Pruritus and Symptoms Assessment for Atopic Dermatitis (PSAAD): Development in Adolescents and Adults and Validation in Adults with Moderate-to-Severe AD

Rebecca Hall,[1] Mark G. Lebwohl,[2] Andrew G. Bushmakin,[3] Eric L. Simpson,[4] Melinda J. Gooderham,[5,6] Andreas Wollenberg,[7] Adam Gater,[1] Jane R. Wells,[1] Joseph C. Cappelleri,[8] Ming-Ann Hsu,[9] Jocelyn Papacharalambous,[10] Elena Peeva,[11] Anna M. Tallman,[10] Weidong Zhang,[12] Linda Chen[10]

[1]Patient-Centered Outcomes, Adelphi Values, Adelphi Mill, Bollington, Cheshire, UK

[2]Kimberly and Eric J. Waldman Department of Dermatology, Icahn School of Medicine at Mount Sinai, New York, NY, USA

[3]PIH Statistics, Pfizer Inc., Groton, CT, USA

[4]Department of Dermatology, Oregon Health and Science University, Portland, OR, USA

[5]SKiN Centre for Dermatology and Probity Medical Research, Peterborough, ON, Canada

[6]Queen's University, Kingston, ON, Canada

[7]Department of Dermatology and Allergy, Ludwig-Maximilians-Universität München, Munich, Germany

[8]Global Biometrics & Data Management, Pfizer Inc., Groton, CT, USA

[9]PIH – Global Health & Value, Pfizer Inc., Groton, CT, USA

[10]Formerly Pfizer Inc., New York, NY, USA

[11]R&D – Inflammation & Immunology, Pfizer Inc., Cambridge, MA, USA

[12]Formerly Pfizer Inc., Cambridge, MA, USA

Running head: PSAAD development and validation in adults

Correspondence: Rebecca Hall

Adelphi Mill, Grimshaw Lane, Bollington, Cheshire, SK10 5JB, UK

rebecca.hall@adelphivalues.com

Word count: 2987 (maximum, 4000; excluding abstract, references, figures, and tables)

Tables/figures: 3/3 (limit: 6 total)

27 **Article type:** Original Research Article

28

Abstract (242 words; maximum, 250)

Background Most patient-reported outcome (PRO) instruments that measure atopic dermatitis (AD) symptoms do not have sufficient documented evidence of content validity to satisfy regulatory agency guidance for inclusion in product-labelling claims in the United States or Europe.

Objective To develop a PRO instrument in accordance with regulatory agency guidance to assess daily AD symptoms during the course of therapy and to establish its content validity and psychometric properties.

Methods The Pruritus and Symptoms Assessment for Atopic Dermatitis (PSAAD) daily diary was developed based on qualitative interviews with US adolescents and adults with mild-to-severe AD. Content validity, test–retest reliability, internal consistency reliability, clinically important difference, clinically important responder, convergent validity and known-group validity were evaluated using correlational and regression methods from a phase 2b data from US adults with moderate-to-severe AD who were treated with abrocitinib.

Results Patient interviews conducted with US adolescents and adults with mild-to-severe AD identified 11 relevant symptoms (itch, dryness, redness, flaking, discolouration, pain, bleeding, cracking, bumps, swelling and weeping/oozing) for inclusion in the PSAAD instrument. All PSAAD psychometric parameters were acceptable based on phase 2b data from US adults with moderate-to-severe AD. Convergent validity and known-group validity were confirmed by significant correlations between PSAAD and six other PRO measures (r = 0.24–0.91, all $p \leq$ 0.01) and Dermatology Life Quality Index category ($p \leq 0.0001$), respectively.

Conclusions Evidence supports the PSAAD instrument validity, reliability, responsiveness and definitions of clinically important changes/differences for adults with moderate-to-severe AD.

Keywords: atopic dermatitis, eczema, patient-reported outcomes, pruritus, daily diary

50

1 Introduction

Among skin diseases, atopic dermatitis (AD) is associated with a major burden of disease [1], and a significant proportion of patients with AD have inadequately controlled disease despite treatment [2]. Patient-reported severity of AD is often incongruous with physician-reported severity, with physicians frequently underestimating the severity of disease [3-5]. Patient-reported symptoms are among the set of core outcome measures recommended by the international Harmonising Outcome Measures for Eczema initiative [6, 7].

Most patient-reported outcome (PRO) instruments that measure AD symptoms do not provide a comprehensive assessment of all symptoms important to patients or do not have documented evidence of content validity (see Table S1 in Online Resource 1) for definitions of psychometric terms) that would be considered sufficient by the United States Food and Drug Administration (FDA) or the European Medicines Agency (EMA) as a clinical trial endpoint to support product-labelling claims. This report details the development of a patient-reported symptom diary in accordance with FDA [8] and EMA [9] PRO guidance using qualitative interviews with adolescents and adults with mild-to-severe AD and evaluation of its psychometric properties using data from a phase 2b study in adults with moderate-to-severe AD [10].

2 Methods

This research was reviewed and approved by the institutional review board at each study site , it and was conducted in accordance with the Helsinki Declaration of 1964 and its later amendments.

2.1 PSAAD Content Development

A review of the literature and of online patient blogs/forums (search of MEDLINE, Embase and PsycINFO as well as Google Scholar and online patient blogs/forums relating to AD and dermatological conditions), physician input and concept elicitation patient interviews were used to identify relevant AD symptoms and the language used by patients to talk about them (see Fig. S1 in Online Resource 1). Based on these findings, a draft 13-item daily diary was developed for completion via an electronic handheld device.

Thirty participants recruited from general practitioner or dermatologist offices in the United States were included in the concept elicitation interviews. Approximately 10 interviews were conducted for each age group (12-14 years, 15-17 years and ≥18 years) to achieve conceptual saturation (i.e. the point at which no new concepts are likely to be

elicited in further interviews) [11-13]. Recruitment quotas were used to ensure adequate representation across sexes, physician- and patient-rated disease severity, racial and ethnic groups and educational achievement (adults only).

To be eligible for interview, patients had to be aged ≥12 years and have a clinical diagnosis of AD (using Hanifin and Rajka criteria [14]), affected percentage of body surface area (%BSA) 2 to 40 (excluding scalp with %BSA ≥2 on body regions other than the palms and the soles) and physician-rated mild, moderate or severe AD. Patients with contact or seborrheic dermatitis; discoid, gravitational/stasis, asteatotic or dyshidrotic eczema; psoriasis; or viral, fungal or bacterial infection were excluded.

Patients participated in two semistructured face-to-face interviews, each lasting approximately 1 hour. Interviewers were experienced in conducting interviews with adolescents and adults and were trained in the use of the interview guide and the electronic diary device. The first interview was designed to explore symptoms experienced by patients (i.e. concept elicitation) through open-ended questions, followed by more probing questions to explore concepts either not mentioned spontaneously or warranting further exploration/clarification. After the first interview, patients completed the draft 13-item daily diary at home once daily for 7 days using a supplied electronic device. The device included an alarm to remind patients to complete the diary each evening within the designated completion window. A second interview was then conducted to evaluate comprehension and relevance of diary content and user acceptability of the electronic instrument (i.e. cognitive debriefing). Interviews were audio-recorded and transcribed verbatim for analysis. Interviews were conducted over two rounds. Updates made to the instrument based on first-round feedback were tested in the second round (Fig. S1 in Online Resource 1).

2.2 PSAAD Psychometric Validation/Quantitative Evaluation in a Phase 2b Clinical Trial

Psychometric evaluation of the PSAAD was performed using data from adults in the United States with moderate-to-severe AD included in a phase 2b study of abrocitinib (NCT02780167) [10]. Accepted methods for psychometric and quantitative evaluation were applied [15-17]. Test–retest reliability was assessed using intraclass correlation coefficient (ICC; with a one-way random effects model), defined as between-patient variability divided by total variability (i.e. between-patient variability plus within-patient variability) [16] using pretreatment data collected for ≥7 days during the screening period. ICC values were considered acceptable if ≥0.70 [18] and excellent if >0.9 [19]. Although patients completed the PSAAD daily, single measurements had acceptable test–retest reliability so internal

consistency reliability was evaluated using Cronbach's coefficient alpha and corrected item-to-total correlations based on data from days –1, 1 (baseline), 8, 15, 29, 43, 57, 85, 92, 99 and 113. Acceptability criteria for Cronbach's coefficient alpha and corrected item-to-total correlations were ≥0.70 [18] and ≥0.40, respectively [20].

Convergent validity was assessed using Pearson correlation coefficients (*r*) between PSAAD and other measures, including pruritus numeric rating scale (NRS; assesses the severity/frequency of itching over the previous 24 hours from no/never itching [0] to worst possible/always or constantly itching [10]), patient global assessment (PtGA; evaluates overall cutaneous disease at time of assessment on 5-point Likert scale ranging from clear [0] to severe [4]), patient global impression of severity (PGIS; daily 11-category scale to assess AD severity over the previous 24 hours, ranging from not present [0] to extremely severe [10]), patient global impression of change (PGIC; weekly 7-category scale to evaluate change in AD severity from baseline (ranging from much better [1] to much worse [7]), Dermatology Life Quality Index (DLQI), Patient-Oriented Eczema Measure (POEM), Investigator's Global Assessment (IGA), Eczema Area and Severity Index (EASI), %BSA and SCORing of AD (SCORAD). Correlation coefficients ≥0.40 were considered supportive of convergent validity; those between 0.30 and 0.40 indicated no evidence for convergent or divergent validity, and those <0.30 indicated divergent validity [16]. Correlations between PSAAD and pruritus NRS, PtGA, IGA, EASI, %BSA or SCORAD were calculated using the average of daily scores from days 1, 8, 15, 29, 43, 57 and 85. Correlation between PSAAD and PGIS was calculated using the average of daily scores from day 1 to day 88. Correlation between PSAAD and PGIC was based on the change from baseline in weekly average of daily PSAAD scores and weekly PGIC scores from week 1 to week 12. Correlation between PSAAD and POEM was based on weekly average of daily PSAAD scores and weekly POEM score for weeks 0, 1, 2, 4, 6, 8 and 12.

PGIS and PGIC are anchors that are recommended by the FDA, along with relevant well-established clinical outcomes, to calculate a clinically meaningful change in a new patient-reported outcome [21]. A clinically important difference (CID; difference between treatment groups considered clinically relevant) threshold in PSAAD total score was estimated by assessing the relationship between PSAAD total score and PGIS using a repeated-measures model and data from the 12-week double-blind part of the phase 2b study (up to day 88). PGIS was assessed daily using an 11-category scale to assess AD severity over the previous 24 hours (not present [0] to extremely severe [10]). Empirical research and historical precedent indicate that a 7-point Likert scale is preferred for important difference calculations [22, 23]. Based on this, CID was defined as the difference in mean PSAAD total score corresponding to

a 1.7-point difference in PGIS (i.e. 10 divided by 6, where 6 is the number of pairwise adjacent categories in PSAAD compared with that in PGIS). Sensitivity analyses for CID were performed using a repeated-measures model to estimate the relationship between PSAAD scores and PGIC and the relationship between PSAAD scores and POEM total scores (assuming that the CID of 3.4 points for POEM [24] would correspond to the CID for PSAAD). These relationships were analysed using PGIS, PGIC and POEM total score each as a continuous anchor (which imposed a linear relationship between outcome and anchor) and as a categorical anchor (which did not impose any functional relationship between outcome and anchor).

Clinically important response (CIR; within-patient change considered clinically relevant according to 'responder' criteria) threshold in PSAAD total score was examined with regard to the relationship between change in PSAAD and Subject Global Impression of Change (SGIC) by a repeated-measures model. SGIC is based on PGIC using the following algorithm: PGIC ≤3, SGIC = 1 (better); PGIC = 4, SGIC = 0 (the same); PGIC ≥5, SGIC = –1 (worse). Difference in change in mean PSAAD score corresponding to a 1-category difference in SGIC was used to define CIR. Standardised effect sizes of CID and CIR for PSAAD total score were obtained by dividing CID and CIR estimates by the standard deviation (SD) of baseline PSAAD total score. Criteria for the impact of an intervention in terms of effect sizes were: 0.2, 'small'; 0.5, 'medium'; 0.8, 'large' [17, 25].

With a repeated measures longitudinal model, known-group validity was determined by examining the relationship between PSAAD and DLQI, a dermatology-specific measure of health-related quality of life that is validated in dermatology clinical trials according to EMA standards [26], and calculating the mean difference in PSAAD between patients with 'no effect at all on patient's life' (DLQI = 0 or 1) and those with at least a 'small effect on patient's life' (DLQI ≥ 2).

3 Results

3.1 PSAAD Development/Qualitative Evaluation of Content Validity

Iterative (repeated) concept elicitation and cognitive debriefing interviews were conducted with 30 adolescents and adults in the United States with mild-to-severe AD (round 1, n = 14; round 2, n = 16). Their disease characteristics were consistent with those of the overall adolescent and adult AD patient population in the United States and included the full range of AD severities and an adequate representation of lower education levels (Table S2 in Online Resource 1).

A review of the literature and patient forums/blogs identified itch (pruritus), dryness (xerosis), redness (erythema), flaking, discolouration, pain (soreness, burning, stinging), bleeding, cracking, swelling/inflammation (oedema), weeping/oozing (fluid/exudate), tightness and thickening as symptoms experienced by patients with AD. Concept elicitation interviews identified the terminology used by patients for AD symptoms and confirmed the relevance of all but two of these symptoms to patient reporting (relevant: itch, dryness, redness, flaking, discolouration, pain, bleeding, cracking, swelling, fluid; not as relevant: tightness, thickening) and identified an additional symptom (bumps) (Fig. 1). Conceptual saturation was achieved across the concept elicitation interviews (Fig. 1).

Itch was by far the most relevant symptom, with all 30 patients reporting it spontaneously. Itch was also reported as the most frequent, most severe and most bothersome symptom. Skin thickening and skin tightening were not considered important symptoms because they were rarely (if at all) mentioned by patients unless probed. Furthermore, more than half the patients did not report skin thickening or skin tightening items as relevant (57% for each); therefore, these symptoms were not included in the final PSAAD. All other symptoms, except for fluid (exudate), were reported by at least half the patients.

Most of the 11 symptoms included in the PSAAD were reported with similar frequency by adults and adolescents, except for fluid and cracking, which were reported slightly more frequently by adult patients. All 11 symptoms were reported across the spectrum of AD severities. Skin dryness, itching and redness were reported by patients as the most frequent symptoms, whereas pain, weeping, itching and bleeding were reported as the most bothersome.

Feedback during cognitive debriefing interviews indicated that instructions, items and response options were consistently interpreted and appeared to be well understood by participants. Completion rates were good, and there were few skipped items or missing days; 57% of patients completed the diary every day during the 7-day period, and the mean number of completions was 6. The majority of patients found the personalised alarm useful or essential to remind them to fill in the diary each day. Patients reported being able to successfully complete the daily diary using the electronic device; the mean time for daily completion was 2 minutes 39 seconds.

3.2 PSAAD Instrument

The final PSAAD is an 11-item instrument designed to provide a comprehensive assessment of symptom severity over the previous 24 hours in adults (aged ≥18 years) and adolescents (aged 12-17 years) with diagnoses of mild-to-

184 severe AD (see www.pfizerpatientreportedoutcomes.com for further information). Each item of the PSAAD assesses
185 the severity of a single symptom on an 11-point NRS, ranging from 0 (none) to 10 (extreme), and contributes
186 equally to the PSAAD total score as depicted in the conceptual framework (Fig. 2). The PSAAD total score is
187 calculated as the average of the responses to each of the 11 items, for a PSAAD total score range of 0 (none) to 10
188 (extreme).

189 **3.3 PSAAD Psychometric Validation/Quantitative Evaluation**

190 The psychometric evaluation of the PSAAD was based on data from adult patients in the United States with
191 moderate-to-severe AD who were enrolled in a phase 2b study for abrocitinib, involving 12 weeks of treatment and
192 a 4-week follow-up period (Table S3, Online Resource 1); 81% of patients completed the PSAAD on >70% of days
193 in the phase 2b study. Test–retest reliability of a single measurement was acceptable with ICC >0.7 (Table 1).
194 Internal consistency reliability was excellent with Cronbach coefficient alpha >0.9 at every time point (Table 1; see
195 Table S4 in Online Resource 1). Convergent validity was confirmed by substantial correlations in the expected
196 direction between PSAAD and other measures (Table 2) ($p \leq 0.01$ for all).

197 Based on anchors PCIS and PGIC, the CID and CIR of PSAAD total score were estimated to be 0.63 and 1.0 points,
198 respectively, which represent approximately 'small' and 'medium' effect sizes of 0.28 and 0.45 (Table 1). The
199 PGIC- and POEM-based estimates of CID (0.65 and 0.64, respectively) were in agreement with the estimate based
200 on PCIS. The close relationship demonstrated between PSAAD total score as a function of PGIS, PGIC or POEM
201 total score as continuous and as categorical anchors supports the linearity assumption in the main CID model
202 (Fig. 3).

203 A positive relationship between PSAAD and DLQI was evident (see Fig. S2 in Online Resource 1), with differences
204 in PSAAD between groups with 'no effect at all on patient's life' (DLQI = 0 or 1) and 'small to extremely large
205 effect on patient's life' (DLQI ≥2) all greater than the CID (0.63) and all statistically significant ($p \leq 0.0001$) (Table
206 3). More severe symptoms according to PSAAD were associated with greater deficits in quality of life according to
207 the DLQI, with DLQI total scores of 0-1 ('no effect'), 2-5 ('small effect'), 6-10 ('moderate effect'), 11-20 ('very
208 large effect') and 21-30 ('extremely large effect'), corresponding to PSAAD overall scores of approximately 2.6,

209 3.3, 4.2, 5.2 and 5.9, respectively (see Fig. S2 in Online Resource). This supports the clinical relevance of the
210 changes observed and the known-group validity of the PSAAD.

211 **4 Discussion**

212 Concept elicitation and conceptual saturation results indicate that the PSAAD captures all the symptoms of AD
213 considered important by patients. Cognitive debriefing interviews confirmed comprehension and relevance of the
214 instrument content among a diverse sample of adolescents and adults with AD in terms of age, sex and physician-
215 rated AD severity (mild to severe). Patient samples were ethnically and racially diverse across black, white,
216 multiracial and other groups in both the qualitative and the quantitative phases. This ensures broad applicability of
217 the measure. However, future work may be needed to evaluate the instrument in younger patients and/or patients
218 living outside the United States.

219 Of note, this analysis defined both the between-group difference and the within-patient change considered to be
220 clinically relevant (CID and CIR, respectively). Although many clinical trials use the former to evaluate treatment
221 effects, which remains important, the FDA has been placing an emphasis on the latter because it represents a
222 meaningful change from the patient perspective [21].

223 Unlike POEM and other more recently developed PROs (ADerm SS, Itch Numeric Rating Scale [v2.0], Skin Pain
224 Numeric Rating Scale [v2.0b] and Peak Pruritus Numerical Rating Scale), the PSAAD provides a comprehensive
225 assessment of AD symptom severity over the previous 24 hours for all symptoms considered important by adults
226 and adolescents with mild-to-severe AD. Furthermore, PSAAD was developed to meet regulatory guidance and—
227 unlike POEM, Itch Numeric Rating Scale (v2.0), Skin Pain Numeric Rating Scale (v2.0b) and Peak Pruritus
228 Numerical Rating Scale—to be included in product-labelling claims in the United States and Europe. These results
229 confirm previous research that itch is a central feature of AD from the patient perspective [27]. Itch was the only
230 symptom reported by all 30 interviewees, all of whom reported it spontaneously, and it was reported by patients in
231 interviews as the most frequent, most severe and most bothersome symptom. Skin dryness and redness were
232 reported by almost all patients, with approximately two-thirds reporting them spontaneously. Although thickening is
233 an important clinical feature associated with AD [28], it was only reported by patients when probed and was

considered not relevant by a majority of patients. Therefore, thickening was not included in the final 11-item PSAAD instrument.

By their nature, patient-reported AD symptoms such as itch are subjective; however, evidence from the qualitative interviews and the phase 2 study supports the reliability, content and construct validity, the definitions of clinically important changes and the use of the PSAAD for assessing symptom severity in adults with moderate-to-severe AD in the United States. As expected, the PSAAD correlates well with POEM, SCORAD and other measures of AD severity, which include a patient-reported subjective assessment of pruritus, but not as well with clinician-assessed objective measures such as EASI, IGA and %BSA, which do not. The lower correlations with EASI, IGA and %BSA may be indicative of divergent validity or lack of evidence to dismiss either convergent validity or divergent validity [16]. Furthermore, the relationship observed between PSAAD and DLQI confirms the substantial detrimental effects of pruritus and other AD symptoms on quality of life.

The PSAAD is a valuable tool for assessing the severity of AD symptoms in clinical studies and perhaps in clinical practice. It demonstrates sufficient validity to be included as an endpoint in clinical trials to support product-labelling claims.

Acknowledgements

We thank the participants of the study.

Funding: This study was sponsored by Pfizer Inc., New York, NY, USA.

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

Medical writing support: Medical writing support under the guidance of the authors was provided by Marianna Johnson, PhD, and Jennifer C. Jaworski, MS, at ApotheCom, San Francisco, CA, USA, and was funded by Pfizer Inc., New York, NY, USA, in accordance with Good Publication Practice (GPP3) guidelines (*Ann Intern Med*. 2015;163:461-464).

Conflicts of Interest: R. Hall, A. Gater and J.R. Wells are employees of Adelphi Values, which has received payment from Pfizer Inc. for conducting the instrument development research activities described in this paper.

M.G. Lebwohl is an employee of Mount Sinai, which receives research funds from Pfizer Inc., AbbVie, Bausch Health (Valeant), Boehringer Ingelheim, Celgene, Eli Lilly, Incyte, Janssen/Johnson & Johnson, LEO Pharma, Medimmune/AstraZeneca, Novartis, SCIderm, UCB and Vidac Pharma. He is also a consultant for Allergan, Almirall (Aqua Pharmaceuticals), Boehringer Ingelheim, Dr. Reddy's Laboratory (Promius), LEO Pharma, Menlo Therapeutics and Verrica.

A.G. Bushmakin, J.C. Cappelleri, M.-A. Hsu and E. Peeva are employees and stockholders of Pfizer Inc.

E.L. Simpson has received grants, personal fees and/or nonfinancial support from Pfizer Inc., AbbVie, Celgene, Eli Lilly, Galderma, GSK, LEO Pharma, Menlo Therapeutics, Novartis, Regeneron, Tioga Pharmaceuticals and Vanda Pharmaceuticals.

M.J. Gooderham has received grants, personal fees and/or nonfinancial support from Pfizer Inc., AbbVie, Akros Pharma, Amgen, Arcutis, Bausch Health (Valeant), Boehringer Ingelheim, Bristol-Myers Squibb, Celgene, Coherus, Dermira, Eli Lilly, Galderma, Incyte, Janssen, Kyowa Kirin, LEO Pharma, MedImmune, Merck, Novartis,

Regeneron, Roche, Sanofi Genzyme, Sun Pharma and UCB.

A. Wollenberg has received grants, personal fees and/or nonfinancial support from Pfizer Inc., Almirall, Anacor, Astellas, Beiersdorf, Bioderma, Celgene, Chugai, Galderma, Hans Karrer, L'Oréal, LEO Pharma, Meda, MedImmune, Merck, Novartis, Pierre Fabre, Regeneron and Sanofi.

J. Papacharalambous, A.M. Tallman, W. Zhang and L. Chen were employees and stockholders of Pfizer Inc. at the time of this research.

Author Contributions: RH, MGL, AGB, AG, JRW, JCC, MH, JP, EP, AMT, WZ, and LC contributed to the conceptualization of the study. RH, AGB, AW, AG, JRW, JCC, MH, EP, WZ, LC performed or supported the formal analysis. All authors (RH, MGL, AGB, ELS, MJG, AW, AG, JRW, JCC, MH, JP, EP,AMT, WZ, and LC) contributed to the data interpretation, critically reviewed the manuscript and approved the submitted version.

Declarations

Ethics statement : This research was reviewed and approved by the institutional review board at each study site. It was conducted in accordance with the Helsinki Declaration of 1964 and its later amendments.

Written informed consent was provided by participants of all studies.

Data Availability: The datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request. Upon request, and subject to certain criteria, conditions and exceptions (see https://www.pfizer.com/science/clinical-trials/trial-data-and-results for more information), Pfizer will provide access to individual de-identified participant data from Pfizer-sponsored global interventional clinical studies conducted for medicines, vaccines, and medical devices (1) for indications that have been approved in the US and/or EU or (2) in programs that have been terminated (ie, development for all indications has been discontinued). Pfizer will also consider requests for the protocol, data dictionary, and statistical analysis plan. Data may be requested from Pfizer trials 24 months after study completion. The de-identified participant data will be

made available to researchers whose proposals meet the research criteria and other conditions, and for which an exception does not apply, via a secure portal. To gain access, data requestors must enter into a data access agreement with Pfizer

References

1. Hay RJ, Johns NE, Williams HC, et al. The global burden of skin disease in 2010: an analysis of the prevalence and impact of skin conditions. J Invest Dermatol. 2014;134:1527–34.
2. Wei W, Anderson P, Gadkari A, et al. Extent and consequences of inadequate disease control among adults with a history of moderate to severe atopic dermatitis. J Dermatol. 2018;45:150–7.
3. Wei W, Anderson P, Gadkari A, et al. Discordance between physician- and patient-reported disease severity in adults with atopic dermatitis: a US cross-sectional survey. Am J Clin Dermatol. 2017;18:825–35.
4. Magin PJ, Pond CD, Smith WT, et al. Correlation and agreement of self-assessed and objective skin disease severity in a cross-sectional study of patients with acne, psoriasis, and atopic eczema. Int J Dermatol. 2011;50:1486–90.
5. Torrelo A, Ortiz J, Alomar A, et al. Atopic dermatitis: impact on quality of life and patients' attitudes toward its management. Eur J Dermatol. 2012;22:97–105.
6. Chalmers JR, Schmitt J, Apfelbacher C, et al. Report from the third international consensus meeting to harmonise core outcome measures for atopic eczema/dermatitis clinical trials (HOME). Br J Dermatol. 2014;171:1318–25.
7. Schmitt J, Spuls P, Boers M, et al. Towards global consensus on outcome measures for atopic eczema research: results of the HOME II meeting. Allergy. 2012;67:1111–7.
8. US Department of Health and Human Services. Guidance for Industry: patient-reported outcome measures: use in medical product development to support labeling claims. https://www.fda.gov/downloads/drugs/guidances/ucm193282.pdf. Accessed 17 Jul 2020.
9. European Medicines Agency. Reflection paper on the regulatory guidance for the use of health-related quality of life (HRQL) measures in the evaluation of medicinal products. https://www.ema.europa.eu/documents/scientific-guideline/reflection-paper-regulatory-guidance-use-healthrelated-quality-life-hrql-measures-evaluation_en.pdf. Accessed 17 Jul 2020.
10. Gooderham MJ, Forman SB, Bissonnette R, et al. Efficacy and safety of oral Janus kinase 1 inhibitor abrocitinib for patients with atopic dermatitis: a phase 2 randomized clinical trial. JAMA Dermatol. 2019;155:1371–9.
11. Guest G, Bunce A, Johnson L. How many interviews are enough? Field Methods. 2016;18:59–82.

12. Francis JJ, Johnston M, Robertson C, et al. What is an adequate sample size? Operationalising data saturation for theory-based interview studies. Psychol Health. 2010;25:1229–45.
13. Turner-Bowker DM, Lamoureux RE, Stokes J, et al. Informing a priori sample size estimation in qualitative concept elicitation interview studies for clinical outcome assessment instrument development. Value Health. 2018;21:839–42.
14. Hanifin JM, Rajka G. Diagnostic features of atopic dermatitis. Acta Derm Venereol. 1980;60:44–7.
15. US Department of Health for Drug and Human Services. Guidance for Industry and FDA Staff: qualification process for drug development tools. https://www.fda.gov/downloads/drugs/guidances/ucm230597.pdf. Accessed 17 Jul 2020.
16. Cappelleri J, Zou K, Bushmakin A, et al. Patient-Reported Outcomes: Measurement, Implementation and Interpretation. Boca Raton, FL: Chapman & Hall/CRC Press; 2013.
17. Cappelleri JC, Bushmakin AG. Interpretation of patient-reported outcomes. Stat Methods Med Res. 2014;23:460–83.
18. Chassany O, Sagnier P, Marquis P, et al. Patient-reported outcomes: the example of health-related quality of life—a European guidance document for the improved integration of health-related quality of life assessment in the drug regulatory process. Drug Info J. 2016;36:209–38.
19. Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. J Chiropr Med. 2016;15:155–63.
20. Ware JE Jr, Gandek B. Methods for testing data quality, scaling assumptions, and reliability: the IQOLA Project approach. International Quality of Life Assessment. J Clin Epidemiol. 1998;51:945–52.
21. US Food and Drug Administration. Discussion Document for Patient-Focused Drug Development Guidance Public Workshop on Guidance 3: select, develop or modify fit-for-purpose clinical outcomes assessments. Workshop date: October 15-16, 2018. https://www.fda.gov/media/116277/download. Accessed 17 Jul 2020.
22. Norman GR, Sloan JA, Wyrwich KW. Interpretation of changes in health-related quality of life: the remarkable universality of half a standard deviation. Med Care. 2003;41:582–92.
23. Norman GR, Sloan JA, Wyrwich KW. The truly remarkable universality of half a standard deviation: confirmation through another look. Expert Rev Pharmacoecon Outcomes Res. 2004;4:581–5.

24. Schram ME, Spuls PI, Leeflang MM, et al. EASI, (objective) SCORAD and POEM for atopic eczema: responsiveness and minimal clinically important difference. Allergy. 2012;67:99–106.
25. Cohen J. Statistical Power Analysis for the Behavioral Sciences. 2nd ed. Hillsdale, NJ: Lawrence Erlbaum Associates; 1988.
26. European Medicines Agency. Guideline on clinical investigation of medicinal products indicated for the treatment of psoriasis. https://www.ema.europa.eu/en/documents/scientific-guideline/guideline-clinical-investigation-medicinal-products-indicated-treatment-psoriasis_en.pdf. Accessed 17 Jul 2020.
27. Charman CR, Venn AJ, Williams HC. The patient-oriented eczema measure: development and initial validation of a new tool for measuring atopic eczema severity from the patients' perspective. Arch Dermatol. 2004;140:1513–9.
28. Eichenfield LF, Tom WL, Chamlin SL, et al. Guidelines of care for the management of atopic dermatitis: section 1. Diagnosis and assessment of atopic dermatitis. J Am Acad Dermatol. 2014;70:338–51.
29. Finlay AY, Khan GK. Dermatology Life Quality Index (DLQI)—a simple practical measure for routine clinical use. Clin Exp Dermatol. 1994;19:210–6.

TABLES

Table 1 Psychometric validation parameters for the PSAAD diary

Parameter	Acceptability criteria	Actual
Test–retest reliability		
Intraclass correlation coefficient	>0.9 excellent, 0.7-0.9 acceptable, <0.7 inadequate	0.81 (for a single measurement)
Internal consistency reliability		
Cronbach coefficient alpha	≥0.70 acceptable	>0.9 (every time point)
Corrected item-to-total correlations	≥0.40 acceptable	>0.5 (every time point)
CID	—	0.63
Effect size, SD units	0.80 large, 0.5 medium, 0.2 small	0.28
CIR	—	1.03
Effect size, SD units	0.80 large, 0.5 medium, 0.2 small	0.45

CID clinically important difference, *CIR* clinically important response, *PSAAD* Pruritus and Symptoms Assessment for Atopic Dermatitis, *SD* standard deviation

Table 2 Convergent validity: correlations between PSAAD diary and other measures

	Pearson correlation coefficient with PSAAD, *r*
Pruritus NRS[a]	0.82
PtGA[a]	0.70
PGIS[b]	0.91
PGIC[c]	0.68
DLQI[d]	0.67
POEM[d]	0.82
IGA[a]	0.38
EASI[a]	0.37
%BSA[a]	0.24
SCORAD[a]	0.60

%BSA percentage of body surface area, *DLQI* Dermatology Life Quality Index, *EASI* Eczema Area and Severity Index, *IGA* Investigator's Global Assessment, *NRS* numeric rating scale, *PGIC* Patient Global Impression of Change, *PGIS* Patient Global Impression of Severity, *POEM* Patient-Oriented Eczema Measure, *PSAAD* Pruritus and Symptoms Assessment for Atopic Dermatitis, *PtGA* Patient Global Assessment, *SCORAD* SCORring Atopic Dermatitis

All correlations were calculated based on means of available data (see footnotes).

Correlation coefficients ≥0.40 were considered supportive of convergent validity, those between 0.30 and 0.40 indicated no evidence for convergent or divergent validity and those <0.30 indicated divergent validity [15]

p values < 0.01 for all

[a]Average of daily scores for days 1, 8, 15, 29, 43, 57 and 85 for both variables

[b]Average of daily scores from day 1 to day 88 for both variables

[c]Change from baseline in weekly average of daily PSAAD scores versus weekly PGIC scores from week 1 to week 12

393 [d]Weekly average of daily PSAAD scores versus weekly POEM score for weeks 0, 1, 2, 4, 6, 8 and 12

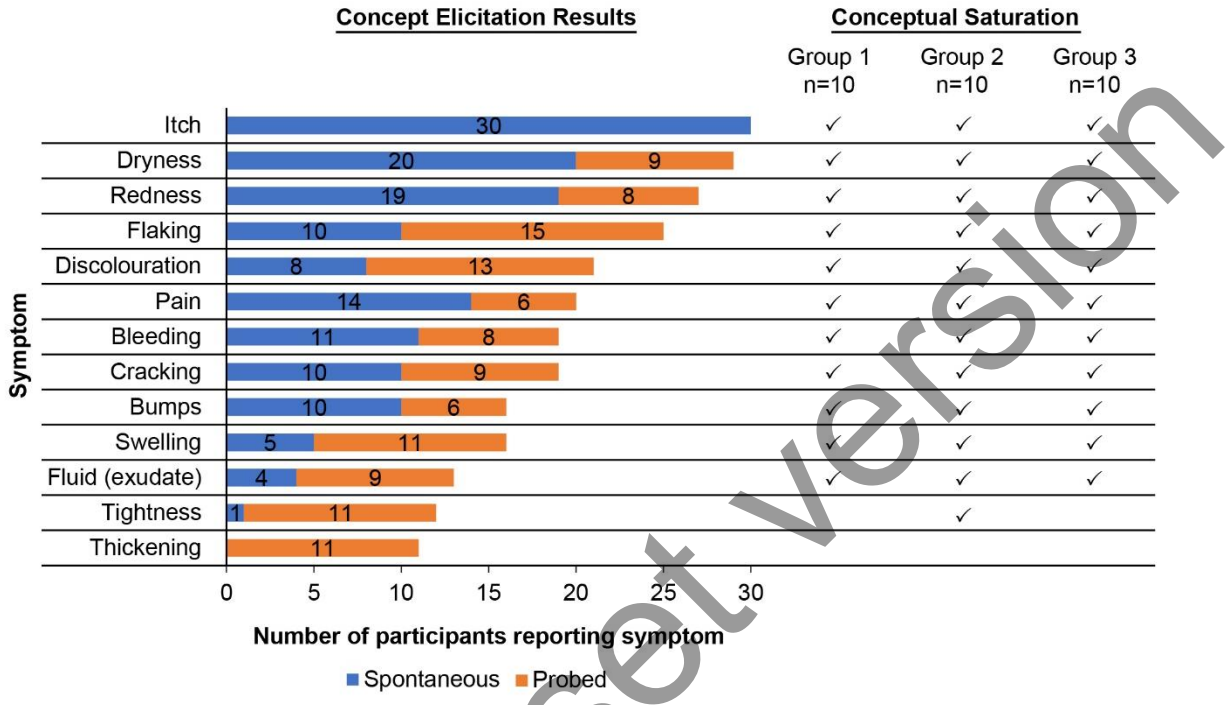
Table 3 Known-group validity

DLQI range [29]	Mean difference in PSAAD total score versus reference	*p* value
0 to 1: no effect at all on patient's life	Reference	N/A
2 to 5: small effect on patient's life	–0.7268	≤ 0.0001
6 to 10: moderate effect on patient's life	–1.6364	≤ 0.0001
11 to 20: very large effect on patient's life	–2.6757	≤ 0.0001
21 to 30: extremely large effect on patient's life	–3.3830	≤ 0.0001

DLQI Dermatology Life Quality Index, N/A not applicable, *PSAAD* Pruritus and Symptoms Assessment for Atopic Dermatitis

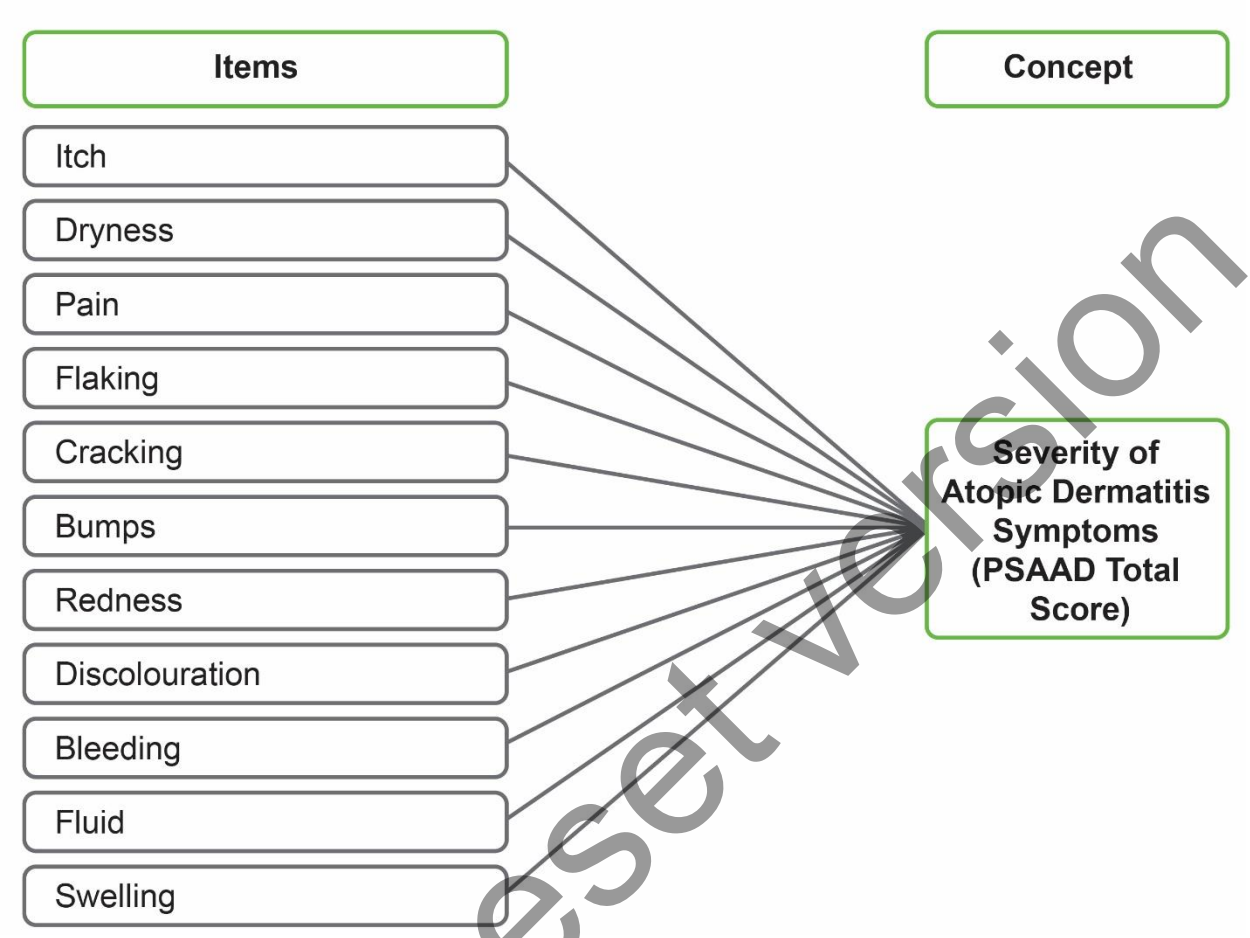
FIGURES

Fig 1 Summary of concept elicitation and conceptual saturation results for atopic dermatitis symptoms



The number of spontaneous (blue) and probed (orange) reports of each symptom are displayed along with the group of concept elicitation transcripts with which each symptom was spontaneously mentioned (checkmarks) to assess conceptual saturation. Note: Interviews were divided into three equally sized groups (Group 1, Group 2, Group 3)

403 **Fig 2** PSAAD diary conceptual framework

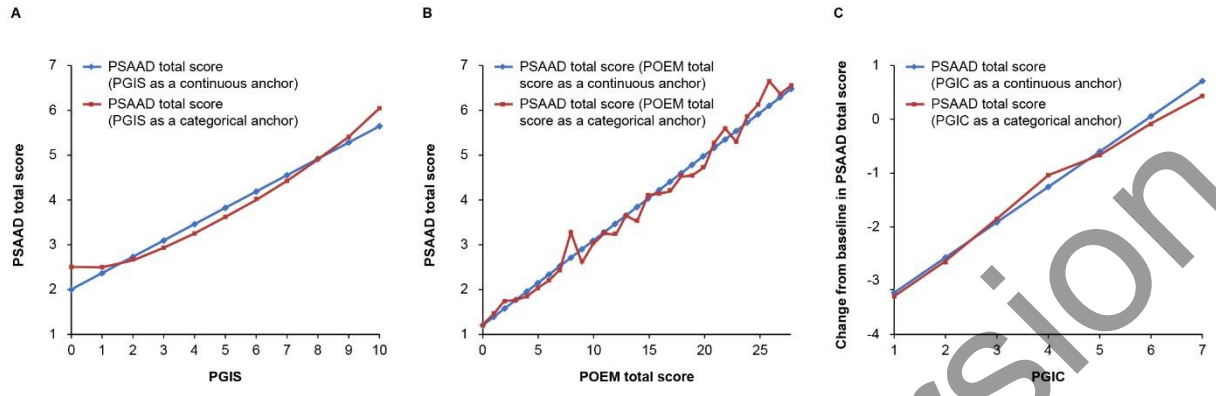


404

405 *AD* atopic dermatitis, *PSAAD* Pruritus and Symptoms Assessment for Atopic Dermatitis

406 **Fig 3** Relationship between (A) PSAAD total score and PGIS, (B) PSAAD total score and POEM total score and

407 (C) change from baseline in PSAAD total score and PGIC



408

409 *PGIC* patient global impression of change, *PGIS* patient global impression of severity, *POEM* Patient-Oriented

410 Eczema Measure, *PSAAD* Pruritus and Symptoms Assessment of Atopic Dermatitis

411 **ELECTRONIC SUPPLEMENTARY MATERIAL**

418 **Table S1** Definitions of psychometric terms

Term	Meaning
Cognitive debriefing interviewing	A qualitative research tool used to determine whether concepts and items are understood by patients in the same way that instrument developers intend. Cognitive interviews involve incorporating follow-up questions in a field test interview to gain a better understanding of how patients interpret questions asked of them. In this method, respondents are often asked to *think aloud* and describe their thought processes as they answer the instrument questions
Concept elicitation	In the development of the patient-reported measure, relevant stakeholders such as patients are queried about important aspects of the disease or condition through one-on-one interviews or focus groups Draft versions of the instructions and items are refined based on additional patient input gathered during iterative sets of interviews, commonly called cognitive interviews
Concept saturation	When sufficient data have been collected to confidently state that the key concepts of importance for the particular patient group being studied have been captured; no new or relevant information is needed
Interpretation of meaningful change and difference	Thresholds for meaningful within-patient change and, separately, between-group difference on the target PRO measure
Reliability	
Internal consistency	Consistency of responses to items on the same multi-item scale, where the items are intended to tap into different aspects of the same underlying concept
Test–retest	Stability of scores over time (at two or more time points) when no change is expected in the concept of interest, whose disease status should be stable
Validity	
Content	Evidence that the instrument measures the concept of interest, including evidence from qualitative studies that the items and domains of an instrument are appropriate and comprehensive relative to its intended measurement concept, population and use. Face validity is one component of content validity and is the degree to which the measurement instrument looks as though it is an adequate reflection of the construct (concept) being measured

Construct	Evidence that relationships among items, domains and concepts conform to a priori hypotheses concerning logical relationships expected to exist with similar or dissimilar measures Includes at least these two major elements: Strength of correlation testing a priori hypotheses (convergent validity with similar measures and divergent or discriminant validity with dissimilar measures) Degree to which the PRO instrument can distinguish between or among groups hypothesized a priori to be different (known-groups validity)

Adapted from US Food and Drug Administration (Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims https://www.fda.gov/downloads/drugs/guidances/ucm193282.pdf. Accessed 17 Jul 2020)

422 **Table S2** Demographics and baseline characteristics of interviewed patients

	PSAAD development population N = 30
Age group, n (%)	
12-14 years	8 (26.7)
15-17 years	7 (23.3)
≥18 years	15 (50.0)
Age	
Overall, mean (range), y	26.3 (12-67)
12-17 years age group, mean , y	14.8
21-67 years age group, mean, y	37.9
Female, n (%)	17 (56.7)
Race, n (%)	
Black	10 (33.3)
White	7 (23.3)
Multiracial	7 (23.3)
Other	6 (20.0)
Non-Hispanic/Latino, n (%)	20 (66.7)
Highest level of education, n (%)	
High school diploma or GED	6 (20.0)
Some years of college	3 (10.0)
College or university degree	2 (6.7)
Graduate or professional degree	4 (13.3)
Not applicable (paediatric patients)	15 (50.0)
Disease duration, mean (range), y	9.6 (2-30)

Patient-reported AD severity, n (%)	
Mild	12 (40.0)
Moderate	10 (33.3)
Severe	8 (26.7)
Physician-reported AD severity, n (%)	
Mild	9 (30.0)
Moderate	13 (43.3)
Severe	8 (26.7)
%BSA per region, mean (range)	
Head and neck	18.9 (10-40)
Upper limbs	10.8 (5-20)
Trunk	6.8 (1.7-17)
Lower limbs	7.4 (2.5-15)
EASI score per region, mean (range)	
Head and neck	2.3 (2-3)
Upper limbs	1.8 (1-2)
Trunk	1.2 (1-2)
Lower limbs	1.4 (1-2)
Top three treatments by type, n (%)	
Topical corticosteroids	28 (93.3)
Emollients/moisturizers	17 (56.7)
Antihistamines	16 (53.3)

%BSA percentage of body surface area, *AD* atopic dermatitis, *DLQI* Dermatology Life Quality Index, *EASI* Eczema Area and Severity Index, *GED* general education diploma, *IGA* Investigator's Global Assessment, *NRS* numeric rating scale, *POEM* Patient-Oriented Eczema Measure, *PSAAD* Pruritus and Symptoms Assessment for Atopic Dermatitis, *SCORAD* SCORing of Atopic Dermatitis, *SD* standard deviation

Table S3 Demographics and baseline characteristics of PSAAD validation population

	PSAAD validation population **N = 105**
Age, mean (range), y	44.4 (18.0-75.0)
Female, n (%)	68 (64.8)
Race, n (%)	
White	73 (69.5)
Black	26 (24.8)
Asian	5 (4.8)
Other	1 (1.0)
BMI, mean (SD), kg/m^2	29.4 (7.3)
Disease duration, mean/median (range), y	24.5/18.9 (1.1-68.8)
Pruritus NRS, mean (SD)	7.8 (2.1)
DLQI, mean (SD)	14.6 (7.8)
POEM, mean (SD)	20.6 (5.9)
EASI, mean (SD)	22.5 (10.5)
%BSA, mean (SD)	34.8 (21.3)
IGA, n (%)	
Moderate – 3	62 (59.0)
Severe – 4	43 (21.3)
SCORAD, mean (SD)	63.7 (12.0)

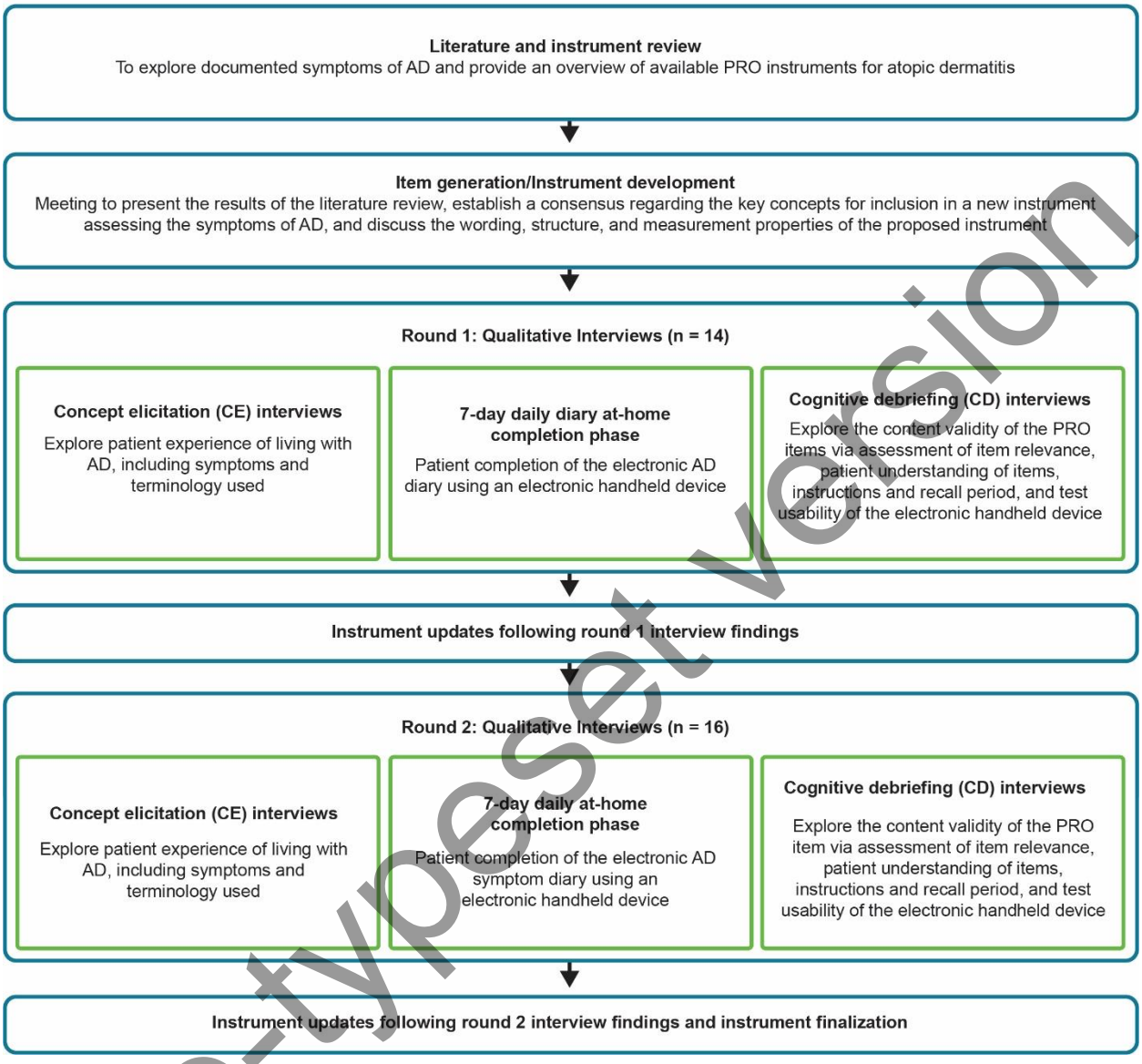
%BSA percentage of body surface area, *BMI* body mass index, *DLQI* Dermatology Life Quality Index, *EASI* Eczema Area and Severity Index, *GED* general education diploma, *IGA* Investigator's Global Assessment, *NRS* numeric rating scale, *POEM* Patient-Oriented Eczema Measure, *SCORAD* SCORing of Atopic Dermatitis, *SD* standard deviation

Table S4 Internal consistency reliability of PSAAD total score

	Cronbach coefficient alpha	Corrected item-to-total correlations		Correlations between items	
Day	**Raw**	**Minimum**	**Maximum**	**Minimum**	**Maximum**
–1	0.92	0.56	0.77	0.21	0.75
1 (baseline)	0.93	0.61	0.81	0.32	0.84
8	0.95	0.65	0.83	0.37	0.82
15	0.96	0.69	0.87	0.42	0.84
29	0.97	0.75	0.92	0.53	0.92
43	0.96	0.71	0.88	0.46	0.91
57	0.97	0.73	0.92	0.57	0.94
85	0.97	0.76	0.90	0.60	0.94
92	0.96	0.74	0.88	0.53	0.91
99	0.96	0.61	0.88	0.44	0.90
113	0.96	0.58	0.89	0.38	0.97

PSAAD Pruritus and Symptoms Assessment for Atopic Dermatitis

434 **Fig S1** Overview of methodology of PSAAD diary development

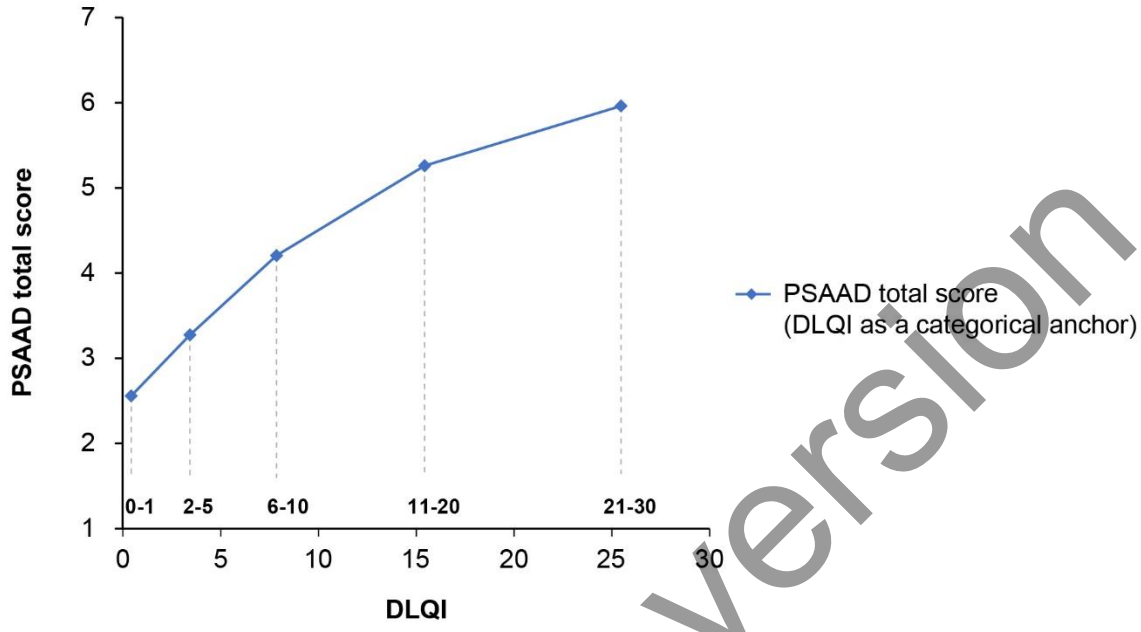


435

436 *AD* atopic dermatitis, *PRO* patient-reported outcome, *PSAAD* Pruritus and Symptoms Assessment for Atopic

437 Dermatitis

438 **Fig S2** Relationship between PSAAD diary and DLQI



439

440 *DLQI* Dermatology Life Quality Index, *PSAAD* Pruritus and Symptoms Assessment for Atopic Dermatitis

441