Dermatol Ther (Heidelb) (2023) 13:2817–2837
https://doi.org/10.1007/s13555-023-00990-4

ORIGINAL RESEARCH

Evaluating the Appropriateness of Existing Health-Related Quality of Life Measures in Lichen Planus

Aoife Mahon-Smith · George Skingley · Lara Ayala-Nunes · Anjali Batish · Rosie Sharp · Christel Naujoks · Eva Schruf · Nicolò Compagno · Santiago G. Moreno

Received: May 17, 2023 / Accepted: July 19, 2023 / Published online: October 5, 2023

ABSTRACT

Introduction: Lichen planus (LP) is an inflammatory skin disorder that can present in various forms across the body, including lesions on the skin (cutaneous LP [CLP]), scalp (lichen planopilaris [LPP]) and mucosal regions (mucosal LP [MLP]). Several existing patient-reported outcome measures (PROMs) were identified for potential use in LP clinical development programs. This study aimed to assess the content validity and psychometric measurement properties of the Dermatology Life Quality Index (DLQI), Epworth Sleepiness Scale (ESS), Scalpdex and Oral Lichen Planus Symptom Severity Measure (OLPSSM) in an LP population.

Methods: Patients completed the PROs at various time points as part of an international Phase 2 clinical study in adults with MLP (n = 37), LPP (n = 37) and CLP (n = 37). Test-retest reliability, construct validity and sensitivity to change were assessed. In addition, qualitative cognitive debriefing interviews were conducted with adults with MLP (n = 20), LPP (n = 19) and CLP (n = 19) in the USA and Germany to examine the PROM content validity.

Results: The DLQI demonstrated adequate reliability and validity, although its ability to detect change was modest and most items were considered not relevant in qualitative interviews.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s13555-023-00990-4.

A. Mahon-Smith (✉) · G. Skingley · L. Ayala-Nunes · A. Batish · R. Sharp
Patient-Centered Outcomes, Adelphi Values Ltd, Bollington, Cheshire, UK
e-mail: aoife.mahon-smith@adelphivalues.com

G. Skingley
e-mail: george.skingley@adelphivalues.com

L. Ayala-Nunes
e-mail: lara.ayala@adelphivalues.com

A. Batish
e-mail: anjalibatish@btinternet.com

R. Sharp
e-mail: rosie.sharp@adelphivalues.com

C. Naujoks · N. Compagno · S. G. Moreno
Novartis Pharma AG, Basel, Switzerland

C. Naujoks
e-mail: christel.naujoks@novartis.com

N. Compagno
e-mail: n.compagno@gmail.com

S. G. Moreno
e-mail: santgumo@gmail.com

E. Schruf
Novartis Pharma GmbH, Nürnberg, Germany
e-mail: eva.schruf@novartis.com

The ESS had good reliability but limited evidence of validity and ability to detect change. Conceptual relevance varied according to the qualitative interview data. The Scalpdex was miscellaneous across domains, but the 'Symptoms' domain performed well overall. Overall, Scalpdex concepts were reported as relevant by most LPP patients interviewed. The OLPSSM demonstrated good psychometric properties and strong evidence of content validity.
Conclusions: The psychometric and qualitative findings support the use of the OLPSSM and Scalpdex within specific LP subtypes but cautioned use of the DLQI. Administration of the ESS is not recommended in LP because of its poor psychometric performance. Given these limitations, further validation of non-specific disease measures is needed and/or the development of additional LP-specific PROMs.
Trial Registration: NCT04300296.

Keywords: Dermatology Life Quality Index; Epworth Sleepiness Scale; Health-related quality of life; Oral Lichen Planus Symptom Severity Measure; Lichen planus; Patient-reported outcomes; Psychometric evaluation; Qualitative evidence; Scalpdex

Key Summary Points

Why carry out this study?

A review of existing patient-reported outcome measures used in dermatological conditions indicated that there were some existing measures that could be appropriate for use in lichen planus; however, further qualitative and psychometric testing was required to address evidence gaps.

This study aimed to assess the content validity and psychometric measurement properties of the Dermatology Life Quality Index, Epworth Sleepiness Scale, Scalpdex and Oral Lichen Planus Symptom Severity Measure across three lichen planus subpopulations: cutaneous lichen planus, lichen planopilaris and mucosal lichen planus.

What was learned from the study?

The findings recommend the use of the Scalpdex and the Oral Lichen Planus Symptom Severity Measure with lichen planopilaris and oral mucosal lichen planus patients, respectively, and the Dermatology Life Quality Index in general lichen planus populations, with caveats. The Epworth Sleepiness Scale demonstrated weak psychometric properties and content validity when utilised with lichen planus patients.

This study highlights the importance of assessing the appropriateness of non-specific disease patient-reported outcome measures in disease-specific populations.

INTRODUCTION

Lichen planus (LP) is an inflammatory skin disorder estimated to affect between 0.5 and 1% of the population worldwide [1, 2]. LP can present in various forms across the body [3]. Cutaneous LP (CLP) lesions are the most common type of LP and are characterized by polygonal purple papules on the skin, often associated with severe itch and typically affecting flexor surfaces including the wrists, ankles and lower back [4]. Lichen planopilaris (LPP) is a follicular variant of LP and is most common in females [2]. LPP can present as painful and itchy patches of hair loss, predominantly localized to the centre of the scalp, along the frontal hair line and/or in the eyebrows [5]. If untreated, LPP can lead to irreversible scarring and alopecia [4]. Mucosal LP (MLP) lesions typically present as asymptomatic bilateral white striations or painful plaques localized in mucosal areas including buccal mucosa, tongue and gingivae, genitalia and conjunctiva [2, 4, 6]. Individuals may be diagnosed with more than one LP subtype, based on the clinical presentation [4].

Given the range of LP signs and symptoms (including itch, pain and a burning sensation at the affected areas) [1, 8–11], LP can have a significant impact on patients' health-related

quality of life (HRQoL) [4]. While qualitative literature is limited, there is evidence that LP patients, particularly CLP and MLP patients, experience psychological impacts including anxiety and depression [12]. Patients with oral MLP also report experiencing significant impacts to daily activities such as discomfort when having certain foods and drinks, which in some cases can result in depression and high levels of stress and anxiety [13, 14]. LPP patients have reported impacts on social interactions and daily activities as a result of scarring and hair loss, causing patients to have low self-esteem and feel self-consciousness [15].

Patient-reported outcome measures (PROMs) are commonly used in routine medical practice and clinical studies to measure symptoms and HRQoL from the patient perspective. It is important that PROMs are appropriate and fit for purpose in terms of content validity and psychometric validity in the context of use [16]. A review of existing PROMs used in LP and other similar dermatological conditions identified several PROMs that could be appropriate for use in LP clinical development programs. Specifically, dermatological measures such as the Dermatology Life Quality Index (DLQI) [17] and Scalpdex [18], and non-specific disease measures such as the Epworth Sleepiness Scale (ESS) [19], have been used to assess HRQoL in LP patients [15, 20–23]. While there is some evidence of content validity and psychometric properties for these measures in some dermatological conditions [23, 24], there is limited evidence to support their use in an LP population [25]. In contrast, while existing LP-specific PROMs such as the recently developed Oral Lichen Planus Symptom Severity Measure (OLPSSM) have strong content validity [8, 26], there is no published additional evidence of psychometric validation in an LP (nor any other) population.

To address the gaps in evidence and align with regulatory standards [16, 27], the current study aimed to assess the content validity and psychometric measurement properties of the DLQI, ESS, Scalpdex and OLPSSM in an LP population through the conduct of qualitative patient interviews and psychometric analysis of data from an international Phase 2 LP clinical study. Aligned with the United States Food and Drug Administration (FDA) patient-focused drug development (PFDD) guidance documents, a mixed-method approach was used to ensure that the patient voice was represented in the evaluation of the select PROMs and in future clinical study design in LP [28–31].

METHODS

Study Design

This study was conducted in two phases: In the quantitative phase the psychometric properties of the DLQI, ESS, Scalpdex and OLPSSM were assessed in an LP population. In the qualitative phase content validity of the measures was evaluated via cognitive debriefing interviews.

Compliance with Ethics Guidelines

Ethical approval and oversight were obtained for the clinical study including exit interviews ([clinicaltrials.gov ID: NCT04300296, EUDRACT: 2019-003588-24]) and the independent qualitative interviews (Western Copernicus Group Independent Review Board [WCG IRB; reference: 20216826]). The studies were performed in accordance with the Helsinki Declaration of 1964 and its later amendments, and all participants provided informed consent indicating their data will be used for medical research purposes and the study results may be published.

Quantitative Phase

The quantitative phase used data collected from a global, randomized, double-blind, placebo-controlled, multi-centre, parallel-group Phase 2 clinical study involving 111 adults with biopsy-proven forms of moderate to severe LP (based on Investigator Global Assessment [IGA] rating of ≥ 3) who were eligible for systemic therapy and not adequately controlled with topical corticosteroids of high-ultrahigh potency in the opinion of the investigator. The study consisted of three cohorts (CLP, MLP and LPP) and two

treatment periods (treatment period 1: baseline to Week 16; treatment period 2: Week 16 to Week 32) (Supplementary Material). For the psychometric analyses, treatment period 1 data were used. The PROMs selected were included as secondary or exploratory study endpoints.

Overview of PROMs

Table 1 provides a brief description of the PROMs included in the planned analyses and the cohorts they were administered to within the clinical study. Licenses to use the PROMs in the clinical study were obtained.

Anchor Measures

Anchor measures were developed and administered in the LP clinical study to the full clinical sample to support psychometric evaluation of the PROMs [16]. This included a five-point patient global impression of severity (PGI-S) item, a five-point patient global impression of change (PGI-C) item, a five-point Investigator's Global Assessment (IGA) scale and Item 1 of the DLQI ('Over the last week, how itchy, sore, painful or stinging has your skin been?'). The PGI-S and the IGA were administered at baseline and at Week 2, 4, 8, 12 and 16; the PGI-C was administered at Week 2, 4, 8, 12 and 16.

Psychometric Analysis

Item- and scale-level psychometric analyses were conducted (Table 2). Unless noted otherwise, Week 4 data were used, as this time point was identified to provide a greater range of scores. As the PROMs were not appropriate for use in all LP types, analyses were conducted with different patient samples, e.g., DLQI and

Table 1 Overview of the PROMs included in the quantitative phase of the study

PROM	Description of measure	Recall period	Response options	Scoring	Clinical study sample
DLQI [17]	Generic measure developed for use across dermatological conditions. Ten items assessing HRQoL across six domains: 'Symptoms and feelings' (Items 1 & 2), 'Daily activities' (Item 3 & 4), 'Leisure' (Item 5 & 6), 'Work and school' (Item 7), 'Personal relationships' (Item 8 & 9), and 'Treatment' (Item 10).	'Over the last week'.	Four-point response scale from 'Very much' to 'Not at all'. Item 7 has a yes/no skip option followed by a 3-point response scale. 'Not relevant' option for items 3-10.	The total score (the sum of 10 item scores) ranges from 0 to 30, with higher scores indicating worse dermatology-specific QoL. The domain scores are calculated by summing item scores in each domain.	Administered to full clinical study sample (CLP, MLP, LPP) (Baseline [n=111], Week 4, 8, 12 and 16 [n=108, respectively]).
ESS [19]	Generic measure including eight items assessing daytime sleepiness.	'Over recent times'.	Four-point response scale from 'Would *never* doze' to '*High* chance of dozing'.	The total score (the sum of all eight item scores) ranges from 0 to 24. A higher score indicates greater 'daytime sleepiness'.	Administered to full clinical study sample (CLP, MLP, LPP) (Baseline [n=111], Week 2, 4, 8, 12 and 16 [n=108, respectively]).
Scalpdex [18]	Developed for use in psoriasis and seborrheic dermatitis. Twenty-three items assessing HRQoL across three domains: 'Symptoms' (Items 1, 3, 8), 'Functioning' (Items 13, 15, 18, 21, 23) and 'Emotions' (Items 2, 4-7, 9-12, 14, 16-17, 19-20, 22).	'Over the past four weeks'.	Five-point response scale from 'Never' to 'All the time'.	All scales (total and domain) were scored by summing items in the scale, with scores converted to a range from 0 to 100. Responses to item 18, "Caring for my scalp condition is inconvenient for me," were reverse scored. Higher scores indicate worse QoL.	Administered to LPP patients only (Baseline, Week 4, 8, 12 and 16 [n=37, respectively]).
OLPSSM [8]	Developed for use in oral lichen planus. Seven items assessing 'soreness' when completing activities of daily living.	'Over the past 24 hours'.	Five-point response scale from 'Not at all sore' to 'Too sore to do'.	The total score (the sum of all seven item scores) ranges from 0 to 28. A higher score indicates greater severity of 'soreness'.	Administered to MLP patients with oral involvement only (Baseline, Week 2, 4, 8, 12 and 16 [n=33, respectively]).

PROM Patient-reported outcome measure, *DLQI* Dermatology Life Quality Index, *HRQoL* Health-related quality of life, *QoL* Quality of life, *ESS* Epworth Sleepiness Scale, *OLPSSM* Oral Lichen Planus Symptom Severity Measure, *OLP* Oral lichen planus

Table 2 Summary of psychometric analyses

Analysis	Description
Dimensionality analyses	
Inter-item correlations	• Inter-item correlations using either polyserial or Spearman correlation coefficients were computed to explore the relationship between items and the latent health-related constructs they aim to measure. Items which correlated with one another >0.90 were flagged for review.
Scale-level analyses	
Internal consistency reliability	• Internal consistency, concerned with the homogeneity of items belonging to the same domain, was evaluated using Cronbach's alpha coefficient for each unidimensional score (≥0.70 for good internal consistency) [39]. • The alpha-if-item-deleted method was also conducted to assess whether the internal consistency of each domain would improve with the removal of each item in turn.
Test-retest reliability	• Test-retest reliability was calculated to evaluate the degree to which each measurement scores were similar over time in a subset of patients defined as having stable LP according to the IGA, PGI-S, PGI-C and DLQI item 1. • It was determined by calculating the intraclass correlation coefficient (ICC) in time in subsets of 'stable' patients (i.e., patients who had no change in the scores of anchor measures between Week 2 to 4 and Week 4 to 8). • The ICC based on a single measurement, absolute agreement, two-way mixed effects model was used [36]. Values less than 0.5 are indicative of poor reliability, values between 0.5 and 0.75 indicate moderate reliability, values between 0.75 and 0.9 indicate good reliability, and values greater than 0.90 indicate excellent reliability. • A simple Pearson's correlation was also calculated as a sensitivity analysis, although not interpreted with as much weight as the ICC. Where data were skewed, Spearman's correlation was calculated. • DLQI Item 1 was not included to assess DLQI total score test-retest reliability due to circularity. Test-retest reliability coefficients for the DLQI and Scalpdex were calculated between Week 4 and 8 only.
Convergent validity	• Convergent validity of the target PROMs was evaluated by examining the correlations between the scores of the target PROMs (DLQI, ESS, Scalpdex, OLPSSM) and anchor measures (PGI-S, PGI-C, and IGA). Correlations were assessed between the domain and total scores, where appropriate. • Measures that were expected to have a moderate to strong correlation coefficient with the ESS, the Scalpdex and the OLPSSM were the DLQI total score and DLQI item 1, as these reflect dermatological symptom experience. Equally, the target measures (the PGI-S, PGI-C, and IGA) were also expected to converge as they are global assessments of LP. • Domains assessing similar or related concepts were expected to correlate between ≥0.30 and ≥0.50.

Table 2 continued

Analysis	Description
	Convergent validity was not calculated for the DLQI because this measure was used as an anchor for the analysis, and this would entail circularity concerns.
Known-groups validity	• Construct validity was also assessed using the known-groups method [40], to evaluate differences in scores among patients who differ on health/disease related variables (None/Mild or Moderate/Severe/Very Severe for the PGI-S; 0-Clear/ 1-Minimal/ 2-Mild or 3-Moderate/ 4-Severe for the IGA). • F-test calculated by one-way ANOVAs (comparison of more than two groups) were used to evaluate if differences were statistically significantly ($p \leq 0.05$). • The magnitude of the differences was considered using between-group effect size estimates (Hedge's g), calculated using the pooled standard deviation as the denominator [41], and based against the reference group as defined above. • These analyses required splitting the total sample into further sub-groups based on the pre-defined disease variables (in this case the PGI-S and IGA), resulting in lower sample sizes in some groups. Where this is the case, *p*-values may be unreliable and in these cases the effect size will have more weight than the *p*-value as the main indicator of known-groups validity.
Ability to detect change	• Ability to detect change analysis, including comparing between and within groups mean change scores for the PGI-S, PGI-C and DLQI Item 1 for patients categorized as 'improved', 'no change' and 'worsened'. • Within-group effect sizes [37] and between-groups one-way ANOVA F-test were calculated to evaluate the magnitude and significance of the differences in change scores between each group, respectively. • Patients were categorized into 'improved', 'no change' and 'worsened' groups as follows: ○ PGI-S and DLQI Item 1: ▪ Improved: ≥1 grade improvement ▪ Stable: No change ▪ Worsened: ≥1 grade worsening ○ PGI-C: ▪ Improved: 'Minimally improved' to 'Very much improved' ▪ Stable: 'No change' ▪ Worsened: 'Minimally worse' to 'Very much worse'

LP Lichen planus, *IGA* Investigator Global Assessment, *PGI-S* Patient Global Impression of Severity, *PGI-C* Patient Global Impression of Change, *DLQI* Dermatology Life Quality Index, *ICC* Intraclass correlation coefficient, *ESS* Epworth Sleepiness Scale, *OLPSSM* Oral Lichen Planus Symptom Severity Measure, *ANOVA* Analysis of variance

ESS with all LP types ($n = 111$), Scalpdex with LPP only ($n = 37$) and OLPSSM with MLP patients with oral LP ($n = 33$). The aim of this study was not to evaluate the structure of the questionnaires; therefore, factor analyses were not conducted.

Qualitative Phase

The qualitative phase assessed the content validity of the PROMs via cognitive debriefing interviews. Given that the DLQI, ESS, Scalpdex and OLPSSM are existing validated measures, only relevance will be reported on, as evidence

of understanding is already available from the original development studies and consequent studies evaluating their use. An overview of the study procedure is provided in Supplementary Material, with further detail described in the subsequent sections.

Sample and Recruitment

A subset of patients ($n = 13$) enrolled in the Phase 2 LP clinical study in the US were invited to participate in an exit interview once they had completed all treatment visits to Week 32 but before their Week 40 follow-up visit. Participation was voluntary and patients could opt-out from taking part in an interview; patients who withdrew from the clinical study early were not eligible to participate in an exit interview. To further enhance the sample size, an additional and independent sample of patients ($n = 45$) were recruited by third-party recruitment agencies via referring clinicians in the US and Germany to participate in a qualitative interview. Inclusion and exclusion criteria for the independent interviews were broadly reflective of the LP clinical study eligibility criteria. Based on previous research, the sample included was deemed sufficient for assessing the content validity of the PROMs [32].

Interview Procedure

Interviews were 60 min and conducted via telephone by trained qualitative interviewers in the patient's native language using a semi-structured interview guide to facilitate the discussions. The cognitive debriefing (CD) section of the interview, which aimed to explore the relevance of the concepts assessed in the PROMs, lasted approximately 30 min and consisted of direct and focused questions.

Qualitative Analysis

All interviews were audio-recorded and transcribed verbatim with identifiable information redacted; the German interviews were further translated to English. Interview transcripts were analysed using Atlas.ti (Version 22) [33] using a framework approach [34]. Dichotomous codes were assigned to each item, instruction, response option(s) and recall period to indicate whether it was understood, relevant and/or appropriate, and why. Further codes captured any suggested changes.

RESULTS

Participant Demographic and Clinical Characteristics

Overviews of the demographic and clinical characteristics for the qualitative interviews ($N = 58$: exit interviews, $n = 13$; independent qualitative interviews, $n = 45$) are presented in Tables 3 and 4, respectively. Age was lower for MLP participants and there was a higher proportion of females, again reflecting the female inclination of LP [35]. Most participants enrolled were in the US and were Black or African American. There was a higher proportion of participants with 'moderate' LP, as confirmed by IGA severity scores at recruitment. The clinical study sample ($N = 111$; $n = 37$ in each LP cohort) was comparable with the qualitative samples; these data will be presented elsewhere.

Quantitative Phase

Item-Level and Dimensionality Analyses

Inter-Item Correlations As expected, items within the DLQI domains (Table 5) correlated well with each other, particularly 'Leisure' ($r = 0.894$) and 'Personal relationships' ($r = 0.890$). Items in the domains 'Symptoms and feelings' ($r = 0.479$) and 'Daily activities' ($r = 0.579$) correlated moderately, however, 'Daily activities' items correlated most strongly with Item 2 ('Embarrassed or self-conscious'), which was part the 'Symptoms and feelings' domain (range: $r = 0.721$–0.848). The ESS (Table 6) had a few weak correlations with the weakest ($r = 0.311$) being observed between Item 2 ('Watching TV') and Item 6 ('Sitting and talking to someone'). Majority of correlations were in the range of $r = 0.60$–0.70. No correlations in the ESS exceeded 0.80. For the Scalpdex

(Tables 7 and 8), inter-item correlations ranged from − 0.226 to 0.935. Items within Scalpdex domains overall correlated moderately, but this varied. Item 19 ('I feel that my knowledge for caring for my scalp is adequate'), Item 20 ('The cost of caring for my scalp condition bothers me') and Item 8 ('My scalp condition bleeds') had the lowest correlations with the remainder of the items, suggesting they measure concepts dissimilar to other items in the Scalpdex. A number of strong correlations were observed, suggesting potential redundancies. As shown in Table 9, the OLPSSM had few weak correlations < 0.40, with the weakest correlation ($r = 0.136$) being observed between Item 1 ('When you brushed your teeth') and Item 6 ('When you talked'). Majority of correlations were in the range of $r = 0.50$–0.60, with Item 2

Table 3 Participant demographic characteristics

Qualitative interview samples				
Exit interview sample[a]	CLP (n=4)	MLP (n=5)	LPP (n=4)	Total (N=13)
Age (years)				
Mean (SD)	58.8 (10.1)	46.8 (15.9)	66.5 (9.0)	56.5 (14.3)
Median	62.5	45.0	68.0	61.0
Minimum, maximum	44, 66	30, 73	64, 75	30, 75
Gender, n (%)				
Female	4 (100%)	4 (80%)	4 (100%)	12 (92%)
Male	-	1 (20%)	-	1 (8%)
Race, n (%)				
Black or African American	4 (100%)	2 (40%)	-	6 (46%)
White	-	1 (20%)	4 (100%)	5 (38%)
Asian	-	1 (20%)	-	1 (8%)
White and American Indian or Alaska Native	-	1 (20%)	-	1 (8%)
Independent interview sample	**CLP (n=15)**	**MLP (n=15)**	**LPP (n=15)**	**Total (N=45)**
Age (years)				
Mean (SD)	54.5 (14.7)	53.5 (11.1)	55.7 (10.5)	54.6 (12.0)
Median	52.0	54.0	54.0	54.0
Minimum, maximum	22, 76	39, 72	43, 75	22, 76
Gender, n (%)				
Female	9 (60%)	11 (73%)	9 (60%)	29 (64%)
Male	6 (40%)	4 (27%)	6 (40%)	16 (36%)
Race, n (%)[b]				
Black or African American	5 (33%)	3 (20%)	4 (27%)	12 (27%)
White	3 (20%)	4 (27%)	5 (33%)	12 (27%)
Hispanic	2 (13%)	2 (13%)	2 (13%)	6 (13%)
Asian	1 (7%)	-	-	1 (2%)
Not asked[b]	4 (27%)	6 (40%)	4 (27%)	14 (31%)
Education level, n (%)				
Elementary school[c]	1 (7%)	-	1 (7%)	2 (4%)
Middle and high school[d]	6 (40%)	7 (47%)	5 (33%)	18 (40%)
College or Associate degree	4 (27%)	5 (33%)	4 (27%)	13 (29%)
Undergraduate degree	1 (7%)	-	5 (33%)	6 (13%)
Graduate degree	3 (20%)	3 (20%)	-	6 (13%)

SD Standard deviation; *CLP* Cutaneous lichen planus; *MLP* Mucosal lichen planus; *LPP* Lichen planopilaris

[a]Education-level data were not collected for the exit interview sample

[b]Collection of racial data as part of surveys or studies in Germany is not permitted

[c]Elementary school defined as including Kindergarten to Grade 5

[d]Middle and high school defined as including Grade 6 to Grade 12, high school diploma/General Education diploma or equivalent

Table 4 Participant clinical characteristics

Qualitative interview samples				
Exit interview sample	**CLP (n=4)**	**MLP (n=5)**	**LPP (n=4)**	**Total (N=13)**
IGA score, n (%)				
Moderate	2 (50%)	4 (80%)	4 (100%)	10 (77%)
Severe	2 (50%)	1 (20%)	-	3 (23%)
Concomitant cutaneous affection, n (%)				
Concomitant MLP	-	-	N/A[a]	-
Concomitant CLP	-	1 (20%)	N/A[a]	1 (8%)
LP treatments[b], n (%)				
Biologic systemic medication	-	-	-	-
Non-biologic systemic medication	2 (50%)	2 (40%)	1 (25%)	5 (38%)
Topical medication	1 (25%)	2 (40%)	-	3 (23%)
Independent interview sample	**CLP (n=15)**	**MLP (n=15)**	**LPP (n=15)**	**Total (N=45)**
IGA score, n (%)				
Mild	-	3 (20%)	1 (7%)	4 (9%)
Moderate	9 (60%)	7 (47%)	9 (60%)	25 (56%)
Severe	6 (40%)	5 (33%)	5 (33%)	16 (36%)
Concomitant cutaneous affection, n (%)				
Concomitant MLP	2 (13%)	-	-	2 (4%)
Concomitant CLP	-	-	1 (7%)	1 (2%)
LP treatments[c], n (%)				
Corticosteroid	6 (40%)	13 (87%)	13 (87%)	32 (71%)
Immunosuppressant	1 (7%)	4 (27%)	4 (27%)	9 (20%)
Mouthwash	8 (53%)	1 (7%)	-	9 (20%)
Disease-modifying anti-rhematic drug	-	4 (27%)	3 (20%)	7 (16%)
Ultraviolet light therapy	-	1 (7%)	5 (33%)	6 (13%)
Anti-inflammatory drug/cream	2 (13%)	2 (13%)	1 (7%)	5 (11%)
Antibiotic	-	4 (27%)	-	4 (9%)
Anaesthetic	3 (20%)	-	-	3 (7%)
Anti-dandruff shampoo	-	2 (13%)	-	2 (4%)
Antihistamine	-	1 (7%)	-	1 (2%)
Diet adjustment	1 (7%)	-	-	1 (2%)
Phosphodiesterase-4 inhibitor	-	-	1 (7%)	1 (2%)
Retinoid	-	-	1 (7%)	1 (2%)

[a]Information about concomitant CLP or MLP in participants enrolled in the LPP cohort was not collected for this study
[b]Previous treatment categories for the clinical study sample were aligned with the clinical study eligibility criteria and thus differ from those collected for the independent interviews
[c]Some participants reported multiple LP treatments
LP Lichen planus, *CLP* Cutaneous lichen planus, *MLP* Mucosal lichen planus, *LPP* Lichen planopilaris, *SD* Standard deviation, *IGA* Investigator Global Assessment

('When you ate food') and Item 7 ('When it was touched') having the strongest correlation (r = 0.889), indicating possible redundancy.

Scale-Level Analyses

Internal Consistency Reliability Internal consistency was examined using Cronbach's alpha to assess the homogeneity of items belonging to the total measure score or domain score (Table 10). As Cronbach's alpha cannot be used for domains with fewer than three items, this was not assessed for DLQI domain scores. Alpha coefficients surpassed 0.70, indicating good internal consistency (DLQI total score = 0.920, ESS total score = 0.859, Scalpdex 'Functioning' domain score = 0.823, Scalpdex 'Emotions' domain score = 0.941, OLPSSM total score = 0.877), except for the Scalpdex 'Symptoms' domain score (0.655). However, this domain is only composed of three items, and

therefore lower reliability was expected. The measure with the highest reliability coefficient was the DLQI total score.

The alpha-if-deleted method was also conducted to assess whether the internal consistency of each total score or domain would improve with the removal of each item in turn (Supplementary Material). The overall internal consistency improved slightly with the removal of: DLQI Item 10 ('Over the last week, how much of a problem has the treatment for your skin been, for example by making your home messy, or by taking up time?') (0.921); ESS Item 6 ('Sitting and talking to someone') (0.864); Scalpdex 'Symptoms' domain Item 8 ('My scalp condition bleeds') (0.695); Scalpdex 'Functioning' domain Item 15 ('My scalp condition affects the color of clothes I wear') (0.844); Scalpdex 'Emotions' domain Item 19 ('I feel that my knowledge for caring for my scalp is adequate') (0.949) and Item 20 ('The cost of caring for my scalp condition bothers me') (0.949). However, given the marginal difference in the Cronbach's alpha coefficient, these results were not considered problematic.

Test-retest Reliability Test-retest reliability was evaluated to examine the stability of scores either between Week 2 and 4, and Week 4 and 8 for the scales (ESS total score & OLPSSM total score) assessed at those three time points, or between Week 4 and Week 8 for the scales

Table 5 Inter-item correlations of the DLQI at Week 4—total sample (n = 108)

Item	1	2	3	4	5	6	7	8	9	10
Item 1: Over the last week, how itchy, sore, painful or stinging has your skin been?	1.000									
Item 2: Over the last week, how embarrassed or self-conscious have you been because of your skin?	0.479	1.000								
Item 3: Over the last week, how much has your skin interfered with you going shopping or looking after your home or garden?	0.458	0.721	1.000							
Item 4: Over the last week, how much has your skin influenced the clothes you wear?	0.548	0.848	0.579	1.000						
Item 5: Over the last week, how much has your skin affected any social or leisure activities?	0.509	0.769	0.804	0.734	1.000					
Item 6: Over the last week, how much has your skin made it difficult for you to do any sport?	0.396	0.715	0.803	0.693	0.894	1.000				
Item 7: Over the last week, how much has your skin affected your work or study?	0.553	0.466	0.666	0.595	0.661	0.780	1.000			
Item 8: Over the last week, how much has your skin created problems with your partner or any of your close friends or relatives?	0.454	0.610	0.640	0.607	0.654	0.718	0.615	1.000		
Item 9: Over the last week, how much has your skin caused any sexual difficulties?	0.412	0.555	0.535	0.712	0.596	0.668	0.572	0.890	1.000	
Item 10: Over the last week, how much of a problem has the treatment for your skin been, for example by making your home messy, or by taking up time?	0.464	0.533	0.467	0.492	0.537	0.718	0.538	0.513	0.425	1.000

<0.40
0.40 ≤ x < 0.50
0.50 ≤ x < 0.60
0.60 ≤ x < 0.70
0.70 ≤ x < 0.80
≥0.80

DLQI Dermatology Life Quality Index

Table 6 Inter-item correlations of the ESS at Week 4—total sample (n = 108)

Item	1	2	3	4	5	6	7	8
Item 1: Sitting and reading	1.000							
Item 2: Watching TV	0.683	1.000						
Item 3: Sitting, inactive in a public place (e.g. a theatre or a meeting)	0.658	0.613	1.000					
Item 4: As a passenger in a car for an hour without a break	0.528	0.567	0.539	1.000				
Item 5: Lying down to rest in the afternoon when circumstances permit	0.640	0.749	0.469	0.500	1.000			
Item 6: Sitting and talking to someone	0.501	0.311	0.463	0.481	0.335	1.000		
Item 7: Sitting quietly after a lunch without alcohol	0.684	0.655	0.641	0.605	0.728	0.666	1.000	
Item 8: In a car, while stopped for a few minutes in the traffic	0.602	0.480	0.527	0.572	0.525	0.653	0.716	1.000

Kindly refer the legend of Table 5 for the significance of color codes
ESS Epworth Sleepiness Scale, *TV* Television

Table 7 Inter-item correlations of the Scalpdex Items 1–12 at Week 4—LPP sample (n = 37)

Item	1	2	3	4	5	6	7	8	9	10	11	12
Item 1: My scalp hurts.	1.000											
Item 2: My scalp condition makes me feel depressed.	0.471	1.000										
Item 3: My scalp itches.	0.619	0.607	1.000									
Item 4: I am ashamed of my scalp condition.	0.374	0.654	0.463	1.000								
Item 5: I am embarrassed by my scalp condition.	0.430	0.670	0.298	0.881	1.000							
Item 6: I am frustrated by my scalp condition.	0.230	0.773	0.350	0.742	0.791	1.000						
Item 7: I am humiliated by my scalp condition.	0.333	0.679	0.529	0.833	0.767	0.630	1.000					
Item 8: My scalp condition bleeds.	0.554	0.060	0.319	0.105	0.114	-0.226	0.252	1.000				
Item 9: I am annoyed by my scalp condition.	0.227	0.801	0.380	0.757	0.757	0.917	0.757	-0.100	1.000			
Item 10: I am bothered by the appearance of my scalp condition.	0.289	0.766	0.390	0.879	0.857	0.930	0.753	-0.086	0.935	1.000		
Item 11: My scalp condition makes me feel self-conscious.	0.205	0.591	0.281	0.908	0.896	0.663	0.749	0.050	0.675	0.792	1.000	
Item 12: I am bothered that my scalp condition is incurable.	0.185	0.630	0.287	0.661	0.781	0.810	0.654	-0.164	0.814	0.751	0.612	1.000
Item 13: My scalp condition affects how I wear my hair (hairstyle, hats).	0.240	0.561	0.424	0.826	0.838	0.821	0.670	-0.158	0.778	0.840	0.755	0.771
Item 14: I am bothered by people's questions about my scalp condition.	0.610	0.473	0.400	0.628	0.667	0.418	0.535	0.325	0.379	0.570	0.532	0.333
Item 15: My scalp condition affects the color of clothes I wear.	0.163	0.328	0.225	0.538	0.601	0.190	0.346	0.226	0.187	0.362	0.519	0.115
Item 16: I am bothered by the persistence/reoccurrence of my scalp condition.	0.449	0.767	0.505	0.586	0.744	0.799	0.579	0.165	0.780	0.780	0.507	0.904
Item 17: I feel stressed about my scalp condition.	0.335	0.833	0.513	0.667	0.750	0.752	0.665	0.055	0.829	0.760	0.700	0.765
Item 18: Caring for my scalp condition is inconvenient for me.	0.569	0.437	0.434	0.482	0.723	0.531	0.630	0.335	0.478	0.555	0.484	0.621
Item 19: I feel that my knowledge for caring for my scalp is adequate.	0.311	0.342	0.123	0.077	0.347	0.461	-0.023	-0.109	0.326	0.363	0.088	0.330
Item 20: The cost of caring for my scalp condition bothers me.	0.062	0.058	0.435	0.214	0.024	0.224	0.271	0.164	0.176	0.171	0.017	-0.054
Item 21: My scalp condition makes my daily life difficult.	0.510	0.632	0.435	0.794	0.773	0.503	0.757	0.301	0.561	0.624	0.722	0.481
Item 22: My scalp condition makes me feel different from others.	0.432	0.588	0.427	0.853	0.886	0.683	0.807	0.150	0.721	0.745	0.845	0.630
Item 23: My scalp condition makes it hard to go to the hairdresser/barber.	0.467	0.552	0.236	0.700	0.753	0.544	0.658	0.317	0.575	0.565	0.685	0.746

Kindly refer the legend of Table 5 for the significance of color codes
LPP Lichen planopilaris

Table 8 Inter-item correlations of the Scalpdex Items 13–23 at Week 4 —LPP sample (n = 37)

Item	13	14	15	16	17	18	19	20	21	22	23
Item 13: My scalp condition affects how I wear my hair (hairstyle, hats).	1.000										
Item 14: I am bothered by people's questions about my scalp condition.	0.648	1.000									
Item 15: My scalp condition affects the color of clothes I wear.	0.662	0.469	1.000								
Item 16: I am bothered by the persistence/reoccurrence of my scalp condition.	0.769	0.537	0.286	1.000							
Item 17: I feel stressed about my scalp condition.	0.736	0.506	0.397	0.769	1.000						
Item 18: Caring for my scalp condition is inconvenient for me.	0.576	0.526	0.448	0.822	0.659	1.000					
Item 19: I feel that my knowledge for caring for my scalp is adequate.	0.136	0.080	0.038	0.504	0.188	0.259	1.000				
Item 20: The cost of caring for my scalp condition bothers me.	0.420	0.253	0.443	-0.001	0.109	0.251	-0.202	1.000			
Item 21: My scalp condition makes my daily life difficult.	0.608	0.774	0.542	0.582	0.729	0.669	-0.057	0.111	1.000		
Item 22: My scalp condition makes me feel different from others.	0.676	0.693	0.403	0.533	0.732	0.647	0.109	0.210	0.827	1.000	
Item 23: My scalp condition makes it hard to go to the hairdresser/barber.	0.639	0.672	0.350	0.767	0.648	0.764	0.081	-0.074	0.893	0.742	1.000

Kindly refer the legend of Table 5 for the significance of color codes
LPP Lichen planopilaris

Table 9 Inter-item correlations of the OLPSSM at Week 4—MLP sample with OLP (n = 33)

Item	1	2	3	4	5	6	7
Item 1: When you brushed your teeth?	1.000						
Item 2: When you ate food?	0.576	1.000					
Item 3: When you drank liquids?	0.558	0.745	1.000				
Item 4: When you smiled?	0.518	0.658	0.820	1.000			
Item 5: When breathed through your mouth?	0.337	0.365	0.595	0.852	1.000		
Item 6: When you talked?	0.136	0.716	0.789	0.875	0.820	1.000	
Item 7: When it was touched?	0.472	0.889	0.518	0.525	0.276	0.564	1.000

Kindly refer the legend of Table 5 for the significance of color codes
OLPSSM Oral Lichen Planus Symptom Severity Measure, *MLP* Mucosal lichen planus, *OLP* Oral lichen planus

Table 10 Internal consistency using Cronbach's alpha for DLQI, ESS, Scalpdex and OLPSSM

Measure	LP subtype	Sample size	Total score	Domain score		
				Symptoms	Functioning	Emotions
DLQI	CLP, MLP, LPP	108	0.920	-	-	-
ESS	CLP, MLP, LPP	108	0.859	-	-	-
Scalpdex	LPP	37	-	0.655	0.823	0.941
OLPSSM	MLP with OLP	33	0.877	-	-	-

DLQI Dermatology Life Quality Index, *ESS* Epworth Sleepiness Scale, *OLPSSM* Oral Lichen Planus Symptom Severity Measure, *CLP* Cutaneous lichen planus, *MLP* Mucosal lichen planus, *LPP* Lichen planopilaris, *OLP* Oral lichen planus

(DLQI total score & Scalpdex 'Total', 'Symptoms', 'Emotions' and 'Functioning' domain scores) not assessed at Week 2.

When stability was defined using the IGA, PGI-S, PGI-C or DLQI item 1, all ICCs surpassed 0.75, indicating good test-retest reliability [36] (Table 11). Pearson's correlation coefficients were similar to the ICCs, providing further evidence of the reproducibility of measure scores in stable participants.

Concurrent Validity The ESS total score had weak correlations (≤ 0.250) with all convergent measures (Table 12). The Scalpdex total score correlated strongly with the DLQI total score (0.801) and moderately with the OLPSSM total score (0.353). Both the Scalpdex total score and the OLPSSM total score correlated moderately with the DLQI Item 1 (range: 0.473–0.504) and the PGI-S (range: 0.609–0.637), while both had weak correlations with the PGI-C (range: 0.173–0.290). The IGA correlated moderately with the OLPSSM total score (0.552) and weakly with the Scalpdex total score (0.030).

Known-Group Validity Known-group analyses compared DLQI total score, ESS total score, Scalpdex total and domain scores and OLPSSM total score, according to groups defined by IGA and PGI-S disease severity scores (Table 13). The DLQI total score, ESS total score, Scalpdex total, 'Symptoms' domain score and Scalpdex 'Emotions' domain score differed significantly ($p < 0.05$) among groups defined by the PGI-S, with moderate to large between-group effect size estimates. In contrast, the difference in mean scores between target PROMs and the IGA was non-significant with negative moderate to small between-group effect size estimates, suggesting that the IGA cannot discriminate between groups. Of note, due to the sample size for the OLPSSM, more weight should be given to the between-group effect size values to interpret validity; as such, OLPSSM scores show

Table 11 Test-retest reliability for the DLQI total score, ESS total score, Scalpdex domain scores and OLPSSM total score

Measure score	Anchor measure[a]	LP subtype	N	Timepoint 1 Mean (SD)	Timepoint 2 Mean (SD)	Reliability (ICC)	95% Confidence Interval		Pearson Correlation Coefficient
							Lower	Upper	
DLQI total score (Week 4-8)	IGA	MLP, CLP, LPP	67	6.3 (5.90)	6.4 (6.26)	0.82	0.72	0.88	0.818
	PGI-S		63	7.0 (6.09)	6.9 (6.35)	0.85	0.76	0.90	0.847
	PGI-C		53	5.6 (5.08)	5.5 (4.60)	0.77	0.63	0.86	0.768
ESS total score (Week 2-4)	IGA	MLP, CLP, LPP	70	7.5 (4.32)	7.3 (4.98)	0.80	0.69	0.87	0.805
	PGI-S		65	6.8 (4.00)	6.2 (4.71)	0.76	0.64	0.85	0.774
	PGI-C		54	7.0 (4.17)	6.5 (4.55)	0.81	0.69	0.88	0.816
ESS total score (Week 4-8)	DLQI Item 1	MLP, CLP, LPP	69	6.9 (4.81)	6.4 (4.19)	0.88	0.81	0.92	0.891
	IGA		67	6.5 (4.77)	6.3 (4.70)	0.90	0.85	0.94	0.904
	PGI-S		63	6.9 (4.85)	6.5 (4.35)	0.89	0.83	0.93	0.898
	PGI-C		53	6.3 (4.45)	6.5 (4.66)	0.90	0.83	0.94	0.900
Scalpdex 'Total' (Weeks 4-8)	DLQI Item 1	LPP	26	48.2 (20.89)	48.6 (17.43)	0.91	0.80	0.96	0.919
	IGA		21	49.5 (20.79)	50.9 (17.09)	0.91	0.80	0.96	0.926
	PGI-S		24	51.2 (19.63)	51.8 (18.28)	0.91	0.80	0.96	0.907
	PGI-C		18	49.2 (22.25)	50.9 (18.96)	0.87	0.70	0.95	0.882
Scalpdex 'Symptoms' (Week 4-8)	DLQI Item 1	LPP	26	33.0 (16.41)	32.4 (15.15)	0.79	0.59	0.90	0.792
	IGA		21	35.7 (15.62)	34.1 (16.44)	0.82	0.61	0.92	0.817
	PGI-S		24	32.6 (15.91)	32.3 (16.17)	0.80	0.59	0.91	0.795
	PGI-C		18	31.5 (17.04)	32.4 (15.63)	0.82	0.59	0.93	0.821
Scalpdex 'Functioning' (Week 4-8)	DLQI Item 1	LPP	26	47.5 (22.90)	46.0 (21.73)	0.91	0.81	0.96	0.909
	IGA		21	46.0 (23.85)	46.7 (23.84)	0.93	0.83	0.97	0.925
	PGI-S		24	50.4 (21.91)	49.8 (21.29)	0.94	0.86	0.97	0.934
	PGI-C		18	45.3 (22.85)	47.2 (19.72)	0.92	0.79	0.99	0.926
Scalpdex 'Emotions' (Week 4-8)	DLQI Item 1	LPP	26	51.5 (23.30)	52.7 (19.40)	0.89	0.77	0.95	0.900
	IGA		21	53.5 (23.59)	55.6 (19.18)	0.89	0.76	0.96	0.914
	PGI-S		24	55.2 (22.41)	56.4 (21.11)	0.87	0.73	0.94	0.872
	PGI-C		18	54.0 (25.05)	55.8 (22.02)	0.85	0.65	0.94	0.855
OLPSSM (Week 2-4)	IGA	MLP with OLP	21	11.0 (6.26)	10.2 (6.07)	0.89	0.75	0.95	0.894
	PGI-S		19	11.4 (5.46)	10.6 (5.68)	0.88	0.71	0.95	0.882
	PGI-C		18	9.9 (5.52)	9.2 (4.63)	0.80	0.55	0.92	0.811
OLPSSM (Week 4-8)	DLQI Item 1	MLP with OLP	21	9.9 (4.97)	10.9 (6.26)	0.83	0.64	0.93	0.863
	IGA		23	10.4 (5.42)	11.0 (7.17)	0.86	0.71	0.94	0.896
	PGI-S		21	11.6 (5.55)	11.3 (7.14)	0.89	0.74	0.95	0.912
	PGI-C		15	10.5 (5.24)	10.5 (6.40)	0.85	0.60	0.95	0.856

DLQI Dermatology Life Quality Index, *ESS* Epworth Sleepiness Scale, *OLPSSM* Oral Lichen Planus Symptom Severity Measure, *LP* Lichen planus, *SD* Standard deviation, *ICC* Intra-class correlation coefficient, *IGA* Investigator Global Assessment, *PGI-S* Patient Global Impression of Severity, *PGI-C* Patient Global Impression of Change, *MLP* Mucosal lichen planus, *CLP* Cutaneous lichen planus, *LPP* Lichen planopilaris, *OLP* Oral lichen planus

[a]Stability defined as no change

evidence of being able to discriminate between groups for the PGI-S known groups and the IGA known groups.

Ability to Detect Change Within-group effect sizes [37] and between-group one-way ANOVA F-test were calculated to evaluate the magnitude and significance of the differences in change scores between each group (improved/worsened versus stable participants) (Table 14).

For the DLQI total score, change scores between groups were statistically significant for both the PGI-S and PGI-C. For the ESS, small effect sizes were observed for all groups in the PGI-S and PGI-C and in the improved group for the DLQI Item 1. However, effect sizes were either non-significant (DLQI Item 1), in an unexpected direction (PGI-S) or similar for the stable and improved/worsened groups (PGI-C), suggesting that the ESS has limited ability to detect change in these anchor measures. For the OLPSSM, both the PGI-S and PGI-C showed a statistically significant difference between groups; however, statistical significance was not achieved for the DLQI Item 1.

For the Scalpdex total score, small effect sizes were found across the three groups for the DLQI, PGI-S and PGI-C, except for a moderate effect in the stable group for the DLQI. The DLQI and the PGI-C demonstrated some

Table 12 Concurrent validity correlations of DLQI, ESS, Scalpdex and OLPSSM with hypothesized convergent measures

Measure score	LP subtype	N	Convergent measure correlation coefficient[a]				
			DLQI total score[b]	DLQI Item 1	PGI-S	PGI-C[b]	IGA
ESS total score	MLP, CLP, LPP	107	0.250	0.167	0.216	0.032	0.171
Scalpdex total score	LPP	37	0.801	0.540	0.609	0.173	0.030
OLPSSM total score	MLP	33	0.353	0.473	0.637	0.290	0.552

DLQI Dermatology Life Quality Index, *ESS* Epworth Sleepiness Scale, *OLPSSM* Oral Lichen Planus Symptom Severity Measure, *LP* Lichen planus, *PGI-S* Patient Global Impression of Severity, *PGI-C* Patient Global Impression of Change, *IGA* Investigator Global Assessment, *MLP* Mucosal lichen planus, *CLP* Cutaneous lichen planus, *LPP* Lichen planopilaris

[a]Light shaded cells indicate hypothesized correlations (> 0.30 or > 0.40) were met, indicating concurrent validity; dark shaded cells indicate hypothesized correlations were not met

[b]Indicates Spearman's correlation, otherwise polyserial

evidence of ability to detect change. The Scalpdex 'Symptoms' score had a large effect size for improved groups in all measures (DLQI Item 1, PGI-S, PGI-C). Worsened groups demonstrated a large effect size in the DLQI Item 1, a moderate effect size in the PGI-C and a small effect size in the PGI-S. A small effect size was observed for the stable groups in all three anchor measures. Change scores between groups were statistically significant for the DLQI Item 1 and PGI-C but not for PGI-S; however, the PGI-S p value may have been impacted by the low sample size for the worsened group. For the Scalpdex 'Functioning' score, all groups (improved, worsened, stable) had small effects sizes in all three anchors (DLQI Item 1, PGI-S, PGI-C). The only statistically significant difference between groups was for the PGI-S. For the Scalpdex 'Emotions' score, the DLQI Item 1 had a statistically significant change between groups with a small effect size reported for the improved and stable groups and a moderate effect size for the worsened group. The PGI-S and PGI-C had small effect sizes for all groups, with the change scores between groups being statistically significant for the PGI-S and not statistically significant for the PGI-C.

Qualitative Phase

DLQI

The DLQI was cognitively debriefed with all exit interviews participants ($n = 13$). Individual items did not perform well in terms of relevance; i.e., most items ($n = 8/10$, 80%) were considered relevant to less than half of participants. The least relevant items were Item 6 ('Over the last week, how much has your skin made it difficult for you to do any sport?') and Item 7 ('Over the last week, has your skin prevented you from working or studying?') ($n = 1/13$, 7.7% per item). The most relevant items were Item 1 ('Over the last week, how itchy, sore, painful or stinging has your skin been?'; $n = 11/11$, 100.0%) and Item 2 ('Over the last week, how embarrassed or self-conscious have you been because of your skin?'; $n = 9/13$, 69.2%), both of which are included in the DLQI 'Symptoms and feelings' domain.

ESS

The ESS was cognitively debriefed with a total of 49 participants (CLP participants during the exit interviews: $n = 4$, all participants during the independent interviews: $n = 45$). Relevance was mixed, with just over half of items ($n = 5/8$, 62.5%) being considered relevant to at least half of participants. The item that demonstrated the highest relevance was Item 5 ('Lying down to rest in the afternoon when circumstances permit'; $n = 43/49$, 87.8%). Item 6 ('Sitting and talking to someone'; $n = 8/48$, 16.7%) demonstrated the lowest relevance. Some participants were also asked additional probes about sleepiness with almost all participants reporting never feeling sleepy because of LP ($n = 19/20$, 95%) and most participants reporting never dozing off or falling asleep due to LP ($n = 11/13$, 84.6%).

Scalpdex

The Scalpdex was cognitively debriefed with a total of 19 LPP participants (exit interviews:

Table 13 Known-group analysis comparisons of DLQI total score, ESS total score, Scalpdex domain scores and OLPSSM total score

Item score	LP subtype	n	Known-group definition[a]	Mean (SD)	Median	Between groups mean difference	Between groups effect size	*p*-value
DLQI	MLP, CLP, LPP	28	IGA Reference	6.3 (5.02)	5	0.5	0.08	0.701
		79	IGA Severe	6.8 (6.31)	5			
		28	PGI-S Reference	3.5 (4.00)	2	4.2	0.75	<0.001
		79	PGI-S Severe	7.7 (6.18)	6			
ESS	MLP, CLP, LPP	28	IGA Reference	5.6 (3.24)	6	1.3	0.29	0.196
		79	IGA Severe	6.9 (5.04)	6			
		28	PGI-S Reference	4.5 (3.95)	3	2.8	0.62	0.006
		79	PGI-S Severe	7.3 (4.69)	6			
Scalpdex 'Total'	LPP	11	IGA Reference	59.2 (15.64)	58	9.0	-0.45	0.222
		25	IGA Severe	50.2(21.49)	55			
		12	PGI-S Reference	41.6 (24.44)	49	17.1	0.92	0.014
		24	PGI-S Severe	58.7 (15.10)	59			
Scalpdex 'Symptoms'		11	IGA Reference	28.8 (22.47)	25	6.9	0.40	0.282
		25	IGA Severe	35.7 (14.73)	33			
		12	PGI-S Reference	20.8 (7.54)	17	19.1	1.27	0.001
		24	PGI-S Severe	39.9 (17.55)	42			
Scalpdex 'Functioning'		11	IGA Reference	59.5 (16.35)	60	-12.1	-0.55	0.139
		25	IGA Severe	47.4 (24.16)	50			
		12	PGI-S Reference	41.3 (27.97)	48	14.7	0.68	0.063
		24	PGI-S Severe	56.0 (18.00)	58			
Scalpdex 'Emotions'		11	IGA Reference	65.2 (15.87)	67	-11.1	-0.50	0.178
		25	IGA Severe	54.1 (24.43)	58			
		12	PGI-S Reference	45.8 (28.71)	55	17.5	0.82	0.026
		24	PGI-S Severe	63.3 (16.54)	63			
OLPSSM	MLP with OLP	5	IGA Reference[b]	8.2 (2.95)	8	2.4	0.45	0.362
		28	IGA Severe	10.6 (5.63)	10			
		7	PGI-S Reference[b]	7.7 (4.31)	7	3.2	0.61	0.162
		26	PGI-S Severe	10.9 (5.47)	11			

DLQI Dermatology Life Quality Index, *ESS* Epworth Sleepiness Scale, *OLPSSM* Oral Lichen Planus Symptom Severity Measure, *LP* Lichen planus, *SD* Standard deviation, *MLP* Mucosal lichen planus, *CLP* Cutaneous lichen planus, *LPP* Lichen planopilaris, *IGA* Investigator Global Assessment, *PGI-S* Patient Global Impression of Severity, *OLP* Oral lichen planus

[a]IGA Reference: 0 = clear, 1 = minimal, 2 = mild group; IGA Severe: 3 = moderate, 4 = severe group. PGI-S Reference: PGI-S none or mild group; PGI-S Severe: moderate, severe or very severe

[b]Indicates that one of the pre-defined groups had a sample size < 10

n = 4, independent interviews: *n* = 15). Relevance was high, with almost all items (*n* = 21/23, 91.3%) being considered relevant to at least half of participants. The most relevant items were Item 3 ('My scalp itches'), Item 6 ('I am frustrated by my scalp condition') and Item 9 ('I am annoyed by my scalp condition') (*n* = 18/19, 94.7% per item). The least relevant item was Item 15 ('My scalp condition affects the color of clothes I wear'; *n* = 7/19, 36.8%).

OLPSSM

The OLPSSM was cognitively debriefed with MLP participants with oral involvement during the exit interviews (*n* = 5). Just over half of the items (*n* = 4/7, 57.1%) were considered relevant to at least half of participants. Almost all participants considered Item 4 ('When you smiled?'; *n* = 4/5, 80.0%) and Item 6 ('When you talked?'; *n* = 4/5, 80.0%) relevant to their experience of MLP, while Item 5 ('When you breathed through your mouth?'; *n* = 2/5, 40.0%) was considered least relevant.

Of note, participant quotes to support the qualitative results are presented in Supplementary Material.

DISCUSSION

There are limited disease-specific PROMs that assess HRQoL in LP patients and a scarcity of

Table 14 Ability to detect change for the DLQI, ESS, Scalpdex and OLPSSM

Measure	Known-groups anchors	Ability to detect change anchor definition	N	Median	Min., Max.	Mean (SD)	*p*-value	Effect Size
DLQI - total score[a,b]	PGI-S: change from baseline	Improved: >=1 grade improvement	47	-3	-14, 4	-3.4 (4.50)	0.002	-0.563
		Stable: no change	44	-1	-16, 15	-1.2 (4.86)		-0.190
		Worsened: >=1 grade worsened	13	1	-8, 16	1.8 (6.14)		0.233
	PGI-C (response at week 16)	Improved: 'Minimally improved' or 'Much improved' or 'Very much improved'	45	-4	-14, 4	-4.1 (4.26)	<0.001	-0.664
		Stable: 'no change'	29	-1	-8, 2	-1.7 (2.78)		-0.282
		Worsened: 'Minimally worse' or 'Much worse' or 'Very much worse'	30	1	-16, 16	1.5 (6.24)		0.215
ESS - total score[a,b]	DLQI Item 1: change from baseline	Improved: >=1 grade improvement	43	-1	-11, 10	-1.1 (3.24)	0.471	-0.246
		Stable: no change	45	0	-7, 10	-0.3 (3.21)		-0.066
		Worsened: >=1 grade worsened	16	-1	-13, 9	0.1 (5.78)		0.029
	PGI-S: change from baseline	Improved: >=1 grade improvement	47	-1	-11, 7	-1.3 (3.32)	0.017	-0.297
		Stable: no change	44	0	-5, 10	0.6 (3.56)		0.140
		Worsened: >=1 grade worsened	13	-1	-13, 7	-1.9 (4.63)		-0.319
	PGI-C (response at week 16)	Improved: 'Minimally improved' or 'Much improved' or 'Very much improved'	45	-1	-11, 6	-1.5 (3.02)	<0.001	-0.320
		Stable: 'no change'	29	-1	-13, 5	-1.3 (3.29)		-0.282
		Worsened: 'Minimally worse' or 'Much worse' or 'Very much worse'	30	1	-5, 10	1.6 (4.24)		0.329
Scalpdex - 'Total' score[a,b]	DLQI Item 1: change from baseline	Improved: >=1 grade improvement	10	-6	-26, 14	-7.3 (13.49)	0.002	-0.472
		Stable: no change	20	-1	-18, 10	-2.1 (8.55)		-0.112
		Worsened: >=1 grade worsened	5	17	0, 23	13.9 (8.88)		0.555
	PGI-S: change from baseline	Improved: >=1 grade improvement	18	-3	-26, 9	-5.7 (10.40)	0.081	-0.362
		Stable: no change	13	4	-17, 23	3.3 (12.60)		0.175
		Worsened: >=1 grade worsened	4	3	-10, 18	3.5 (11.79)		0.320
	PGI-C (response at week 16)	Improved: 'Minimally improved' or 'Much improved' or 'Very much improved'	14	-3	-26, 10	-6.9 (11.73)	0.007	-0.363
		Stable: 'no change'	11	-2	-13, 9	-2.5 (7.48)		-0.128
		Worsened: 'Minimally worse' or 'Much worse' or 'Very much worse'	10	8	-17, 23	7.8 (11.70)		0.412
Scalpdex – 'Symptoms' score[a,b]	DLQI Item 1: change from baseline	Improved: >=1 grade improvement	10	-25	-42, 17	-18.3 (17.92)	<0.001	-1.205
		Stable: no change	20	-8	-17, 8	-5.4 (8.23)		-0.414
		Worsened: >=1 grade worsened	5	17	8, 33	18.3 (10.86)		0.814
	PGI-S: change from baseline	Improved: >=1 grade improvement	18	-8	-42, 17	-11.1 (14.85)	0.089	-0.910
		Stable: no change	13	0	-25, 33	-1.9 (16.37)		-0.097
		Worsened: >=1 grade worsened	4	4	-8, 25	6.3 (17.18)		0.300
	PGI-C (response at week 16)	Improved: 'Minimally improved' or 'Much improved' or 'Very much improved'	14	-17	-42, 0	-19.0 (11.05)	<0.001	-1.127
		Stable: 'no change'	11	0	-17, 17	-3.0 (10.05)		-0.256
		Worsened: 'Minimally worse' or 'Much worse' or 'Very much worse'	10	8	-8, 33	10.0 (12.91)		0.580
Scalpdex – 'Functioning' score[a,b]	DLQI Item 1: change from baseline	Improved: >=1 grade improvement	10	-5	-35, 30	-2.5 (20.58)	0.067	-0.137
		Stable: no change	20	-5	-30, 25	-2.0 (12.07)		-0.083
		Worsened: >=1 grade worsened	5	20	0, 20	15.0 (8.66)		0.489
	PGI-S: change from baseline	Improved: >=1 grade improvement	18	-5	-35, 30	-5.1 (15.10)	0.028	-0.293
		Stable: no change	13	10	-15, 25	8.5 (13.45)		0.435
		Worsened: >=1 grade worsened	4	0	-10, 20	2.5 (12.58)		0.104
	PGI-C (response at week 16)	Improved: 'Minimally improved' or 'Much improved' or 'Very much improved'	14	0	-35, 30	-2.1 (19.49)	0.071	-0.093
		Stable: 'no change'	11	-5	-15, 10	-5.0 (7.75)		-0.194
		Worsened: 'Minimally worse' or 'Much worse' or 'Very much worse'	10	13	-10, 25	9.5 (12.35)		0.389
Scalpdex – 'Emotions' score[a,b]	DLQI Item 1: change from baseline	Improved: >=1 grade improvement	10	-3	-25, 12	-6.7 (13.90)	0.014	-0.362
		Stable: no change	20	0	-23, 13	-1.5 (9.87)		-0.071
		Worsened: >=1 grade worsened	5	17	-3, 27	12.7 (11.82)		0.504
	PGI-S: change from baseline	Improved: >=1 grade improvement	18	-2	-25, 13	-4.4 (11.42)	0.246	-0.249
		Stable: no change	13	5	-23, 27	2.6 (13.82)		0.128
		Worsened: >=1 grade worsened	4	3	-10, 17	3.3 (12.17)		0.356
	PGI-C (response at week 16)	Improved: 'Minimally improved' or 'Much improved' or 'Very much improved'	14	-2	-25, 12	-6.1 (12.10)	0.041	-0.285
		Stable: 'no change'	11	-3	-13, 13	-1.5 (9.11)		-0.070
		Worsened: 'Minimally worse' or 'Much worse' or 'Very much worse'	10	8	-23, 27	6.8 (13.73)		0.354
OLPSSM – total score[a,b]	DLQI Item 1: change from baseline	Improved: >=1 grade improvement	14	-3	-17, 10	-3.1 (6.76)	0.095	-0.450
		Stable: no change	12	0	-15, 11	-0.8 (6.33)		-0.148
		Worsened: >=1 grade worsened	7	4	-5, 10	3.4 (5.06)		0.929
	PGI-S: change from baseline	Improved: >=1 grade improvement	10	-7	-17, 3	-6.6 (6.35)	<0.001	-1.105
		Stable: no change	17	0	-8, 7	0.0 (3.92)		0.000
		Worsened: >=1 grade worsened	6	9	-1, 11	6.2 (5.34)		1.603
	PGI-C (response at week 16)	Improved: 'Minimally improved' or 'Much improved' or 'Very much improved'	10	-7	-17, -1	-7.5 (5.34)	<0.001	-1.336
		Stable: 'no change'	11	0	-8, 10	0.8 (5.23)		0.312
		Worsened: 'Minimally worse' or 'Much worse' or 'Very much worse'	12	3	-3, 11	3.1 (4.27)		0.385

**

Table 14 continued

SD Standard deviation; *Min* Minimum, *Max* Maximum, *DLQI* Dermatology Life Quality Index, *PGI-S* Patient Global Impression of Severity, *PGI-C* Patient Global Impression of Change, *ESS* Epworth Sleepiness Scale, *OLPSSM* Oral Lichen Planus Symptom Severity Measure

[a]Change between baseline and Week 16

[b]Negative values indicate improvement

psychometric evidence for the use of generic HRQoL PROMs in this population. The analyses described in this study evaluated the content validity and psychometric properties of the DLQI, ESS, Scalpdex and OLPSSM to assess appropriateness of use in clinical trials with LP patients. Importantly, the mixed methods approach adopted allows for the patient voice to be represented not only in this study but in future clinical study designs, as recommended by the PFDD guidance documents [28–31] and followed the FDA recommendation for evidence-based rationale when proposing a clinical outcome assessment (COA) as fit for purpose [30]. Specifically, the approach adopted allowed for the assessment of whether the PROMs capture all important aspects of the concept of interest; that the method of scoring is appropriate and sufficiently sensitive to reflect clinically meaningful change within the context of use; that respondents understand the items as intended; that differences in scores can be interpreted in terms of impact on patient's experience and that scores correspond to specific health experiences of patients [30]. The study also included exit interviews, which the FDA have noted as a valuable tool to contribute cumulative evidence on aspects of the patient experience; inform development or refinement of COAs; add greater depth to data in diseases, such as LP, that do not have much qualitative patient input; and to obtain patient input on meaningful outcomes [29].

While the DLQI is one of the most widely used PROM in multiple dermatological indications and has also been commonly used with LP patients [17], content and psychometric evidence of its appropriateness in LP patients for usage in clinical studies is limited [21]. The current study on the one hand supports the use of the DLQI in LP patients, as findings provide strong evidence of reliability and construct validity. The DLQI domain 'Symptoms and feelings' performed particularly well. On the other hand, the psychometric data do not confidently support that the DLQI can detect change over time in the specific context of use for adults with LP as high inter-item correlations between some items suggest potential redundancies. The qualitative interview data further suggest that patients did not consider most items relevant to their disease experience of LP. Given the modular nature of the DLQI, the study data support the use of the 'Symptoms and feelings' domain as an independent module with LP patients, where necessary and appropriate.

Even though the ESS demonstrated evidence of reliability in other populations, convergent validity was poor in this study. Furthermore, known-group comparisons showed evidence of the ESS' ability to discriminate between groups for the PGI-S but not the IGA; ability to detect change was limited or null. These findings suggest that the ESS may not be appropriate for use in clinical trials with LP patients. This is supported by the qualitative findings where most participants reported that they never felt sleepy or wanted to fall asleep because of their LP, although some patients did spontaneously report sleep-related impacts, such as sleep disturbance (i.e., sleep quality and/or sleep quantity). It is suggested that measures that assess sleep rather than daytime sleepiness should be used in clinical studies with LP patients. However, further research is needed to ascertain whether sleep is a meaningful and important concept of LP, as data are scarce [20].

The Scalpdex performed relatively well when psychometrically evaluated in the study's LPP patient sample, demonstrating evidence of internal consistency, test-retest reliability and convergent validity (although only weak

correlations with PGI-C and IGA). There was mixed evidence to differentiate between known groups and to report an ability to detect change. Not all items may be appropriate for use with LPP patients. For example, inter-item correlations for Item 19 and Item 20 were much weaker than the rest of the items, while Item 15 demonstrated weak correlations with the other 'Function' domain items and Item 8 had overall very weak correlations including other 'Symptoms' domain items, which is particularly concerning as the 'Symptoms' domain only consisted of three items. These findings are not surprising as the Scalpdex was originally developed with patients with seborrheic dermatitis and scalp psoriasis [18]. Clinical characteristics present in these patients, such as desquamation and bleeding [23], may not be relevant to LPP patients. This finding is supported by the qualitative CD interviews and the original Scalpdex development study whereby the impact of desquamation, as assessed via Item 15, was reported as not relevant by a high percentage of patients [18]. Based on the study findings, it is suggested that the Scalpdex may be used with caution with LPP patients and that further evidence is needed when it is used in clinical trials. A potential further limitation of the Scalpdex is its length with 23 items that might be viewed as burdensome for many patients, particularly if some items are deemed not relevant. Similar to the DLQI, the Scalpdex 'Symptoms' domain performed better than the measure as a whole, but caution should be taken if the acceptable performance of the measure total score is purely driven by the 'Symptoms' domain-specific items.

Lastly, the OLPSSM, as psychometrically evaluated in MLP patients with oral involvement, had evidence of good reliability, construct validity and ability to detect change over time (PGI-S and PGI-C). It is not surprising that the OLPSSM performed well as it was designed specifically for patients with oral lichen planus and has been previously used within similar populations [8, 38]. However, despite the psychometric validity of this measure, it is worth noting that not all items may be relevant to all patients with oral involvement. For example, Item 4 and Item 5 have been noted in the literature and supported by the qualitative interviews in the current study as triggers least likely to cause soreness and are associated more with patients with severe OLP [8]. Furthermore, inter-item correlations between Item 1 and Item 6 were weak, suggesting that these two items might measure dissimilar concepts while correlations between Item 2 and Item 5 were very high, suggesting potential redundancy. Lastly, the OLPSSM is limited in its use to patients with oral involvement [8, 38], leaving a gap for other LP patients. Overall, the data suggest that the OLPSSM is a valid HRQoL PROM for use with patients with OLP.

Study Limitations

Given the potential limitation of a relatively small sample size of some LP cohorts in the current study, particularly for the OLPSSM and Scalpdex, future research in a larger sample size is recommended to strengthen the findings. Further research is also recommended to review other existing HRQoL measures that may be used in LP patients.

Conclusion

The results of our study contribute to the literature by providing novel insights into the appropriateness of existing PROMs commonly used with LP patients. Our study further highlights the need for additional psychometric evaluation and qualitative evidence to assess whether PROMs under consideration are "fit for purpose" for use in future LP clinical studies and support the development of additional LP specific HRQoL PROMs.

ACKNOWLEDGEMENTS

The authors thank the participants of this study.

Medical Writing, Editorial Assistance and Other Assistance Charlotte Panter, Molly Clifford, Carl Cooper and Sophi Tatlock

contributed to the study design, and study conduct. Jan Stochl and Mike Greenwood contributed to the conduct and review of psychometric analyses. All contributors, except Carl Cooper, are employees of Adelphi Values Ltd. Carl Cooper was an employee of Adelphi Values Ltd. at the time the work was performed.

Author Contributions. Adelphi Values authors (Aoife Mahon-Smith, George Skingley, Lara Ayala-Nunes, Anjali Batish, Rosie Sharp) contributed to the study design, data collection, interpretation of data and preparation of the manuscript for publication. All sponsor authors (Christel Naujoks, Nicolò Compagno, Eva Schruf and Santiago G. Moreno) contributed to defining the scope of the research, including study design, and interpretation of study results in the manuscript.

Funding. Adelphi Values Ltd. were commissioned by Novartis Pharma AG to conduct this research and the sponsor contributed to the study design, data collection and preparation of the manuscript for publication. The sponsor is funding the journal's rapid service fees.

Data Availability. The datasets generated and/or analyzed during the current study are not publicly available in order to protect participant confidentiality.

Declarations

Conflict of Interest. Christel Naujoks and Santiago G. Moreno are employees of Novartis Pharma AG at the time of performing the research and preparing the manuscript. Eva Schruf is an employee of Novartis Pharma GmbH at the time of performing the research and preparing the manuscript. Nicolò Compagno was an employee of Novartis Pharma AG at the time of performing the research and preparing the manuscript and is now an employee of F. Hoffmann-La Roche Ltd, Basel, Switzerland. Aoife Mahon-Smith, Lara Ayala, George Skingley and Rosie Sharp are employees of Adelphi Values Ltd., a health outcomes agency commissioned by Novartis Pharma AG to conduct this research. Anjali Batish was an employee of Adelphi Values at the time of performing the research and is now an employee of the National Institute for Health and Care Research (NIHR). The authors declare that there are no competing interests.

Ethical Approval. All participants provided informed consent indicating their data will be used for medical research purposes and the study results may be published. The studies were performed in accordance with the Helsinki Declaration of 1964 and its later amendments. Ethical approval and oversight for the clinical study, including exit interviews was obtained as part of clinical study procedures (clinicaltrials.gov ID: NCT04300296, EUDRACT: 2019-003588-24). The Western Copernicus Group Independent Review Board (WCG IRB), a centralized IRB, provided ethical approval and oversight to conduct the independent qualitative interviews in the US and Germany.

REFERENCES

1. Le Cleach L, Chosidow O. Lichen planus. N Engl J Med. 2012;366(8):723–32.

2. Weston G, Payette M. Update on lichen planus and its clinical variants. Int J Women's Dermatol. 2015;1(3):140–9.

3. Thandar Y, Maharajh R, Haffejee F, Mosam A. Treatment of cutaneous lichen planus (Part 1): a review of topical therapies and phototherapy. Cogent Med. 2019;6(1):1582467.

4. Gorouhi F, Davari P, Fazel N. Cutaneous and mucosal lichen planus: a comprehensive review of clinical subtypes, risk factors, diagnosis, and prognosis. Sci World J. 2014;2014: 742826.

5. Assouly P, Reygagne P. Lichen planopilaris: update on diagnosis and treatment. Semin Cutan Med Surg. 2009;28(1):3–10.

6. Ingafou M, Leao JC, Porter SR, Scully C. Oral lichen planus: a retrospective study of 690 British patients. Oral Dis. 2006;12(5):463–8.

7. Usatine RP, Tinitigan M. Diagnosis and treatment of lichen planus. Am Fam Physician. 2011;84(1): 53–60.

8. Burke LB, Brennan MT, Ni Riordain R, Madsen LS. Novel oral lichen planus symptom severity measure for assessing patients' daily symptom experience. Oral Dis. 2019;25(6):1564–72.

9. Cassol-Spanemberg J, Blanco-Carrión A, Rodríguez-de Rivera-Campillo ME, Estrugo-Devesa A, Jané-Salas E, López-López J. Cutaneous, genital and oral lichen planus: a descriptive study of 274 patients. Med Oral Patol Oral Cir Bucal. 2019;24(1):e1–7.

10. Lepe K, Nassereddin A, Salazar FJ. Lichen Planopilaris. StatPearls [Internet]. 2021. https://www.ncbi.nlm.nih.gov/books/NBK470325/.

11. Tadakamadla J, Kumar S, Lalloo R, Johnson NW. Qualitative analysis of the impact of oral potentially malignant disorders on daily life activities. PLoS ONE. 2017;12(4): e0175531.

12. Hsu D-Y, Chien W-C, Chung C-H, et al. Risk of anxiety and depression in patients with lichen planus: a nationwide population-based study. J Affect Disord. 2022;300:255–62.

13. Nassab A, Navabi N, Pour M, Charrosta N, Hashemipour M. Quality of life in patients with chronic oral mucosal conditions: a qualitative research. Pesqui Bras Odontopediatria Clín Integrada. 2021;21: e0092.

14. López-Jornet P, Camacho-Alonso F. Quality of life in patients with oral lichen planus. J Eval Clin Pract. 2010;16(1):111–3.

15. Nasimi M, Ahangari N, Lajevardi V, Mahmoudi H, Ghodsi SZ, Etesami I. Quality of life and mental health status in patients with lichen planopilaris based on dermatology life quality index and general health questionnaire-28 questionnaires. Int J Women's Dermatol. 2020;6(5):399–403.

16. US Food and Drug Administration. Patient-Reported Outcome Measures: Use in medical product development to support labeling claims, Silver Spring, ML, 2009.

17. 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.

18. Chen SC, Yeung J, Chren MM. Scalpdex: a quality-of-life instrument for scalp dermatitis. Arch Dermatol. 2002;138(6):803–7.

19. Johns MW. A new method for measuring daytime sleepiness: the Epworth sleepiness scale. Sleep. 1991;14(6):540–5.

20. Adamo D, Ruoppo E, Leuci S, Aria M, Amato M, Mignogna MD. Sleep disturbances, anxiety and depression in patients with oral lichen planus: a case–control study. J Eur Acad Dermatol Venereol. 2015;29(2):291–7.

21. Balcı DD, İnandı T. Dermatology life quality index scores in lichen planus: comparison of psoriasis and healthy controls. Turkderm-Turk Arch Dermatol Venereol. 2008;42(4):127–30.

22. Fiocco Z, Kupf S, Patzak L, et al. Quality of life and psychopathology in lichen planus: a neglected disease burden. Acta Derm Venereol. 2021;101(12): adv00619.

23. Sampogna F, Linder D, Piaserico S, et al. Quality of life assessment of patients with scalp dermatitis using the Italian version of the Scalpdex. Acta Derm Venereol. 2014;94(4):411–4.

24. Chernyshov PV. The evolution of quality of life assessment and use in dermatology. Dermatology. 2019;235(3):167–74.

25. Wiriyakijja P, Fedele S, Porter SR, Mercadante V, Ni RR. Patient-reported outcome measures in oral lichen planus: a comprehensive review of the literature with focus on psychometric properties and interpretability. J Oral Pathol Med. 2018;47(3): 228–39.

26. Ni Riordain R, Hodgson T, Porter S, Fedele S. Validity and reliability of the chronic oral mucosal diseases questionnaire in a UK population. J Oral Pathol Med. 2016;45(8):613–6.

27. US Food and Drug Administration. Patient-Focused Drug Development: Methods to identify what is important to patients guidance for industry, food and drug administration staff, and other stakeholders. Silver Spring, MD. 2022.

28. US Food and Drug Administration. Patient-Focused Drug Development: Collecting comprehensive and representative input - guidance for industry, food and drug administration staff, and other stakeholders. In: (FDA) UFaDA, ed2020.

29. US Food and Drug Administration. Patient-Focused Drug Development: Methods to identify what is important to patients, guidance for industry, food and drug administration staff, and other stakeholders (Guidance 2). 2022.

30. US Food and Drug Administration. Patient-Focused Drug Development: Selecting, developing, or modifying fit-for-purpose clinical outcome assessments (draft guidance). In: Services USDoHaH, ed2022.

31. US Food and Drug Administration. Guidance for industry: Patient-Reported Outcome Measures: Use in medical product development to support labeling claims. 2009.

32. Willis GB. Cognitive interviewing: a tool for improving questionnaire design. Sage Publications; 2004.

33. ATLAS.ti Scientific Software Development GmbH [ATLAS.ti 22 Windows]. Retrieved from https://atlasti.com [computer program]. 2022.

34. Hsieh H, Shannon S. Three approaches to qualitative content analysis. Qual Health Res. 2005;15(9):1277–88.

35. Laga AC, Haefner HK, Granter SR. Noninfectious inflammatory disorders of the vulva. In: Diagnostic gynecologic and obstetric pathology. Elsevier; 2018. p. 22–52.

36. Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. J Chiropr Med. 2016;15(2):155–63.

37. Kazis LE, Anderson JJ, Meenan RF. Effect sizes for interpreting changes in health status. Med Care. 1989;27(3 Suppl):S178-189.

38. Brennan MT, Madsen LS, Saunders DP, et al. Efficacy and safety of a novel mucoadhesive clobetasol patch for treatment of erosive oral lichen planus: a phase 2 randomized clinical trial. J Oral Pathol Med. 2022;51(1):86–97.

39. Nunnally J, Bernstein I. Psychometric theory. 3rd ed. New York: McGraw-Hill Inc.; 1994.

40. Cohen J. Statistical power analysis for the behavioral sciences. Routledge; 2013.

41. Hedges LV. Distribution theory for Glass's estimator of effect size and related estimators. J Educ Stat. 1981;6(2):107–28.