



BRIEF REPORT

Development, Psychometric Validation and Responder Definition of Worst Itch Scale in Children with Severe Atopic Dermatitis

Amy S. Paller · Gil Yosipovitch · Stephan Weidinger ·
Dana DiBenedetti · Diane Whalley · Abhijit Gadkari · Isabelle Guillemin ·
Haixin Zhang · Laurent Eckert · Jingdong Chao · Ashish Bansal ·
Chien-Chia Chuang · Dimitri Delevry

Received: July 14, 2022 / Accepted: August 26, 2022 / Published online: October 21, 2022
© The Author(s) 2022

ABSTRACT

Introduction: Itch associated with atopic dermatitis (AD) has a profoundly negative effect on patients of all ages. Therefore, itch is a main target for AD therapeutic approaches, and treatments are perceived as beneficial when they achieve an itch reduction. In the absence

of a validated scale for children aged 6–11 years that is suitable for assessing itch intensity in clinical trial settings, the Worst Itch Scale was developed.

Methods: Qualitative interviews, comprising concept elicitation and cognitive debriefing, were conducted to develop and evaluate the content validity of the Worst Itch Scale. Psychometric assessments used data from the LIBERTY AD PEDS phase 3 trial of dupilumab in patients aged 6–11 years with severe AD. These included test–retest reliability, construct validity, known-groups validity and responsiveness.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s13555-022-00804-z>.

A. S. Paller (✉)
Departments of Dermatology and Pediatrics,
Northwestern University Feinberg School of
Medicine, Chicago, IL, USA
e-mail: apaller@northwestern.edu

A. S. Paller
Department of Pediatrics, Northwestern University
Feinberg School of Medicine, Chicago, IL, USA

G. Yosipovitch
Department of Dermatology and Itch Center,
University of Miami, Miami, FL, USA

S. Weidinger
Department of Dermatology and Allergy, Center for
Inflammatory Skin Diseases, University Hospital
Schleswig-Holstein, Campus Kiel, Germany

D. DiBenedetti
RTI Health Solutions, Research Triangle Park, NC,
USA

D. Whalley
RTI Health Solutions, Wilmslow Road, Didsbury,
Manchester, UK

A. Gadkari
Novartis Pharmaceuticals, East Hanover, NJ, USA

I. Guillemin
IQVIA, Courbevoie, France

H. Zhang · J. Chao · A. Bansal · D. Delevry
Regeneron Pharmaceuticals, Inc, 777 Old Saw Mill
River Rd, Tarrytown, NY, USA

L. Eckert
Sanofi Aventis, 1 Avenue Pierre Brossolette, 91380
Chilly-Mazarin, France

C.-C. Chuang
Sanofi Genzyme, 50 Binney Street, Cambridge, MA,
USA

Thresholds for clinically meaningful change were defined using anchor- and distribution-based methods.

Results: The Worst Itch Scale consisted of two items asking about ‘worst itching’ experienced ‘last night’ and ‘today’. Worst Itch Scale scores showed large, positive correlations with existing patient-reported outcome (PRO) measures of itch, and weaker correlations with clinician-reported outcome (ClinRO) measures assessing objective signs of AD. Improvements in Worst Itch Scale scores were highly correlated with improvements in other itch PROs and moderately correlated with improvements in ClinROs. The responder definition based on the primary anchor, a 1-point improvement in the Patient Global Impression of Disease, was 2.84. Supportive anchors produced response estimates ranging from 2.43 to 4.80 points.

Conclusions: The Worst Itch Scale is a fit-for-purpose (e.g. well-defined, reliable, responsive and valid) scale for evaluating worst itch intensity in children aged 6–11 years with severe AD. The within-patient threshold for defining a clinically meaningful response was a ≥ 3 –4-point change in the Worst Itch Scale score.

Trial Registration: NCT03345914.

Keywords: Atopic dermatitis; Clinically meaningful; Dupilumab; Numerical rating scale; Paediatric; Pruritus; Psychometric; Responder threshold; Validation

Key Summary Points

Why carry out this study?

To develop a fit-for-purpose (well-designed, reliable and valid) scale to assess itch intensity in clinical trial settings for children aged 6–11 years.

To provide a responder definition to identify children with a meaningful change in itch.

What was learned from the study?

The Worst Itch Scale is a well-defined, reliable, sensitive and valid scale for evaluating worst itch intensity in children aged 6–11 years with severe AD.

The most appropriate within-patient threshold for defining a clinically relevant response was a ≥ 3 –4-point change on the Worst Itch Scale score.

DIGITAL FEATURES

This article is published with digital features, including a video abstract to facilitate understanding of the article. To view digital features for this article go to <https://doi.org/10.6084/m9.figshare.20655396>.

INTRODUCTION

Atopic dermatitis (AD) is a chronic, relapsing, inflammatory skin condition affecting patients of all ages [1]. Moderate-to-severe AD is characterised by intense, persistent and debilitating itch (pruritus), which can have a profoundly negative impact on patients’ lives [2–5]. Reducing itch in patients with AD is a major therapeutic goal and an important marker of treatment benefit [6]. However, itch can be accurately and reliably evaluated only by the patient. For adults and adolescents with moderate-to-severe AD, the Peak Pruritus Numerical Rating Scale (NRS) has been established as a reliable, sensitive and validated tool allowing patients to evaluate worst itch intensity [7, 8]. In adults, a change from baseline of ≥ 2 –4 points in Peak Pruritus NRS has been defined as a clinically meaningful improvement in itch [8].

Recent attempts to address the lack of available assessment tools for children have included

modifications of existing tools (such as Itchy-QoL and ItchyQuant) for use with patients as young as 6–7 years of age [9] and the generation and validation of the PROMIS Itch Questionnaire-Child (PIQ-C) for itch intensity and impact in children ≥ 8 years of age [10]. The Worst Itch Scale was developed as a fit-for-purpose (i.e. well-defined, reliable and valid) scale to assess itch intensity in clinical trial settings [11] for children aged 6–11 years. The scale measures itch intensity during daytime and nighttime (when itch can be most intense and cause sleep disturbances). The purpose of the current study was to develop a fit-for-purpose (well-designed, reliable and valid) scale to assess itch intensity in clinical trial settings for children aged 6–11 years, and to empirically derive a responder definition (i.e. a clinically meaningful within-patient improvement) for worst itch in this age group, which can be used to identify children who have experienced a meaningful improvement in itch.

METHODS

Development and Content Validation of Worst Itch Scale

Three iterative rounds of qualitative semi-structured interviews, comprising concept elicitation and cognitive debriefing, were conducted to develop and evaluate the content validity of the Worst Itch Scale. Children were eligible to participate in the study if they met the following criteria, as reported by their caregiver (parent or guardian): between the ages of 6 and 11 years, inclusive; diagnosed with AD (or eczema) by a physician; experienced symptoms of AD for a period of at least 1 year; recently (in the previous month) experienced at least moderate itching related to AD, as reported by the caregiver during screening in response to the question, ‘Over the last month, would you describe your child’s dermatitis/eczema-related itching as mild, moderate, severe or extremely severe, at its worst?’; inadequate response to topical AD medication(s) in the previous

6 months, as reported by the caregiver during screening in response to the question, ‘In the past 6 months, have any prescription or over-the-counter topical medications completely cleared up your child’s eczema?’ (possible responses were yes or no).

The first round of interviews aimed to identify any initial problems with question wording or response options. The second and third rounds were intended to provide further conceptual support, optimise items, test the adequacy of modifications and collect additional qualitative data about final items before quantitative evaluation.

Psychometric Evaluation of Worst Itch Scale

Psychometric evaluation of the Worst Itch Scale was conducted using data from the randomised, double-blinded, placebo-controlled LIBERTY AD PEDS (R668-AD-1652; ClinicalTrials.gov identifier: NCT03345914) phase 3 trial of dupilumab in paediatric patients aged 6–11 years with severe AD that was inadequately controlled with topical medications or for whom topical treatment was medically inadvisable [12]. The study was conducted in accordance with the provisions of the Declaration of Helsinki, the International Council for Harmonisation Good Clinical Practice (ICH GCP) guideline and applicable regulatory requirements; the protocol was reviewed and approved by institutional review boards/ethics committees. For all patients, written informed consent was obtained from a parent or legal guardian and written informed assent was obtained from the patient. The full study design of LIBERTY AD PEDS has been reported previously [12].

The Worst Itch Scale was included as an outcome in LIBERTY AD PEDS. The two Worst Itch Scale items (described below) were completed by children once daily in the evening via electronic diary throughout the study. Whenever possible, the child read and completed the two items alone. When required, a caregiver (parent or other) read the questions and

response options aloud to the child, but caregivers were instructed not to influence or question the responses provided by the child.

The psychometric evaluation of the Worst Itch Score was performed in accordance with classical psychometric test theory and guidelines recommended by the US Food and Drug Administration (FDA) [13] and was based on evaluations previously conducted in adolescents and adults with AD [7, 8]. Psychometric evaluation included assessment of test–retest reliability, construct validity, known-groups validity and responsiveness. Data were pooled across treatment arms.

Several patient-reported outcome (PRO) and clinician-reported outcome (ClinRO) measures from LIBERTY AD PEDS were used for the psychometric evaluation of the Worst Itch Scale. The PRO measures included Patient Global Impression of Disease (PGID), Patient Global Impression of Change (PGIC), Children’s Dermatology Life Quality Index (CDLQI), SCORing Atopic Dermatitis (SCORAD) itch Visual Analogue Scale (VAS), SCORAD sleeplessness VAS and Patient-Oriented Eczema Measure (POEM). The PGID asked participants about their itching in the last 7 days and was scored on a 5-point scale (not itchy at all, a little itchy, medium itchy, pretty itchy, very itchy). The PGIC was designed to measure perceived change in itching since starting medication and was scored on a 5-point scale (much better, a little better, the same, a little worse, much worse). The ClinRO measures included SCORAD objective score, Eczema Area and Severity Index (EASI) and Investigator’s Global Assessment (IGA).

Test–retest reliability intraclass correlation coefficients (ICC) for the weekly average of the daily worst itch were computed using scores from the last 2 weeks of the treatment period (Week 15 [test] and Week 16 [retest]), when a patient’s underlying condition and intensity of symptoms are expected to remain stable, and using baseline (test) and Week 2 (retest) scores of a subgroup of patients with no change on the PGID (Table 1). Although no standards exist for judging the magnitude of reliability coefficients for individual items, it is generally recommended that ICCs should be ≥ 0.70 for multi-item scales [14].

Establishing a Clinically Meaningful Within-Patient Change Threshold for Worst Itch Scale

Analyses were conducted using LIBERTY AD PEDS data pooled across treatment arms to define a clinically meaningful within-patient change (i.e. responder definition) using both anchor- and distribution-based methods. Consistent with recommendations that the anchor-based method is preferred [13, 15, 16] and patient-reported global status measures are the most appropriate [17], the primary anchor was a PGID improvement of 1 point. Supportive anchors were: PGIC improvement ‘a little better’ or ‘much better’; EASI response 50–74, 75–89 and 90–100 (50–74%, 75–89% and 90–100% improvement in EASI from baseline, respectively); IGA score 0 or 1; IGA improvement ≥ 2 points. The distribution-based one-half standard deviation (SD) and standard error of measurements (SEM) of the Worst Itch Scale score at baseline were calculated.

RESULTS

Development and Content Validation of Worst Itch Scale

Three iterative rounds of interviews were conducted with a total of 22 children with AD aged 6–11 years (mean [SD] age 8.7 [1.8] years; 54.5% male) and their caregivers (one parent or guardian). The study sample was racially and ethnically diverse, with approximately 55% White, 23% Black, 18% Hispanic and 5% other, as reported by the caregiver. Among the child participants, 14 (63.6%) had moderate AD-related itch and 8 (36.4%) had severe AD-related itch, per caregiver report. Efforts were made to recruit a diverse group of participants with regard to age, given the developmental and cognitive differences expected in children across the span of 6–11 years of age.

Besides Peak Pruritus NRS, which has been used to assess itch in adult and adolescent patients with AD [8, 18], several new draft items were developed in collaboration with clinical outcome assessment experts that had different

Table 1 Summary of key measurement properties of the weekly average of daily worst itch scores in paediatric patients with atopic dermatitis aged 6–11 years

Measurement property	Phase 3 Study (R668-AD-1652) (N = 361)	
Test–retest reliability	ICC ^a at prespecified timepoints [<i>n</i>]	
	Week 15 (test) and Week 16 (retest) (<i>n</i> = 342)	0.95
	Baseline (test) and Week 2 (retest) for subgroup with no change in PGID (<i>n</i> = 117)	0.76
Construct validity	Pearson correlation coefficient with weekly average of the daily worst itch at baseline, Week 16	
	SCORAD itch VAS (<i>n</i> = 361, 344)	0.69*, 0.78*
	CDLQI itch item score (<i>n</i> = 361, 345)	0.56*, 0.58*
	PGID (<i>n</i> = 361, 345)	0.65*, 0.67*
	SCORAD objective score (<i>n</i> = 361, 344)	0.24*, 0.48*
	EASI total score (<i>n</i> = 361, 345)	0.22*, 0.47*
	IGA (<i>n</i> = 361, 345)	NC ^b , 0.46*
Known-groups validity	Mean weekly daily worst itch scores per known group (ANOVA F-statistic; <i>p</i> -value)	
	PGID (not itchy at all, a little itchy, medium itchy, pretty itchy, very itchy)	
	Baseline (<i>n</i> = 1, 11, 70, 139, 140)	6.00, 5.75, 6.39, 7.63, 8.87 (67.61; <i>p</i> < 0.0001)
	Week 16 (<i>n</i> = 30, 149, 85, 52, 29)	1.39, 2.99, 5.42, 5.98, 7.73 (75.61; <i>p</i> < 0.0001)
	CDLQI bands (no effect, small, moderate, very large, extremely large effect)	
	Baseline (<i>n</i> = 2, 46, 91, 98, 124)	5.21, 6.22, 6.47, 7.18, 8.26 (26.65; <i>p</i> < 0.0001)
	Week 16 (<i>n</i> = 85, 147, 64, 23, 24)	2.36, 4.02, 5.41, 6.63, 7.18 (39.66; <i>p</i> < 0.0001)
	POEM bands (clear, mild, moderate, severe, very severe)	
	Baseline (<i>n</i> = 1, 5, 70, 186, 99)	6.29, 5.77, 6.86, 7.74, 8.71 (22.87; <i>p</i> < 0.0001)
	Week 16 (<i>n</i> = 65, 92, 96, 66, 25)	2.06, 3.28, 4.69, 6.08, 7.71 (56.08; <i>p</i> < 0.0001)

Table 1 continued

Measurement property		Phase 3 Study (R668-AD-1652) (N = 361)
Responsiveness	Change in weekly average of the daily worst itch score from baseline to Week 16: mean (SD), median (<i>n</i> = 345)	−3.49 (2.60), −3.45
	Effect size of change from baseline to Week 16 (SD at baseline units)	−2.32
	Standardised response mean (SD of change units)	−1.34
	Pearson correlation coefficient with weekly average of the daily worst itch change from baseline to Week 16	
	SCORAD itch VAS (<i>n</i> = 344)	0.70*
	CDLQI itch score (<i>n</i> = 345)	0.45*
	PGID (<i>n</i> = 345)	0.58*
	PGIC (<i>n</i> = 345)	0.50*
	SCORAD objective score (<i>n</i> = 344)	0.46*
	EASI total score (<i>n</i> = 345)	0.41*
	IGA (<i>n</i> = 345)	0.40*
	Mean change in weekly average of the daily worst itch score by change in PGID subgroups (ANOVA F-statistic; <i>p</i> -value)	
	Improved, stable, worsened (<i>n</i> = 265, 61, 19)	−4.09, −1.55, −1.35 (37.10; <i>p</i> < 0.0001)
	Effect sizes of change: improved vs stable, improved versus worsened, stable versus worsened	1.70, 1.82, 0.13

Negative change scores denote improvement on the Worst Itch Scale

ANOVA analysis of variance, CDLQI Children's Dermatology Life Quality Index, EASI Eczema Area and Severity Index, ICC intraclass correlation coefficient, IGA Investigator's Global Assessment, NC not computed, PGIC Patient Global Impression of Change, PGID Patient Global Impression of Disease, POEM Patient-Oriented Eczema Measure, SCORAD SCORing Atopic Dermatitis, SD standard deviation, VAS visual analogue scale

**p* < 0.01, Pearson correlation coefficients

^aTwo-way mixed-effects ANOVA with absolute agreement for single measures was used to compute test–retest reliability ICCs [21]

^bCorrelational analysis was not appropriate due to the baseline requirement (i.e. IGA = 4)

options for item wording (e.g. severity/intensity of worst itch, frequency of itching/scratching), response scale (e.g. 0–10 NRS, 4-point verbal rating scale, with and without figures) and recall period (24 h, last night and today). Across the three rounds of interviews, 4 children (3 aged 6 years and 1 aged 7 years) preferred that the interviewer read the items aloud while they followed along silently, although they were able

to follow along and respond to questions as the interviewer read aloud.

Poorer performing items were deleted. Respondent feedback guided the selection of items for revision and additional testing, and confirmed the final items. For example, none of the participants could accurately interpret and use the recall period description 'the previous 24 h' when selecting a response. The recall

period description ‘from the time you went to bed last night until right now’ was lengthy, wordy and confusing to almost all participants. Since many children had difficulty accurately selecting a score to represent worst itch over the entire 24-h time period, the most appropriate format for the Worst Itch Scale consisted of two separate items: one asking about ‘worst itching’ experienced ‘last night’ and one asking about ‘worst itching’ experienced ‘today’. Both items were rated by the child using an 11-point NRS on which 0 = ‘no itching’ and 10 = ‘worst itching possible’, with figures depicting the 0 and 10 anchors on the scale. Both items tested well within the context of the interviews, including participants who had limited reading ability (e.g. younger ages, poor readers). The highest (i.e. worst) response of the two items was taken as the daily worst itch score. In the LIBERTY AD PEDS study, the daily worst itch scores were averaged over 1-week intervals to take into account the potential variation in itch between the daily scores.

Psychometric Evaluation of Worst Itch Scale

Patient Characteristics

The analysis sample included 361 randomised patients (mean [SD] age 8.4 [1.7] years; 49.9% male) from the LIBERTY AD PEDS study [12] who received at least one dose of dupilumab or placebo and had at least one post-baseline Worst Itch Scale assessment during the treatment period. The study sample was racially and ethnically diverse, with approximately 69% White, 17% Black or African American, 8% Asian and 5% other. Patient demographics and background have been reported previously [12].

Test–Retest Reliability

Test–retest reliability ICCs for the weekly average of the daily worst itch during the last 2 weeks of the treatment period (Week 15 [test] and Week 16 [retest]), and at the start of the study (baseline [test] and Week 2 [retest]), were above the 0.70 criterion in both cases, at 0.95 and 0.76, respectively, indicating that the weekly average of the daily worst itch scores was

stable during the time when the patients’ disease was stable.

Construct Validity

As expected, strong, positive correlations ($r \geq 0.50$; based on Cohen’s guidelines [19]) were observed at baseline between the weekly average of the daily worst itch score and the SCORAD itch item (r , 0.69), PGID (r , 0.65) and the CDLQI itch item (r , 0.56) (Table 1). By Week 16, the magnitude of correlation coefficients increased in strength to 0.78, 0.67 and 0.58, respectively. Correlations with the ClinRO instruments were lower than with the PRO measures at both timepoints, with small correlations ($r < 0.30$; based on Cohen’s guidelines [19]) at baseline and moderate correlations ($0.30 \leq r < 0.50$; based on Cohen’s guidelines [19]) at Week 16.

Known-Groups Validity

Known-Groups Analysis of Variance (ANOVA) analyses were conducted to evaluate the discriminating ability of the weekly average worst itch at baseline and Week 16 (Table 1). Results for the known-groups comparisons were in the anticipated direction and statistically significant for all omnibus tests and most pairwise tests (except for the pairwise comparisons at baseline, likely due to lack of variability).

Sensitivity to Change

Responsiveness or sensitivity to change was evaluated by computing correlations of change from baseline to Week 16 in the weekly average of the daily worst itch scores and the supporting outcome measures, effect-size estimates of change and ANOVAs (Table 1).

Clinically Meaningful Within-Patient Change Threshold

The response threshold estimates and distribution-based estimates are presented in Table 2. The primary anchor was a 1-point improvement in the PGID. The responder definition estimated as the mean of the weekly average of the daily worst itch scores corresponding to the primary anchor was 2.84. The responder definitions

based on other anchors (PGIC, EASI response and IGA score) ranged from 2.43 for the PGIC to 4.80 for EASI 90–100. The highest estimates obtained (4.49, 4.71 and 4.80) were from the most stringent criteria (PGIC ‘much better’, IGA score 0/1 and EASI 90–100, respectively). As expected, the distribution-based one-half SD and SEM were much lower (less than 1.0) than the anchor-based threshold estimates.

DISCUSSION

The results of this study show that the Worst Itch Scale is a fit-for-purpose, well-defined and reliable tool to evaluate intensity of worst itch among paediatric patients aged 6–11 years with AD. The findings from the anchor- and distribution-based response analyses suggest that the most appropriate definition of a clinically meaningful response on the weekly average of the daily worst itch scores is ≥ 3 –4 points.

Cognitive interviews were conducted to develop and evaluate the content validity of the Worst Itch Scale. Children were included across the whole range of 6–11 years and the children in the lower end of the range (6–7 years), for whom self-assessment of itch is particularly challenging, were well represented by 6 of the 22 participants (27.3%). The most appropriate format for the Worst Itch Scale consisted of two separate items: one asking about ‘worst itching’ experienced ‘last night’ and one asking about ‘worst itching’ experienced ‘today’. The results of the psychometric evaluation of the weekly average of the daily worst itch scores using data from the LIBERTY AD PEDS clinical trial [12] provide confirmation of strong measurement properties in children aged 6–11 years with severe AD.

Test-retest reliability was well above the recommended 0.70, indicating that the weekly average of the daily worst itch scores was stable over the time when the patients’ disease was stable. The results consistently supported the construct validity of the weekly average of the daily worst itch scores, demonstrating that the weekly average provides a measure of itch intensity from the patients’ perspective. All correlations were statistically significant, and

the largest convergent correlations were, as expected, between the weekly average of the daily worst itch scores and the PRO measures designed to assess itch intensity or severity. Correlations with EASI total score and IGA score were lower, as might be expected for ClinRo instruments, which were not developed for the assessment of symptoms from the patient perspective and assess multiple clinical signs with varying association to pruritus. Furthermore, correlations tended to be lower at baseline than at Week 16, again, as expected, due to relatively low variability in scale scores resulting from study inclusion criteria designed to enroll patients with severe AD.

In the known-groups analyses, the consistent patterns of scores and highly significant differences between patients at the levels of the known groups tested at both timepoints (except for the pairwise comparisons at baseline) provide strong support for the discriminating ability of the weekly average of daily worst itch scores. Additionally, the patterns of mean change, which were in the anticipated direction and statistically significantly different across the PGID subgroups, as well as the strong correlations and large effect sizes of change, provide compelling evidence to support the responsiveness of the weekly average of the daily worst itch score. As expected, correlations between change in the weekly average of daily worst itch scores and change scores in the PRO measures, which assessed similar constructs, tended to be moderate to strong ($r \geq 0.30$), ranging from 0.45 to 0.70. These were stronger than those observed between the change in weekly average of daily worst itch scores and ClinRO measures, which ranged from 0.40 to 0.46. Overall, the results indicated that improvements in worst itch intensity reported by patients using the weekly average of the worst itch score corresponded with improvements reported by patients and clinicians, and were also related to changes in similar concepts as assessed by different instruments. Although the Worst Itch Scale score provides only a single score of itch intensity, it did compare well with CDLQI. The moderate correlation with CDLQI suggests that worst itch score reflects the impact of itch on a patient’s quality of life. However, it is important

Table 2 Estimates for thresholds for meaningful within-person change on the weekly average of the daily worst itch score using data from LIBERTY AD PEDS (R668-AD-1652; NCT03345914)

Method	Change in Weekly Average of the Daily Worst Itch Scores from Baseline to Week 16 (N = 361)	
	Mean	Median
Anchor-based		
PGID improvement of 1 point (n = 94)	−2.84	−2.83
Supportive anchor-based		
PGIC improvement		
A little better (n = 85)	−2.43	−2.08
Much better (n = 221)	−4.49	−4.83
EASI response		
EASI 50–74 (n = 71)	−3.00	−3.00
EASI 75–89 (n = 105)	−3.79	−3.83
EASI 90–100 (n = 98)	−4.80	−5.31
IGA response		
IGA 0 or 1 (n = 92)	−4.71	−5.21
IGA improvement of ≥ 2 points (n = 215)	−4.22	−4.29
Distribution-based		
One-half SD at baseline	0.75	
SEM ^a	0.34 to 0.74	

Negative change scores denote improvement on the Worst Itch Scale

EASI Eczema Area and Severity Index, *ICC* intraclass correlation coefficient, *IGA* Investigator’s Global Assessment, *PGIC* Patient Global Impression of Change, *PGID* Patient Global Impression of Disease, *SD* standard deviation, *SEM* standardised error of measurement

^aComputed as $SD\sqrt{(1-r)}$, where SD at baseline = 1.50 and r (test–retest ICC) = 0.95 (Week 15 to Week 16) and 0.76 (baseline to Week 2 for sample with no change in PGID)

that the true impact of itch on quality of life is accurately captured using CDLQI or another itch-specific impact measure.

The results of the anchor-based analyses using the primary patient-reported anchor, PGID, as well as supportive PRO and ClinRO anchors, provide evidence that a range of ≥ 3–4 points is an appropriate threshold for identifying clinically meaningful improvement in the weekly average of the daily worst itch. The distribution-based estimates of less than 1 point were lower compared with the anchor-based estimates. As the anchor-based method is

preferred by the FDA [13], this was taken as the primary method. Thresholds for meaningful improvement in itch derived in children aged 6–11 years are comparable to those observed in adults [8].

A limitation of this analysis was that thresholds were derived using empirically driven data for patients with severe AD aged 6–11 years, and so may not be appropriate for extrapolation to other AD severity classifications or conditions. The construct validity was conducted using correlations of the Worst Itch Scale with several items, including the CDLQI

itch and SCORAD itch VAS. The CDLQI and SCORAD measures have been fully validated [20]; however, the individual itch items have not been validated separately.

CONCLUSION

The Worst Itch Scale is a well-defined, reliable, sensitive and valid scale for evaluating worst itch intensity in children aged 6–11 years with severe AD. The most appropriate within-patient threshold for defining a clinically relevant response was a ≥ 3 –4-point change on the Worst Itch Scale.

ACKNOWLEDGEMENTS

Funding. The study was sponsored by Sanofi and Regeneron Pharmaceuticals, Inc. The journal Rapid Service Fee was sponsored by Sanofi and Regeneron Pharmaceuticals, Inc.

Medical Writing and Editorial Assistance. Medical writing and editorial assistance was provided by Liselotte van Delden, of Excerpta Medica, funded by Sanofi and Regeneron Pharmaceuticals, Inc., according to the Good Publication Practice guideline.

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.

Author Contributions. DDB, DW, AG, IG, LE, JC and AB contributed to manuscript concept and design. All authors interpreted the data, provided critical feedback on the manuscript, approved the final manuscript for submission and are accountable for the accuracy and integrity of the manuscript. The authors acknowledge Lauren Nelson for earlier contributions to the data analysis.

Disclosures. Amy Paller has served as an investigator for AbbVie, AnaptysBio, Eli Lilly, Incyte, Janssen, Krystal Biotech, Regeneron Pharmaceuticals, Inc. and UCB; as a consultant with honorarium for AbbVie, Acrotech, Almirall, Amgen, Amryt Pharma, Arcutis Biotherapeutics, Arena Pharmaceuticals, Azitra, BioCryst, BiomX, BMS, BridgeBio, Castle Creek Biosciences, Catawba Research, Eli Lilly, Excure, Gilead, Incyte, Janssen, Johnson & Johnson, Kamari Pharma, LEO Pharma, Novartis, OM Pharma, Pfizer, Pierre Fabre, RAPT Therapeutics, Regeneron Pharmaceuticals, Inc., Sanofi, Seanergy and UCB; and on data and safety monitoring boards with AbbVie, Abeona, Bausch Health, Galderma and Novan. Gil Yosipovitch has served as an advisory board member for Arcutis and Galderma, BELLUS Health, Eli Lilly, Kiniksa Pharmaceuticals, Novartis, Pfizer, Regeneron Pharmaceuticals, Inc., Sanofi, and Trevi Therapeutics; and received research grants from Eli Lilly, Kiniksa, LEO Pharma, Novartis, Pfizer and Sanofi-Regeneron Pharmaceuticals, Inc. Stephan Weidinger has received institutional research grants from LEO Pharma, L'Oréal, Pfizer and Sanofi; has served as consultant and/or lecturer for AbbVie, Almirall, LEO Pharma, Lilly, Pfizer, Regeneron Pharmaceuticals Inc. and Sanofi; and is involved in performing clinical trials with many pharmaceutical companies that manufacture drugs used for the treatment of psoriasis and atopic eczema. Dana DiBenedetti and Diane Whalley are employees of RTI Health Solutions, which received funding for the conduct of interviews and psychometric evaluation from Regeneron Pharmaceuticals, Inc. Abhijit Gadkari was an employee at Regeneron Pharmaceuticals, Inc. at the time of the study, may have held stock and/or stock options in Regeneron Pharmaceuticals, Inc. and is a current employee at Novartis Pharmaceuticals. Isabelle Guillemin was an employee at Sanofi at the time of the study, may have held stock and/or stock options in Sanofi and is current employee at IQVIA. Haixin Zhang, Jingdong Chao, Ashish Bansal and Dimitri Delevry are employees and shareholders of Regeneron Pharmaceuticals, Inc. Laurent Eckert and Chien-Chia Chuang are employees

at and may hold stock and/or stock options in Sanofi.

Compliance with Ethics Guidelines. The study was conducted in accordance with the provisions of the Declaration of Helsinki, the ICH GCP guideline and applicable regulatory requirements; the protocol was reviewed and approved by institutional review boards/ethics committees. For all patients, written informed consent was obtained from a parent or legal guardian, and written informed assent was obtained from the patient.

Data Availability. Qualified researchers may request access to study documents (including the clinical study report, study protocol with any amendments, blank case report form, statistical analysis plan, Worst Itch scale) that support the methods and findings reported in this manuscript. Individual anonymized participant data will be considered for sharing once the product and indication has been approved by major health authorities (e.g. FDA, EMA, PMDA, etc.), if there is legal authority to share the data and there is not a reasonable likelihood of participant re-identification. Submit requests to <https://vivli.org/>

Open Access. This article is licensed under a Creative Commons Attribution-Non-Commercial 4.0 International License, which permits any non-commercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc/4.0/>.

REFERENCES

1. Langan SM, Irvine AD, Weidinger S. Atopic dermatitis. *Lancet*. 2020;396(10247):345–60 (**Published correction appears in Lancet**. 2020;396(10253):758).
2. Brenninkmeijer EE, Legierse CM, Sillevius Smitt JH, Last BF, Grootenhuis MA, Bos JD. The course of life of patients with childhood atopic dermatitis. *Pediatr Dermatol*. 2009;26(1):14–22.
3. Sánchez-Pérez J, Daudén-Tello E, Mora AM, Lara Surinyac N. Impact of atopic dermatitis on health-related quality of life in Spanish children and adults: the PSEDA study. *Actas Dermosifiliogr*. 2013;104(1):44–52.
4. Simpson EL, Bieber T, Eckert L, et al. Patient burden of moderate to severe atopic dermatitis (AD): insights from a phase 2b clinical trial of dupilumab in adults. *J Am Acad Dermatol*. 2016;74(3):491–8.
5. Ng MS, Tan S, Chan NH, Foong AY, Koh MJ. Effect of atopic dermatitis on quality of life and its psychosocial impact in Asian adolescents. *Australas J Dermatol*. 2018;59(2):e114–7.
6. De Bruin-Weller M, Biedermann T, Bissonnette R, et al. Treat-to-target in atopic dermatitis: an international consensus on a set of core decision points for systemic therapies. *Acta Derm Venereol*. 2021;101(2):00402.
7. Yosipovitch G, Guillemin I, Eckert L, et al. Validation of the Peak Pruritus Numerical Rating Scale (NRS) in adolescent moderate-to-severe atopic dermatitis patients for use in clinical trials. *Exp Dermatol*. 2018;27(Suppl 2):S98.
8. Yosipovitch G, Reaney M, Mastey V, et al. Peak Pruritus Numerical Rating Scale: psychometric validation and responder definition for assessing itch in moderate-to-severe atopic dermatitis. *Br J Dermatol*. 2019;181(4):761–9.
9. Kong HE, Francois S, Smith S, et al. Pruritus assessment tools for 6 to 7-year-old children: KidsItchyQoL and ItchyQuant. *Pediatr Dermatol*. 2021;38(3):591–601.
10. Paller AS, Lai JS, Jackson K, et al. Generation and validation of the patient-reported outcome measurement information system itch questionnaire-child (PIQ-C) to measure the impact of itch on life quality. *J Invest Dermatol*. 2022;142(5):1309–17.e1.
11. Ständer S, Augustin M, Reich A, et al. Pruritus assessment in clinical trials: consensus recommendations from the International Forum for the Study of Itch (IFSI) Special Interest Group Scoring Itch in

- Clinical Trials. *Acta Derm Venereol.* 2013;93(5): 509–14.
12. Paller AS, Siegfried EC, Thaçi D, et al. Efficacy and safety of dupilumab with concomitant topical corticosteroids in children 6 to 11 years old with severe atopic dermatitis: a randomized, double-blinded, placebo-controlled phase 3 trial. *J Am Acad Dermatol.* 2020;83(5):1282–93.
 13. Food and Drug Administration. Guidance for industry. Patient-reported outcome measures: use in medical product development to support labeling claims. 2009. <https://www.fda.gov/regulatory-information/search-fda-guidance-documents/patient-reported-outcome-measures-use-medical-product-development-support-labeling-claims>. Accessed 11 May 2022.
 14. Nunnally JC, Bernstein IH. *Psychometric theory*. 3rd ed. New York: McGraw-Hill; 1994.
 15. McLeod LD, Coon CD, Martin SA, Fehnel SE, Hays RD. Interpreting patient-reported outcome results: US FDA guidance and emerging methods. *Expert Rev Pharmacoecon Outcomes Res.* 2011;11(2): 163–9.
 16. Wyrwich KW, Norquist JM, Lenderking WR, Acaster S. Methods for interpreting change over time in patient-reported outcome measures. *Qual Life Res.* 2013;22(3):475–83.
 17. Coon CD, Cook KF. Moving from significance to real-world meaning: methods for interpreting change in clinical outcome assessment scores. *Qual Life Res.* 2018;27(1):33–40.
 18. Phan NQ, Blome C, Fritz F, et al. Assessment of pruritus intensity: prospective study on validity and reliability of the visual analogue scale, numerical rating scale and verbal rating scale in 471 patients with chronic pruritus. *Acta Derm Venereol.* 2012;92(5):502–7.
 19. Cohen J. A power primer. *Psychol Bull.* 1992;112(1):155–9.
 20. Rehal B, Armstrong AW. Health outcome measures in atopic dermatitis: a systematic review of trends in disease severity and quality-of-life instruments 1985–2010. *PLoS ONE.* 2011;6(4): e17520.
 21. McGraw KO, Wong SP. Forming inferences about some intraclass correlation coefficients. *Psychol Methods.* 1996;1(1):30–46.