



Psychometric Validation of the Psoriasis Symptoms and Impacts Measure (P-SIM): A Novel Patient-Reported Outcome Instrument for Patients with Plaque Psoriasis, Using Reported Data from the BE RADIANT Phase 3b Trial

Alice B. Gottlieb · Richard B. Warren · Matthias Augustin ·
Llenalia Garcia · Christopher Cioffi · Luke Peterson · Christopher Pelligra ·
Valerie Ciaravino
Received: May 27, 2021 / Accepted: June 18, 2021 / Published online: September 2, 2021
© The Author(s) 2021

ABSTRACT

Introduction: This analysis assessed the psychometric properties of the Psoriasis Symptoms and Impacts Measure (P-SIM), a novel patient-reported outcome (PRO) tool designed to

capture patient experiences of the signs, symptoms and impacts of psoriasis.

Methods: Blinded data from the BE RADIANT phase 3b trial of bimekizumab were analysed. In BE RADIANT, patients were randomised 1:1 to bimekizumab 320 mg every 4 weeks (Q4W) or secukinumab 300 mg (weekly until Week 4, then Q4W). Three items (itching, skin pain and scaling) of the P-SIM were electronically assessed throughout the trial and were scored from 0 to 10 (none to very severe signs/symptoms/impacts). Test–retest reliability was determined using intraclass correlations. Convergent validity was assessed between P-SIM and other relevant PRO and clinician-reported outcome (ClinRO) scores. Known-groups validity was assessed by comparing mean P-SIM item scores between patient subgroups based on the Psoriasis Area and Severity Index (PASI)/Investigator’s Global Assessment (IGA) scores. Responsiveness was assessed via correlations between changes from baseline in P-SIM item scores and other relevant PRO and ClinRO scores. Anchor-based responder analyses and empirical cumulative distribution function (eCDF) curves determined responder thresholds.

Results: The three P-SIM items yielded high intraclass coefficients (> 0.70). By Week 48, the three P-SIM items had moderate (> 0.30 and ≤ 0.50) to strong (> 0.50) correlations with other PROs and weaker correlations with ClinROs, demonstrating good convergent validity. For almost all known-group comparisons, statistically significant between-subgroup score

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s12325-021-01836-1>.

A. B. Gottlieb (✉)
Icahn School of Medicine at Mount Sinai, New York,
NY, USA
e-mail: alicegottliebderm@gmail.com

R. B. Warren
Dermatology Centre, Salford Royal NHS Foundation
Trust, Manchester NIHR Biomedical Research
Centre, The University of Manchester, Manchester,
UK

M. Augustin
Institute for Health Services Research in
Dermatology and Nursing (IVDP), University
Medical Center Hamburg-Eppendorf (UKE),
Hamburg, Germany

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

differences were seen across all three P-SIM items. Changes from baseline in the P-SIM and other relevant outcomes were above the acceptable threshold of ≤ 0.30 , demonstrating sensitivity to change. Anchor-based analyses determined a \geq four-point reduction from baseline to indicate marked clinically meaningful improvement for the P-SIM.

Conclusion: These results support the validity, reliability and sensitivity to change of the P-SIM in assessing key symptoms (itching, skin pain and scaling) in patients with moderate to severe plaque psoriasis.

Trial registration: NCT03536884.

Keywords: Bimekizumab; Patient-reported outcome; Psoriasis; Psychometric validation

Take-home message

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, three key items of the P-SIM (itching, skin pain and scaling) were completed at pre-defined patient visits using a hand-held device on a 0–10 numerical rating scale.

In this research, the psychometric properties of the P-SIM were evaluated using data from BE RADIANT, a phase 3b, multi-centre, randomised, secukinumab-controlled, double-blinded, parallel-group trial that investigated the efficacy and safety of bimekizumab in the treatment of moderate to severe plaque psoriasis.

What was learned from this study?

In these analyses, the itching, skin pain and scaling items 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 at pre-defined patient visits on a hand-held device over a period of 48 weeks.

Anchor-based analyses confirmed that a four-point decrease in item scores of the P-SIM was indicative of a marked clinically meaningful improvement when completed at pre-defined patient visits on a hand-held device; this threshold can be used to assess treatment effects over 48 weeks in patients with moderate to severe plaque psoriasis.

INTRODUCTION

Psoriasis is a chronic, immune-mediated disease which manifests in the skin or joints [1]. Plaque psoriasis makes up approximately 80–90% of all cases of psoriasis and typically presents as red, raised patches of skin with silvery-white scales and well-defined edges [2]. The clinical manifestations of plaque psoriasis have been shown to severely impact patients' quality of life and emotional well-being [3].

To date, several patient-reported outcome (PRO) measures have been developed to assess the symptoms and functional impacts of skin diseases, the most commonly used being the Dermatology Life Quality Index (DLQI) [4]. However, many PRO tools, such as DLQI, do not accurately capture the signs and symptoms of psoriasis reported by patients, due to: long recall periods, the exclusion of certain symptoms, lack of specificity and the lack of measurements to assess the severity of symptoms [5].

For these reasons, the Psoriasis Symptoms and Impacts Measure (P-SIM) was developed to specifically capture patients' experiences of key

signs and symptoms of plaque psoriasis. The P-SIM was designed as an electronic diary with a 24 h recall period. It consists of 14 items which aim to cover the breadth of psoriasis symptoms; these items were derived from literature searches, clinical expert opinion and interviews with patients [6]. Content validation of the P-SIM has previously been conducted alongside the BE ABLE 2 phase 2b study (NCT03010527) [6, 7]. Additionally, psychometric properties of the P-SIM have been evaluated based on blinded data from the BE VIVID (NCT03370133) and BE READY (NCT03410992) phase 3 trials of bimekizumab, in which daily assessments using the P-SIM were conducted [8].

Here, we evaluate the psychometric properties of the P-SIM in the context of BE RADIANT (NCT03536884), a phase 3b, multi-centre, randomised, secukinumab-controlled, double-blinded, parallel-group trial investigating the efficacy and safety of bimekizumab in the treatment of moderate to severe plaque psoriasis.

Unlike in the BE VIVID and BE READY clinical trials, in which the P-SIM was completed daily, in BE RADIANT, three of the 14 P-SIM items (itching, skin pain and scaling) were completed by patients on electronic tablets at pre-defined study visits. It is therefore necessary to further evaluate the instrument's properties in the context of BE RADIANT, to assess whether the three selected P-SIM item scores can reliably measure symptoms in patients with moderate to severe plaque psoriasis, and to assess the reliability of a single measurement rather than a weekly average. Responder definition (RD) thresholds for the P-SIM items are also estimated.

METHODS

Study Design

Analyses were conducted using blinded data from BE RADIANT, in which patients were randomised 1:1 to bimekizumab 320 mg every 4 weeks (Q4W) or secukinumab 300 mg Q4W (Fig. 1). At Week 16, patients receiving bimekizumab continued on bimekizumab Q4W or

received bimekizumab Q8W. Adults with moderate to severe plaque psoriasis were enrolled, with baseline Psoriasis Area and Severity Index (PASI) ≥ 12 , body surface area (BSA) affected by psoriasis $\geq 10\%$, and Investigator's Global Assessment (IGA) score ≥ 3 on a 5-point scale.

The study protocol, amendments and patient informed consent were reviewed by a national, regional or Independent Ethics Committee or Institutional Review Board. BE RADIANT (NCT03536884) was conducted in accordance with the current version of the applicable regulatory and International Conference on Harmonisation–Good Clinical Practice requirements, the ethical principles that have their origin in the principles of the Declaration of Helsinki, and the local laws of the countries involved.

Measurements and Outcomes

The full P-SIM consists of 14 items are listed in Table 1. Each item was scored for severity or impact level at its worst over the past 24 h on a scale from 0 (no symptoms/impacts) to 10 (very severe). In BE RADIANT, three symptom items (skin itching, skin pain and skin scaling) from the full P-SIM were assessed, as highlighted in Table 1. These three items were used as key secondary endpoints in the BE VIVID and BE READY phase 3 trials, because they are subjective concepts that are most accurately assessed directly by the patient. These concepts can complement clinician-reported concepts and provide a holistic and comprehensive assessment of psoriasis signs and symptoms. It was therefore deemed appropriate to only include these items of the P-SIM for assessment in the BE RADIANT trial. Additionally, by including only three of the 14 P-SIM items, the time burden on patients would be reduced.

The three-item P-SIM was administered and completed by patients at baseline and at Weeks 4, 8, 12, 16, 32 and 48, as shown in Table 2, on electronic tablets. Other PRO instruments (DLQI and the Patient Global Assessment of Psoriasis [PGAP]) and clinician-reported outcome (ClinRO) instruments for the assessment of treatment efficacy (PASI and IGA) were also

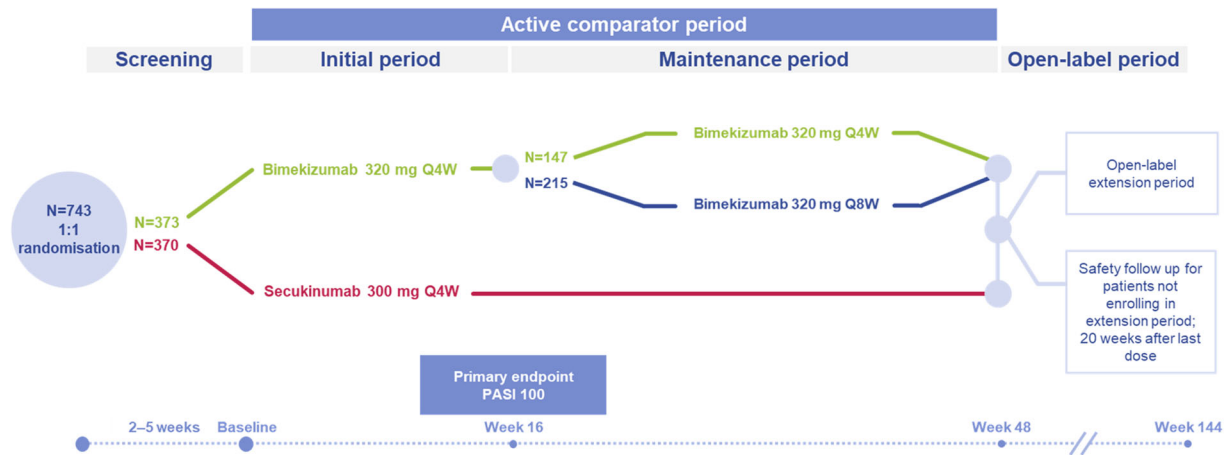


Fig. 1 In BE RADIANT, patients were randomised 1:1 to bimekizumab 320 mg Q4W or secukinumab 300 mg weekly to Week 4 then Q4W. Patients randomised to bimekizumab 320 mg Q4W either continued to receive Q4W dosing at Week 16 or switched to Q8W maintenance dosing. Secukinumab dosing was 300 mg at Weeks 0, 1, 2, 3 and 4, and Q4W thereafter. After completing the 48-week double-blinded period, patients could enrol in a

96-week open-label extension period to assess the long-term safety, tolerability and efficacy of bimekizumab. Patients who did not enrol in the open-label extension underwent a safety follow-up visit 20 weeks after their last dose of study treatment. *PASI 100* 100% improvement from baseline Psoriasis Area and Severity Index, *Q4W* every 4 weeks, *Q8W* every 8 weeks

used in BE RADIANT. PGAP and DLQI were administered at the same pre-defined study visits as the P-SIM, and PASI and IGA scores were recorded at every on-site study visit, as shown in Table 2. The PGAP consists of a multiple-choice question answered on a verbal rating scale: “How severe are your psoriasis-related symptoms right now?” the patient may respond that they have “no symptoms” (1), “mild symptoms” (2), “moderate symptoms” (3), “severe symptoms” (4), or “very severe symptoms” (5). The DLQI consists of 10 items, each scored from 0 to 3, with 3 representing the highest impact. A total score, ranging from 0 to 30, is obtained by adding the 10 item scores together, with higher scores indicating increased impact on health-related quality of life [9]. In this analysis, DLQI total score and DLQI Item 1 score (question: “Over the last week, how itchy, sore, painful or stinging has your skin been?”, answers: “Very much”; “A lot”; “A little” or “Not at all”) were used as part of the psychometric validation. Details of PASI and IGA score measurements can be found in the Supplementary Methods.

Statistical Analysis

Analyses were performed including all randomised patients, unless otherwise noted. Only blinded datasets were used; therefore, treatment group allocation was not considered in any of the analyses. Observed case data were used for the analyses, and no imputation was performed for missing data.

Compliance and Completion Rates

The compliance rate for each item at a given visit was calculated as the number of patients who had a non-missing item score for that visit divided by the number of randomised patients who were receiving study treatment or who had a non-missing P-SIM score at that visit. The completion rate for each item at a given visit was calculated as the number of patients with a non-missing item score at that visit divided by the number of patients randomised at baseline.

Table 1 Items of the P-SIM

Item	Response scale
1. During the past 24 h, how severe was your worst skin itching?	0 (No itching)—10 (very severe itching)
2. During the past 24 h, how severe was your worst skin redness?	0 (No skin redness)—10 (very severe skin redness)
3. During the past 24 h, how severe was your worst skin pain?	0 (No skin pain)—10 (very severe skin pain)
4. During the past 24 h, how severe was your worst skin burning?	0 (No skin burning)—10 (very severe skin burning)
5. During the past 24 h, how severe was your worst skin scaling?	0 (No skin scaling)—10 (very severe skin scaling)
6. During the past 24 h, how severe was your worst skin cracking?	0 (No skin cracking)—10 (very severe skin cracking)
7. During the past 24 h, how severe was your worst skin dryness?	0 (No skin dryness)—10 (very severe skin dryness)
8. During the past 24 h, how severe was your worst skin irritation?	0 (No skin irritation)—10 (very severe skin irritation)
9. During the past 24 h, how severe was your worst skin sensitivity?	0 (No skin sensitivity)—10 (very severe skin sensitivity)
10. During the past 24 h, how severe was your worst skin lesions (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 thickening (hardening or roughening)?	0 (No skin thickening)—10 (very severe skin thickening)

Table 1 continued

Item	Response scale
12. During the past 24 h, at its worst, how severe was your psoriasis-related fatigue (weariness, tiredness)?	0 (No fatigue)—10 (worst possible fatigue)
13. During the past 24 h, at its worst, how much embarrassment 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 choice of clothing?	0 (No impact on my choice of clothing)—10 (completely impacted my choice of clothing)

Bolded items are those included in this analysis
P-SIM Psoriasis Symptoms and Impacts Measure

Psychometric Validation

Construct Validity

Convergent validity was assessed by calculating the correlation between the three P-SIM item scores and scores from the other outcomes measuring similar concepts. Spearman’s rank correlation coefficients and corresponding *p* values were calculated between the P-SIM itching, skin pain and scaling item scores and PROs (DLQI total score, DLQI Item 1 score and PGAP score) and ClinROs (PASI total score and IGA score) at baseline and Weeks 16, 32 and 48. Only pairs of outcome measures assessed on the same visit (shown in Table 2) were included in this analysis. A correlation coefficient > 0.3 and ≤ 0.5 signified moderate convergent validity and a coefficient > 0.5 signified strong convergent validity [10]. It was hypothesised that the three items of the P-SIM would have moderate to strong correlations with PASI, IGA, DLQI and PGAP total scores, apart from baseline correlations between the three items of the

Table 2 Timing and frequency of completed measures through treatment period

Measure	Screening	Baseline	Weeks													
			1	2	4	8	12	16	20	24	28	32	36	40	44	48
P-SIM		×			×	×	×	×				×				×
PGAP		×			×	×	×	×				×				×
DLQI		×			×	×	×	×				×				×
PASI	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×
IGA	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×

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

P-SIM and both PASI and IGA. These correlations would be slightly weaker as there would be limited variability in these anchors due to the inclusion criteria of BE RADIANT (IGA was required to be moderate or severe and $PASI \geq 12$). It was also hypothesised that the itching and skin pain items of the P-SIM would have a strong correlation with DLQI Item 1.

To assess the known-groups validity of P-SIM items (ability to differentiate between different patients subgroups), the mean scores of each of the three P-SIM items at Weeks 16, 32 and 48 were compared between subgroups of patients based on absolute PASI total score thresholds (≤ 1 , > 1 to ≤ 3 , > 3 to < 5 and ≥ 5) and IGA scores (0: "clear"; 1: "almost clear"; 2 or 3: "mild or moderate"; and 4: "severe"). PASI and IGA were used for the assessment of known-group validity as they are well-accepted clinical measures of psoriasis disease severity and were used to define primary efficacy endpoints in bimekizumab phase 3 studies. As expected due to the inclusion criteria of the BE RADIANT trial ($PASI \geq 12$, $IGA \geq 3$), there was low variability in scores for both measures at baseline; assessment of known-groups validity therefore focussed on Weeks 16, 32 and 48. The absolute PASI cut-off values considered to define subgroups (1, 3 and 5) have been shown to provide reliable estimates of disease activity that can be used to define treatment goals for psoriasis treatment and facilitate clinical decision-making [11]. Additionally, the absolute PASI cut-off

value of 12 was used to define moderate to severe plaque psoriasis in the BE RADIANT inclusion criteria. The response options of the 5-point IGA scale were used as the criteria to define subgroups of patients with clinically different psoriasis severity levels, an approach shown to be a valid and reliable measure of psoriasis severity [12].

It was hypothesised that patients with higher PASI or IGA scores would have higher mean and median P-SIM item scores. *p* values from an analysis of variance (ANOVA; *F* test) and the Kruskal–Wallis test were used to compare distributions between the known groups.

Sensitivity to Change

Spearman correlation coefficients and corresponding *p* values were calculated between changes from baseline to Weeks 16, 32 and 48 in P-SIM item scores and those in other relevant outcomes [PASI (percent and absolute change), IGA, PGAP, DLQI Item 1 and DLQI total score]. A correlation coefficient ≥ 0.3 demonstrated acceptable sensitivity to change [13, 14].

Test–Retest Reliability

For each of the three items of the P-SIM, test–retest reliability was assessed using intraclass correlation coefficients (ICCs). These were calculated as the ratio of between-patient variance to the total variance [15]. ICCs were calculated at time intervals between baseline and Week 4 and between Weeks 12 and 16 among

Table 3 Final response categorisation by anchor for RD threshold analyses

Anchor	No meaningful change/ improvement (0)	Improvement level (1)	Improvement level (2)	Improvement level (3)
PGAP change from baseline to Weeks 16, 32 and 48	≥ 0 categories	= - 1 category	≤ - 2 categories	N/A
DLQI Item 1 change from baseline to Weeks 16, 32 and 48	≥ 0 categories	= - 1 category	≤ - 2 categories	N/A
DLQI change from baseline in total score to Weeks 16, 32 and 48	> - 4 points	> - 8 to ≤ - 4 points	≤ - 8 points	N/A
PASI percent reduction from baseline to Weeks 16, 32 and 48	< 75%	≥ 75% to < 100%	= 100%	N/A
IGA change from baseline to Week 16	≥ 5 points	> 1 to < 5 points	≤ 1 point	N/A
IGA change from baseline to Weeks 16, 32 and 48	≥ 0 categories	= - 1 category	≤ -2 categories	≤ -3 categories
IGA level at Weeks 16, 32 and 48	Mild/moderate/severe	Clear/almost clear	N/A	N/A

Categories that were collapsed due to small sample sizes ($n < 15$) were: (1) PGAP groups reporting “moderate,” “severe,” and “very severe symptoms” at all post-baseline visits; (2) DLQI Item 1 groups reporting “a lot” and “very much” at all post-baseline visits; (3) DLQI total score groups with a total score of “6–10,” “11–20,” and “21–30” at all post-baseline visits. *DLQI* Dermatology Life Quality Index, *IGA* Investigator’s Global Assessment, *PASI* Psoriasis Area and Severity Index, *PGAP* Patient Global Assessment of Psoriasis, *RD* responder definition

stable study participants (defined as those with the same PGAP score over these intervals). An ICC value ≥ 0.70 demonstrated acceptable test–retest reliability [15].

Determination of Responder Definition Thresholds

Anchor-based methods were used to define thresholds for within-patient clinically meaningful changes, in accordance with the FDA PRO guidance [16]. Distribution-based methods were also used to provide additional information about the variability of the relevant scale scores, as per the FDA guidance [16]. Triangulation was conducted by examining the results from all the anchor-based analyses, with the support of the estimates from the distribution-based methods, to define a range of values which could be used to interpret within-patient

meaningful change scores for each of the three P-SIM items. Anchor-based analyses were conducted at Weeks 16, 32 and 48. At each of these visits, only anchors with Spearman correlations ≥ 0.30 and with sample size ≥ 50 were considered for the RD determination. The anchors chosen are shown in Table 3. Empirical cumulative distribution functions (eCDF) and probability density function (PDF) curves of observed changes for each of the three P-SIM items were plotted separately at Weeks 16, 32 and 48 for each response group for each of the anchors. Distribution-based analyses were conducted using the standard error of measurement and using the standard deviation (SD) of the item scores at baseline and pre-defined study visits. If the SD was vastly different over time, the baseline SD was used. If not, the mean of all SDs across the relevant time points was used.

Table 4 Patient demographics and baseline disease characteristics

Characteristic	All randomised patients (<i>n</i> = 743)
Age, years, mean (SD)	45.0 (14.5)
Gender, <i>n</i> (%)	
Male	486 (65.4)
Race, <i>n</i> (%)	
White	695 (93.5)
Black or African American	10 (1.3)
Asian	19 (2.6)
American Indian or Alaska native	2 (0.3)
Native Hawaiian or other Pacific islander	2 (0.3)
Other or mixed	15 (2.0)
Geographic region, <i>n</i> (%)	
North America	291 (39.2)
Central/Eastern Europe	255 (34.3)
Western Europe	152 (20.5)
Asia/Australia	45 (6.1)
Duration of disease, years, mean (SD)	17.8 (12.7)
BSA affected by psoriasis, mean (SD)	24.3 (14.9)
Nail involvement, <i>n</i> (%)	383 (51.5)
Scalp involvement, <i>n</i> (%)	694 (93.4)
Palmoplantar involvement, <i>n</i> (%)	216 (29.1)
P-SIM item 1—itching, mean (SD)	6.7 (2.7)
P-SIM item 3—skin pain, mean (SD)	4.6 (3.2)
P-SIM item 5—scaling, mean (SD)	6.7 (2.4)
DLQI total score, mean (SD)	11.0 (6.9)

Table 4 continued

Characteristic	All randomised patients (<i>n</i> = 743)
PASI total score, mean (SD)	19.9 (7.1)
IGA, <i>n</i> (%)	
(3) Moderate	503 (67.7)
(4) Severe	232 (31.2)
PGAP, <i>n</i> (%)	
(3) Moderate symptoms	202 (27.2)
(4) Severe symptoms	348 (46.8)

BSA body surface area, *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

Results

Patient Disposition and Baseline Characteristics

In BE RADIANT, 743 patients were randomised. Baseline demographics and disease characteristics are shown in Table 4. At baseline, the mean age of study participants was 45.0 years and the majority of participants were male (65.4%) and white (93.5%). The majority of study participants were from North America (39.2%). Mean scores for the itching, skin pain and scaling items were 6.7, 4.6 and 6.7, respectively. Mean BSA affected by psoriasis was 24.3% and the mean PASI score was 19.9.

P-SIM Compliance Rates and Completion Rates

At each study visit, all participants completed either all three P-SIM items or none (i.e., there was no item-level missingness). P-SIM compliance rates for itching, skin pain and scaling ranged from 98.9% at baseline to 96.9% at Week 48, with an overall average across all visits of 99.1% (Supplementary Table 1). P-SIM completion rates for itching, skin pain and scaling were also high (above 92.4%) at each

time point apart from Week 48 (54.5%), with an overall average across all visits of 90.8% (Supplementary Table 2). The lower completion rate at Week 48 was due to the 252 study participants who had not yet reached the Week 48 visit at the time of the interim data-cut used in these analyses.

Psychometric Validation

Construct Validity

As hypothesised, the three P-SIM items had moderate to strong correlations with PGAP, DLQI total score and DLQI Item 1 at baseline and Weeks 16, 32 and 48 (Supplementary Table 3). A weak correlation (≤ 0.30) was observed at baseline between the three P-SIM item scores and the PASI and IGA scores, as expected due to study inclusion criteria. At Weeks 16, 32 and 48, the three P-SIM items had moderate to strong correlations with PASI total score and IGA, with the exception of the skin pain item score at Week 16, for which correlations were below the acceptable threshold of 0.3.

When examining known-groups defined using PASI total score (Fig. 2) and IGA score (Fig. 3), the P-SIM responded as expected across all three items, with higher P-SIM item scores reporting more severe symptoms at Weeks 16, 32 and 48. Statistically significant differences in P-SIM item scores were observed between the known-groups across all three P-SIM items.

Sensitivity to Change

Spearman correlation coefficients between changes from baseline in the three P-SIM items and in PRO measurements (PGAP, DLQI Item 1 and DLQI total score) were strong (> 0.5) or moderate (> 0.3 and ≤ 0.5) at Weeks 16, 32 and 48 (Table 5). In contrast, correlation coefficients with changes in ClinROs showed a weak correlation (≤ 0.30 , except for the P-SIM scaling item with PASI total score at Week 48 [0.31], as expected due to the inclusion of a scaling component in PASI).

Test–Retest Reliability

Only ICCs calculated between Weeks 12 and 16 are reported, as between baseline and Week 4 the sample size was small ($n = 79$) and patients may have been misclassified as “stable” during this time interval because of improvements in some of their symptoms (but not to the point of changing the PGAP category) given that both study treatments were highly effective in relieving symptoms in the first few weeks of treatment. It was therefore not possible to draw conclusions from the ICCs calculated between baseline and Week 4. Through Weeks 12–16, ICC values were above the acceptability threshold of 0.70 (Table 6).

Determination of Responder Definition Thresholds

Among all proposed anchors, only PGAP score, DLQI Item 1 score and DLQI total score were included in the anchor-based analysis; PASI percent and absolute change from baseline and IGA change from baseline all had correlation coefficients well below the acceptable threshold of 0.30 [13, 14], with changes from baseline in P-SIM item scores across all time points. Therefore, it was deemed inappropriate to include them in the anchor-based analysis.

Table 7 shows the mean changes in P-SIM items by response category used for each anchor. The mean changes obtained for one level of improvement were considered to be appropriate potential RD thresholds for the three items of the P-SIM based on the eCDF plots, which consistently separated from those in the ‘no meaningful change’ groups (Supplementary Figs. 1 and 2).

Values obtained from the PGAP and DLQI Item 1 score change from baseline were prioritised for four main reasons: the PGAP is the preferred anchor of the US FDA; both the PGAP and DLQI Item 1 are measured using ordinal scales with directly interpretable verbal descriptors; the resulting false positive rate, which should be minimised, was the highest with the ‘no meaningful improvement group’ based on DLQI change in total score; and these anchors assess proximal concepts that are

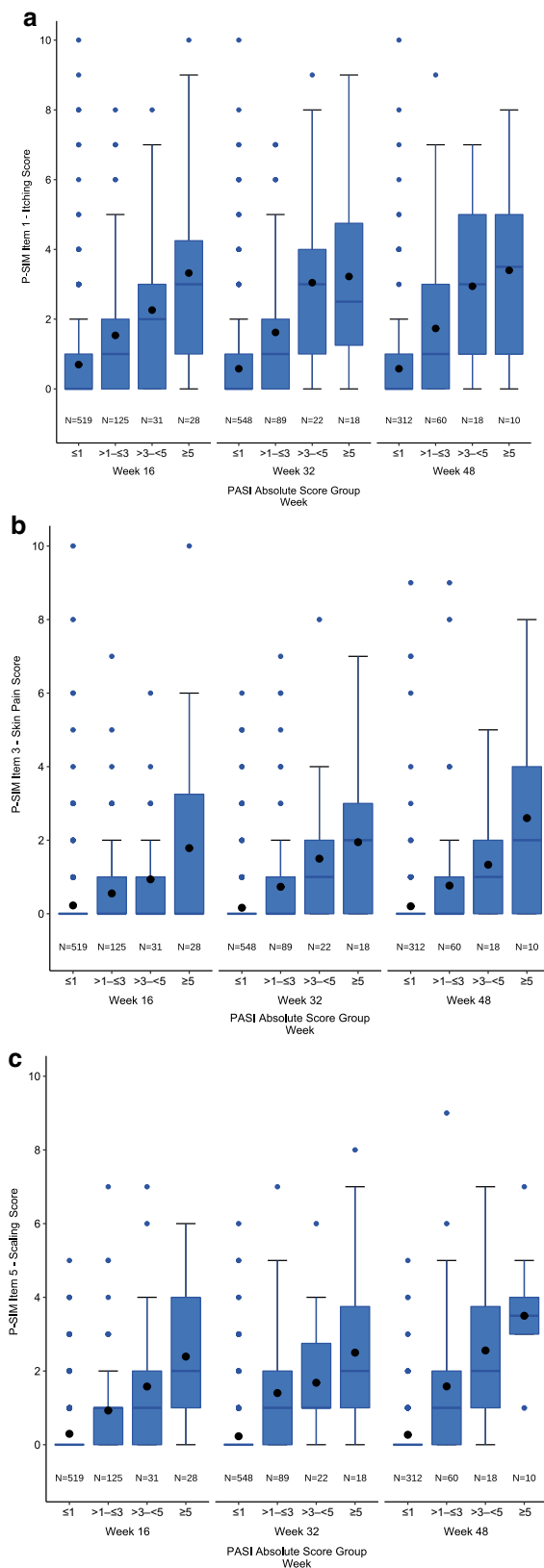


Fig. 2 P-SIM item scores at Week 16, 32 and 48 by PASI total score. **a** P-SIM Item 1—itching. **b** P-SIM Item 3—skin pain. **c** P-SIM Item 5—scaling. Black circular markers indicate mean scores and blue circular markers indicate outliers. *PASI* Psoriasis Area and Severity Index, *P-SIM* Psoriasis Symptoms and Impacts Measure

similar to the concepts measured by the selected P-SIM items.

The P-SIM item RD thresholds found using mean score changes ranged from 2.40 (Item 3: skin pain at Week 16) to 5.82 (Item 5: scaling at Week 48) (Table 7). Distribution-based values were lower compared to those obtained with the anchor-based approach and were considered as lower bounds of possible RD thresholds that would be less conservative. The mean change in P-SIM item scores obtained with one-point improvement on DLQI Item 1 and PGAP were close to the four-point FDA-recommended threshold, which was confirmed upon the inspection of the eCDF and PDF plots (the three P-SIM item eCDF curves for PGAP and DLQI Item 1 are presented in Supplementary Fig. 1 and 2, respectively). Based on these results, it was determined that a four-point improvement threshold represented a marked clinical improvement in patients with moderate to severe plaque psoriasis.

DISCUSSION

Convergent validity assessment showed that the three items of the P-SIM had moderate to strong correlations with PGAP score, DLQI Item 1 score and DLQI total score, but weak correlations with PASI score and IGA score at baseline and Week 16 (for the skin pain item only). The weak correlations at baseline could be explained by the low variability of these anchors at baseline due to BE RADIANT inclusion criteria. The exact reason for the weak correlation between the skin pain item and PASI and IGA at Week 16 is unknown; however, it could be because ClinROs mainly focus on the appearance of the skin, therefore skin pain would be inadequately

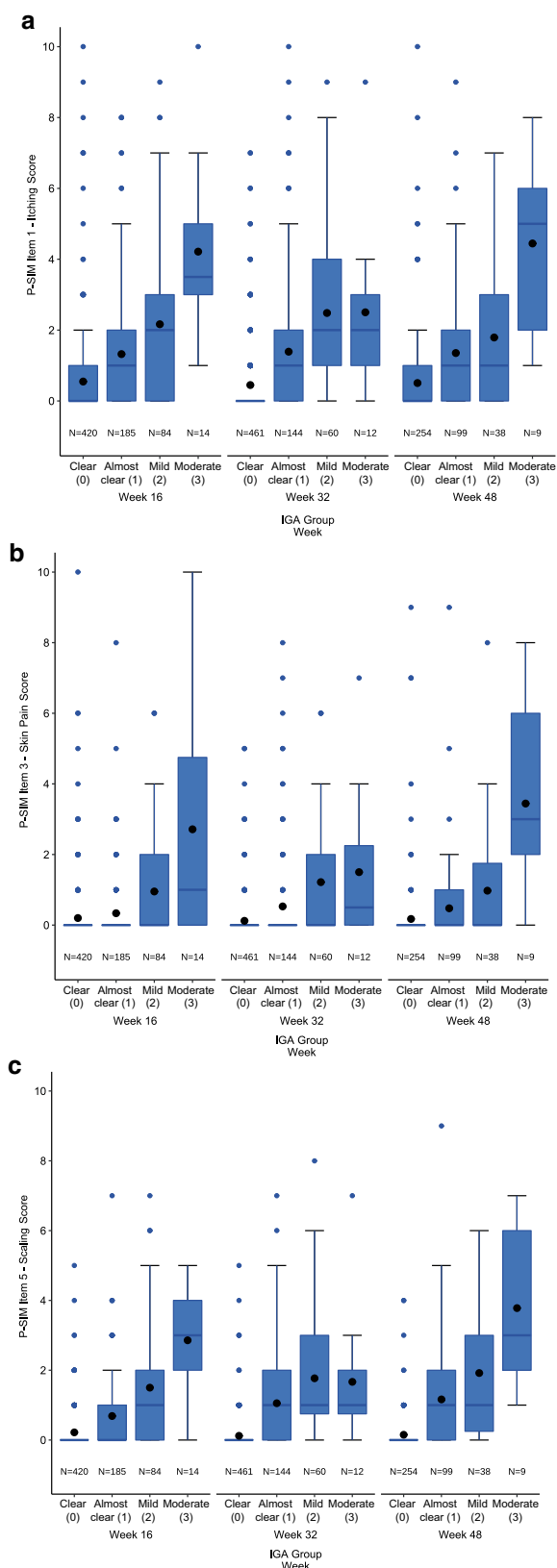


Fig. 3 P-SIM item scores at Week 16, 32 and 48 by IGA score subgroup. **a** P-SIM Item 1—itching. **b** P-SIM Item 3—skin pain. **c** P-SIM Item 5—scaling. Black circular markers indicate mean scores and blue circular markers indicate outliers. IGA Investigator’s Global Assessment, P-SIM Psoriasis Symptoms and Impacts Measure

captured. Overall, these results indicate that the three P-SIM items have good convergent validity.

Statistically significant score differences were observed across the three items of the P-SIM between known groups based on PASI total score and IGA score, demonstrating the ability of the P-SIM to discriminate between known subgroups.

When examining sensitivity to change, correlation coefficients between changes in the P-SIM items and PRO measures were strong or moderate at Weeks 16, 32 and 48. However, correlation coefficients were weak between the P-SIM items and ClinRO measures (except for the scaling item and PASI total score at Week 48). Such low correlations between these clinician-rated measures and P-SIM items are likely due to two reasons. First, the variability of change scores from baseline was greatly reduced due to the use of highly effective treatments in this study, resulting in almost no patients with worsening and only a very small percentage of patients presenting no meaningful improvement based on these clinical measures. Additionally, the ClinRO scales may be less sensitive in distinguishing between different levels of improvement as compared to the PRO scales considered.

Generally, the P-SIM items demonstrated good test–retest reliability, with all ICC values well above the acceptability threshold of 0.70 between Week 12 and Week 16, demonstrating that the P-SIM is consistent in its measurement of signs, symptoms and impacts of plaque psoriasis.

RD thresholds were estimated in accordance with the FDA guidance [16]. For the anchor-based approach, the mean item score changes in the one level of improvement anchor group

Table 5 Spearman correlations between changes from baseline in P-SIM item scores and other clinician- and patient-reported outcomes

P-SIM Item	PGAP (<i>n</i> , <i>r</i> , <i>p</i> value)	DLQI Item 1 (<i>n</i> , <i>r</i> , <i>p</i> value)	DLQI total (<i>n</i> , <i>r</i> , <i>p</i> value)	PASI total (percent change) (<i>n</i> , <i>r</i> , <i>p</i> value)	PASI total (absolute change) (<i>n</i> , <i>r</i> , <i>p</i> value)	IGA (<i>n</i> , <i>r</i> , <i>p</i> value)
Week 16						
Item 1—itching	696, <i>0.47</i> , < 0.001	696, 0.66, < 0.001	699, <i>0.48</i> , < 0.001	695 , 0.18 , < 0.001	695 , 0.15 , < 0.001	695 , 0.19 , < 0.001
Item 3—skin pain	696, <i>0.41</i> , < 0.001	696, 0.55, < 0.001	699, <i>0.57</i> , < 0.001	695 , 0.06 , 0.147	695 , 0.11 , 0.003	695 , 0.08 , 0.034
Item 5—scaling	696, <i>0.52</i> , < 0.001	696, <i>0.48</i> , < 0.001	699, <i>0.46</i> , < 0.001	695 , 0.16 , < 0.001	695 , 0.14 , < 0.001	695 , 0.18 , < 0.001
Week 32						
Item 1—itching	671, <i>0.50</i> , < 0.001	671, 0.66, < 0.001	674, <i>0.52</i> , < 0.001	669 , 0.20 , < 0.001	669 , 0.13 , 0.001	669 , 0.19 , < 0.001
Item 3—skin pain	671, <i>0.39</i> , < 0.001	671, 0.54, < 0.001	674, <i>0.58</i> , < 0.001	669 , 0.08 , 0.049	669 , 0.11 , 0.004	669 , 0.11 , 0.006
Item 5—scaling	671, <i>0.58</i> , < 0.001	671, 0.53, < 0.001	674, <i>0.48</i> , < 0.001	669 , 0.24 , < 0.001	669 , 0.15 , < 0.001	669 , 0.23 , < 0.001
Week 48						
Item 1—itching	397, <i>0.52</i> , < 0.001	397, 0.68, < 0.001	398, <i>0.54</i> , < 0.001	398 , 0.22 , < 0.001	398 , 0.24 , < 0.001	398 , 0.23 , < 0.001
Item 3—skin pain	397, <i>0.41</i> , < 0.001	397, 0.60, < 0.001	398, <i>0.62</i> , < 0.001	398 , 0.09 , 0.081	398 , 0.17 , 0.001	398 , 0.14 , 0.006
Item 5—scaling	397, <i>0.57</i> , < 0.001	397, 0.51, < 0.001	398, <i>0.44</i> , < 0.001	398, <i>0.31</i> , < 0.001	398 , 0.20 , < 0.001	398 , 0.30 , < 0.001

Bold values indicates a weak correlation ($r \leq 0.30$), italic values indicates indicate a moderate correlation ($0.30 < r \leq 0.50$), and a lack of emphasis indicates a strong correlation ($r > 0.50$)

n, *r*, *p* number of study participants included in the analysis, correlation coefficients and corresponding *p* values, respectively *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

were lower for the skin pain item than for itching or scaling. This was likely due to the higher proportion of patients reporting little or no skin pain at baseline, whereas higher levels of scaling and itching were observed at baseline. This floor effect may impact the magnitude of

reductions seen in the skin pain item scores, relative to the itching and scaling items.

Limitations of this study include the high proportion of white (93.5%) patients included in the randomised set. The lack of racial diversity in the randomised set limits the generalis-

Table 6 Test–retest reliability of each P-SIM item score using intraclass correlation coefficients by response category ‘improvement level 1’

P-SIM Item	Week 12–Week 16 ICC ($n = 522$)
Item 1—itching	0.94
Item 3—skin pain	0.91
Item 5—scaling	0.95

Only study participants who were stable (having the same response to the Patient Global Assessment of Psoriasis at Week 12 and Week 16 at a given visit) are included. The intraclass correlation coefficient is calculated as the ratio of between-patient variance to total variance using analysis of variance

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

ability of the results. For example, post-inflammatory hyper- and hypopigmentation are generally of more concern in dark-skinned individuals [17]. Another potential limitation is the low completion rate at Week 48. It was necessary to determine the RD in advance of database unblinding to minimise bias. Therefore, all analyses were conducted on a blinded interim dataset. As a result, 252 patients had not yet completed Week 48 at the time of the data-cut-off, impacting the completion rates at Week 48. However, results at Week 32 were generally consistent with those at Week 48, indicating that the impact of this lower completion rate at Week 48 is likely minimal.

In BE RADIANT, three symptom items (itching, skin pain and scaling) from the full P-SIM were assessed. Although other items of the P-SIM, such as redness and lesions, are also subjective concepts which could be assessed by the patient, changes in patient experiences of itching, skin pain and scaling are key outcomes in the evaluation of efficacy of treatments for plaque psoriasis and have previously been used to provide patient-reported evidence complementary to traditional ClinRO outcomes. This assertion is supported by findings from concept

elicitation research reported in the literature [5, 18].

In this study, a high proportion of patients had nail or scalp involvement (51.5% and 93.4%, respectively). Both nail and scalp involvement have been shown to severely impact patient quality of life; in future studies, the P-SIM should be used to assess experiences of patients with lesions involving the nails and scalp [19]. Based on this psychometric validation, a four-point RD threshold on an 11-point rating scale is proposed. Responder analyses using the four-point threshold to assess treatment effect may need to be restricted to patients reporting a score of ≥ 4 at baseline. Complementary analyses describing the proportion of patients reaching a score of 0 (no symptom) would allow all symptomatic patients to be included.

A strength of this psychometric validation work was the completion of the P-SIM during patient visits to the study site based on a pre-defined schedule. This resulted in all three item scores being obtained from a single assessment, which may have improved the completion rates as compared to a more regular assessment. Additionally, the use of verbal rating scales as anchors in this analysis may have more accurately captured patient experiences of psoriasis, especially for complex concepts such as skin pain.

CONCLUSIONS

In conclusion, the three items of the P-SIM assessed here demonstrated robust psychometric properties, including test–retest reliability, convergent and known-group validity and sensitivity to change in the concepts that they intended to measure. The anchor-based analysis in this study suggested that a four-point threshold for the three P-SIM items can be used to define marked within-patient meaningful change in patients with psoriasis treated over a period of 48 weeks.

Table 7 Mean changes in P-SIM item by response category within each anchor and each time point

Anchor	Improvement level (1) mean (95% CI)		
	Item 1—itching	Item 3—skin pain	Item 5—scaling
PGAP change from baseline to Weeks 16, 32 and 48			
Week 16	– 3.67 (– 4.16, – 3.19)	– 2.40 (– 2.86, – 1.96)	– 4.40 (– 4.84, – 3.95)
Week 32	– 3.69 (– 4.18, – 3.20)	– 2.77 (– 3.26, – 2.29)	– 4.28 (– 4.71, – 3.80)
Week 48	– 4.16 (– 4.72, – 3.59)	– 3.06 (– 3.67, – 2.45)	– 4.41 (– 4.90, – 3.95)
DLQI Item 1 change from baseline to Weeks 16, 32 and 48			
Week 16	– 4.46 (– 4.78, – 4.18)	– 3.12 (– 3.46, – 2.77)	– 5.34 (– 5.64, – 5.02)
Week 32	– 4.31 (– 4.64, – 4.00)	– 2.90 (– 3.26, – 2.57)	– 5.15 (– 5.48, – 4.86)
Week 48	– 4.40 (– 4.84, – 3.97)	– 2.69 (– 3.14, – 2.27)	– 5.19 (– 5.60, – 4.79)
DLQI change from baseline in total score to Weeks 16, 32 and 48			
Week 16	– 5.01 (– 5.47, – 4.54)	– 3.32 (– 3.81, – 2.89)	– 5.63 (– 6.01, – 5.25)
Week 32	– 5.16 (– 5.61, – 4.72)	– 3.33 (– 3.85, – 2.83)	– 5.68 (– 6.10, – 5.29)
Week 48	– 5.30 (– 5.84, – 4.76)	– 3.38 (– 3.99, – 2.78)	– 5.82 (– 6.24, – 5.36)

95% CIs were calculated using bootstrapping methods

CI confidence interval, *DLQI* Dermatology Life Quality Index, *PGAP* Patient Global Assessment of Psoriasis, *P-SIM* Psoriasis Symptoms and Impacts Measure

ACKNOWLEDGEMENTS

The authors thank the patients, the investigators and their teams who took part in this study. The authors also acknowledge Susanne Wiegatz, M.Sc., UCB Pharma, Monheim, Germany for publication coordination and Ruth Moulson, MPH, Costello Medical, London, UK, for medical writing and editorial assistance based on the authors' input and direction. This study was funded by UCB Pharma. RBW is supported

by the Manchester NIHR Biomedical Research Centre.

Funding. This study, as well as the Journal's Rapid Service and Open Access fees, was sponsored by UCB Pharma, Brussels, Belgium. This article was based on the original study BE RADIANT (NCT03536884) sponsored by UCB Pharma. Support for third-party writing assistance for this article, provided by Ruth Moulson, Costello Medical, UK, was funded by UCB

Pharma in accordance with Good Publication Practice (GPP3) guidelines (<http://www.ismpp.org/gpp3>).

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: Alice B. Gottlieb, Richard B. Warren, Matthias Augustin, Llenalia Garcia, Christopher Cioffi, Luke Peterson, Christopher Pelligra, Valerie Ciaravino; substantial contributions to analysis and interpretation of the Alice B. Gottlieb, Richard B. Warren, Matthias Augustin, Llenalia Garcia, Christopher Cioffi, Luke Peterson, Christopher Pelligra, Valerie Ciaravino; drafting the article or revising it critically for important intellectual content: Alice B. Gottlieb, Richard B. Warren, Matthias Augustin, Llenalia Garcia, Christopher Cioffi, Luke Peterson, Christopher Pelligra, Valerie Ciaravino; final approval of the version of the article to be published: Alice B. Gottlieb, Richard B. Warren, Matthias Augustin, Llenalia Garcia, Christopher Cioffi, Luke Peterson, Christopher Pelligra, Valerie Ciaravino.

Disclosures. Alice B. Gottlieb: 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. 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. Matthias Augustin:

Consulting fees from AbbVie, Almirall, Amgen, Biogen, Boehringer Ingelheim, Celgene, Centocor, Eli Lilly, GSK, Hexal, Janssen, LEO Pharma, Medac, Merck, MSD, Mundipharma, Novartis, Pfizer, Sandoz, UCB Pharma and Xenoport. Llenalia Garcia, Valerie Ciaravino: Employees of UCB Pharma. Luke Peterson, Christopher Cioffi: 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 protocol, amendments, and patient informed consent were reviewed by a national, regional, or Independent Ethics Committee or Institutional Review Board. BE RADIANT (NCT03536884) was conducted in accordance with the current version of the applicable regulatory and International Conference on Harmonisation-Good Clinical Practice 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 RADIANT (NCT03536884) phase 3 clinical trial 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-

NonCommercial 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

- Boehncke WH, Schön MP. Psoriasis. *Lancet*. 2015;386(9997):983–94.
- National Health Service (NHS). Symptoms—psoriasis 2021 [updated 09–05–2018]. <https://www.nhs.uk/conditions/psoriasis/symptoms/>.
- Augustin M, Radtke MA. Quality of life in psoriasis patients. *Expert Rev Pharmacoecon Outcomes Res*. 2014;14(4):559–68.
- Basra MK, Fenech R, Gatt RM, Salek MS, Finlay AY. The Dermatology Life Quality Index 1994–2007: a comprehensive review of validation data and clinical results. *Br J Dermatol*. 2008;159(5):997–1035.
- Lebwohl M, Swensen AR, Nyirady J, Kim E, Gwaltney CJ, Strober BE. The Psoriasis Symptom Diary: development and content validity of a novel patient-reported outcome instrument. *Int J Dermatol*. 2014;53(6):714–22.
- Gottlieb AB, Ciaravino V, Cioffi C, Peterson L, Warren RB. 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.
- Blauvelt A, Papp KA, Merola JF, Gottlieb AB, Cross N, Madden C, et al. Bimekizumab for patients with moderate to severe plaque psoriasis: 60-week results from BE ABLE 2, a randomized, double-blinded, placebo-controlled, phase 2b extension study. *J Am Acad Dermatol*. 2020;83(5):1367–74.
- Warren RB, Gottlieb AB, Merola JF, et al. 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. *Dermatol Ther (Heidelb)*. 2021;. <https://doi.org/10.1007/s13555-021-00570-4>.
- Finlay AY, Khan GK. Dermatology Life Quality Index (DLQI)—a simple practical measure for routine clinical use. *Clin Exp Dermatol*. 1994;19(3):210–6.
- Cohen J. *Statistical power analysis for the behavioural sciences*. New York: Routledge; 1988.
- Gordon KB, Reich K, Crowley JJ, Korman NJ, Murphy FT, Poulin Y, et al. Disease activity and treatment efficacy using patient-level Psoriasis Area and Severity Index scores from tildrakizumab phase 3 clinical trials. *J Dermatol Treat*. 2020;. <https://doi.org/10.1080/09546634.2020.1747590>.
- Langley RG, Feldman SR, Nyirady J, van de Kerkhof P, Papavassilis C. 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(1):23–31.
- Hays RD, Farivar SS, Liu H. Approaches and recommendations for estimating minimally important differences for health-related quality of life measures. *COPD*. 2005;2(1):63–7.
- Revicki D, Hays RD, Cella D, Sloan J. Recommended methods for determining responsiveness and minimally important differences for patient-reported outcomes. *J Clin Epidemiol*. 2008;61(2):102–9.
- Cappelleri JC, Zou KH, Bushmakin AG, et al. *Patient-reported outcomes: measurement, implementation and interpretation*. Boca Raton: CRC Press; 2013.
- 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.
- 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(3):405–23.

-
18. Rentz AM, Skalicky AM, Burslem K, Becker K, Kaschinski D, Esser D, et al. The content validity of the PSS in patients with plaque psoriasis. *J Patient Rep Outcomes*. 2017;1(1):4.
 19. Khobzey K, Liskova I, Szegedi A, Pavlovsky L, Lunder T, Kingo K, et al. Effectiveness of adalimumab in the treatment of scalp and nail affection in patients with moderate to severe plaque psoriasis in routine clinical practice. *Acta Dermatovenerol Alp Pannonica Adriat*. 2017;26(1):11–4.