psychosocial characteristics of patients with psoriasis. *Clin Exp Dermatol*. 2020;45:705–11.
 37. Shields KE, Lyerly AD. Exclusion of pregnant women from industry-sponsored clinical trials. *Obstet Gynecol*. 2013;122:1077–81.
 38. Kaufman BP, Alexis AF. Psoriasis in skin of color: insights into the epidemiology, clinical presentation, genetics, quality-of-life impact, and treatment of psoriasis in non-white racial/ethnic groups. *Am J Clin Dermatol*. 2018;19:405–23.
 39. Kerr GS, Qaiyumi S, Richards J, et al. Psoriasis and psoriatic arthritis in African-American patients—the need to measure disease burden. *Clin Rheumatol*. 2015;34:1753–9.
 40. Alexis AF, Blackcloud P. Psoriasis in skin of color: epidemiology, genetics, clinical presentation, and treatment nuances. *J Clin Aesthet Dermatol*. 2014;7:16–24.