BJD
British Journal of Dermatology

Peak Pruritus Numerical Rating Scale: psychometric validation and responder definition for assessing itch in moderate-to-severe atopic dermatitis*

G. Yosipovitch,[1] M. Reaney,[2] V. Mastey,[3] L. Eckert,[4] A. Abbé,[4] L. Nelson,[5] M. Clark,[6] N. Williams,[5] Z. Chen,[3] M. Ardeleanu,[3] B. Akinlade,[3] N.M.H. Graham,[3] G. Pirozzi,[7] H. Staudinger,[7] S. Plaum,[7] A. Radin[3] and A. Gadkari[3]

[1]Dr Phillip Frost Department of Dermatology and Cutaneous Surgery and Miami Itch Center, Miller School of Medicine, University of Miami, Miami, FL, U.S.A.
[2]Sanofi, Guildford, U.K.
[3]Regeneron Pharmaceuticals, Tarrytown, NY, U.S.A.
[4]Sanofi, Chilly-Mazarin, France
[5]RTI Health Solutions, Research Triangle Park, NC, U.S.A.
[6]RTI Health Solutions, Ann Arbor, MI, U.S.A.
[7]Sanofi, Bridgewater, NJ, U.S.A.

Summary

Background Moderate-to-severe atopic dermatitis (AD) is a chronic disease characterized by intense, persistent and debilitating itch, resulting in sleep deprivation, signs of anxiety and depression, impaired quality of life and reduced productivity. The Peak Pruritus Numerical Rating Scale (NRS) was developed and validated as a single-item, patient-reported outcome (PRO) of itch severity.

Objectives To describe the content validity and psychometric assessment (test–retest reliability, construct validity, known-groups validity, sensitivity to change) of the Peak Pruritus NRS, and to derive empirically a responder definition to identify adults with a meaningful change in itch.

Methods Content validity was assessed through in-depth patient interviews. Psychometric assessments used data from phase IIb and phase III dupilumab clinical trials and included test–retest reliability, construct validity, known-groups validity and sensitivity to change in patients with moderate-to-severe AD.

Results Interview participants indicated that the Peak Pruritus NRS was a relevant, clear and comprehensive assessment of itch severity. Peak Pruritus NRS scores showed large, positive correlations with existing PRO measures of itch, and weak or moderate correlations with clinician-reported measures assessing objective signs of AD. Peak Pruritus NRS score improvements were highly correlated with improvements in other itch PROs, and moderately correlated with improvements in clinician-reported measures assessing objective signs of AD. The most appropriate threshold for defining a clinically relevant, within-person response was ≥ 2–4-point change in the Peak Pruritus NRS.

Conclusions The Peak Pruritus NRS is a well-defined, reliable, sensitive and valid scale for evaluating worst itch intensity in adults with moderate-to-severe AD.

Correspondence
Gil Yosipovitch.
E-mail: yosipog@gmail.com

Accepted for publication
4 February 2019

Funding sources
This research was sponsored by Sanofi and Regeneron Pharmaceuticals, Inc.

Conflicts of interest
G.Y. has been on advisory boards for and received honoraria from Sanofi and Regeneron Pharmaceuticals, Inc. M.R., L.E., A.A, G.P., S.P. and H.S. are employees of Sanofi and may hold stock and/or stock options in the company. V.M. is an employee and shareholder of Regeneron Pharmaceuticals, Inc. L.N., M.C. and N.W. are employees of RTI Health Solutions. Z.C., M.A., B.A., N.M.H.G., A.R. and A.G. are employees and shareholders of Regeneron Pharmaceuticals, Inc.

*Plain language summary available online

DOI 10.1111/bjd.17744

What's already known about this topic?

- Moderate-to-severe atopic dermatitis is characterized by persistent and debilitating itch, which can greatly impair quality of life.
- A validated, brief patient-reported outcome measure is needed to quantify the intensity of itch accurately and reliably in patients with atopic dermatitis in clinical trials.

What does this study add?

- The Peak Pruritus Numerical Rating Scale (NRS) is a well-defined, reliable, fit-for-purpose measure to evaluate patient-reported intensity of worst itch in the previous 24 h for adults with moderate-to-severe atopic dermatitis.
- Clinical response is indicated by a $\geq$ 2–4-point change from baseline in Peak Pruritus NRS score.

What are the clinical implications of this work?

- This study provides practising clinicians and clinical trialists with a validated patient-reported outcome measure to assess itch, a hallmark symptom of atopic dermatitis and a crucial marker of treatment benefit.

Moderate-to-severe atopic dermatitis (AD) is a chronic disease with significant patient burden that is underappreciated as a public health concern.[1–3] AD is characterized by intense, persistent and debilitating itch (pruritus), which can result in marked sleep deprivation, symptoms of anxiety and depression, impaired health-related quality of life and reduced daily productivity.[3–8] Itch can be a constant presence in patients' lives in terms of both duration and intensity. Almost two-thirds (62·9%) of patients with moderate-to-severe AD report itching 12 h a day or more, while a similar proportion (60·5%) rate their itch as severe or unbearable.[5] Itch and its downstream effects are responsible for much of the disease burden borne by patients with moderate-to-severe AD.[4] As a result, itch has a profound negative impact on the lives of these patients.[5] Hence, reducing the itch–scratch cycle is an important treatment goal for patients with moderate-to-severe AD[4,6] and forms a key aspect of reducing the overall severity of AD.

As the subjective nature of itch perception precludes a universal clinical profile,[9] the presence and intensity of itch are most accurately reported by patients directly. A special working group of the International Forum for the Study of Itch identified a need for an itch questionnaire applicable to randomized controlled trials that can support comparisons of itch parameters.[9] The Peak Pruritus Numerical Rating Scale (NRS) was developed to assess one parameter or dimension of itch (i.e. itch severity) in clinical trials of drugs in development for patients with moderate-to-severe AD. Itch intensity has been endorsed by the global Harmonising Outcome Measures for Eczema (HOME) initiative as a core symptom of AD, and is recommended for repeated measurement.[10,11] One of the goals of the HOME initiative is the standardization of outcome instruments, thus enabling trials to be compared and combined in meta-analyses. To this end, evidence of a reliable and valid assessment of itch in patients with AD would help further the goals of the HOME initiative.

This article describes the content validation and psychometric assessment of the Peak Pruritus NRS in patients with moderate-to-severe AD. Psychometric assessment included test–retest reliability, construct validity, known-groups validity and sensitivity to change. Finally, we describe the empirical derivation of a responder definition for the Peak Pruritus NRS to identify patients who have experienced a meaningful change in itch.[12,13]

Materials and methods

Peak Pruritus Numerical Rating Scale

The Peak Pruritus NRS is a single self-reported item designed to measure peak pruritus, or 'worst' itch, over the previous 24 h based on the following question: 'On a scale of 0 to 10, with 0 being "no itch" and 10 being "worst itch imaginable", how would you rate your itch at the worst moment during the previous 24 hours?'

The item was developed with input from clinical dermatology experts and confirmed by patients with AD on the basis of ease of use and understanding.[14] Although others have developed similar items,[15–17] this, to our knowledge, is the first comprehensive report summarizing the reliability, validity, responsiveness and threshold of meaningful within-person change (i.e. response definition) in a single-item global assessment of peak pruritus. The validation of the Peak Pruritus NRS followed three steps: (i) evaluation of content validity via qualitative interviews, (ii) preliminary assessment of psychometric measurement properties and (iii) confirmation of psychometric measurement properties.

Evaluating the content validity of the Peak Pruritus Numerical Rating Scale

Content validation was conducted through qualitative analysis of in-depth one-to-one patient interviews to test the relevance, wording, applicability, usability, recall period and response options of the Peak Pruritus NRS in patients with moderate-to-severe AD. Interview participants had to be at least 18 years of age, have physician-reported diagnosis of AD, have had AD for at least 3 years, have had moderate-to-severe itching related to AD in the past month, and not have participated in a related clinical trial or focus group in the past 6 months.

Moderate-to-severe itching related to AD in the past was measured by response to the survey questions (i) 'At its worst, would you describe the dermatitis-related itching as mild, moderate, severe or extremely severe?' (participants must answer 'moderate', 'severe' or 'extremely severe' to qualify) and (ii) 'When was your last itching episode, as related to atopic dermatitis: in the past 2 weeks, past month, past 2 months or more than 2 months ago?' (participants must answer 'in the past 2 weeks' or 'in the past month' to qualify).

We aimed to recruit a sample of patients with diversity of sex, race, ethnicity and level of education. Interviews included concept elicitation about AD symptoms and cognitive debriefing regarding the Peak Pruritus NRS. All interviews were conducted in English by the same pair of skilled, experienced interviewers following a semistructured interview guide, and verbatim responses were transcribed. Patients provided signed written informed consent, and all study materials were reviewed and approved by the institutional review board of RTI International (Raleigh, NC, U.S.A.).

Evaluating the psychometric measurement properties of the Peak Pruritus Numerical Rating Scale

Psychometric assessments were conducted in line with the U.S. Food and Drug Administration (FDA) guidance on patient-reported outcomes (PROs).[18] Data were drawn from a randomized, placebo-controlled, double-blinded multicentre phase IIb clinical trial (ClinicalTrial.gov Identifier: NCT01859988)[19] and two identically designed, randomized, placebo-controlled, double-blinded multicentre phase III clinical trials of dupilumab for moderate-to-severe AD (SOLO 1: NCT02277743; SOLO 2: NCT02277769).[20] Preliminary psychometric assessments were conducted in the phase IIb clinical trial, with confirmatory analyses conducted using pooled data from the two phase III clinical trials.

In all trials, Peak Pruritus NRS was administered via an Interactive Voice Response System and completed daily from baseline through week 16. The analysis population for the psychometric work consisted of all randomized patients who received at least one dose of study medication and had at least one postbaseline Peak Pruritus NRS measurement in the treatment period. Given the day-to-day fluctuation in worst itch scores, daily scores were averaged over 1-week intervals from baseline through week 16 to obtain weekly scores. Baseline scores were calculated using the average of the daily scores from the 7 days immediately preceding randomization. A minimum of four scores in the preceding 7 days was required to calculate the baseline score.

Psychometric evaluation included assessment of test–retest reliability, construct validity, known-groups validity and sensitivity to change of the Peak Pruritus NRS. Both PRO and clinician-reported outcome (ClinRO) measures were used for the psychometric assessment of Peak Pruritus NRS (Tables 1 and 2). PRO measures included the Pruritus Categorical Scale (PCS),[21] Average Pruritus NRS,[22] itch visual analogue scale (VAS) of Scoring Atopic Dermatitis (SCORAD),[23] the itch item of the Dermatology Life Quality Index (DLQI)[23] and a global measure of disease status that we have termed the Patient Global Assessment of Disease Status (PGADS). ClinRO measures included the Eczema Area and Severity Index (EASI) and the Investigator's Global Assessment (IGA).[24–26]

Test–retest reliability of the Peak Pruritus NRS was evaluated using intraclass correlation coefficients (ICCs) of week 15

Table 1 Patient- and clinician-reported outcome measures used in Peak Pruritus (worst itch) Numerical Rating Scale (NRS) validation assessments: construct and known-groups validity

Outcome measure	Type of outcome measure	Response scale	Recall period	Analysis
Construct validity				
Average Pruritus NRS	PRO	Average itch on an 11-point scale: 0–10	24 h	Correlational analysis between the Peak Pruritus NRS scores and scores on each outcome measure at baseline confirming a priori hypotheses
SCORAD itch VAS	PRO	Average itch with a range of 0–10	Last 3 days	
DLQI itch item	PRO	How itchy, sore, painful or stinging skin has been on a four-point scale: 0 (not at all) to 3 (very much)	Past week	
PCS	PRO	Overall itch on a four-point scale: 0 (absence of pruritus) to 3 (severe pruritus)	24 h	
EASI	ClinRO	Range: 0–72 points	Current	
IGA	ClinRO	Five-point scale: 0 (clear) to 4 (severe)	Current	
Known-groups validity				
PCS bands	PRO	Four-point scale: 0 (absence of pruritus) to 3 (severe pruritus)	24 h	Known-groups ANOVA at week 16 comparing mean Peak Pruritus NRS scores confirming a priori hypotheses
DLQI bands	PRO	Range: 0–30 points	Past week	
PGADS	PRO	Five-point scale: 1 (poor) to 5 (excellent)	Current	

ANOVA, analysis of variance; ClinRO, clinician-reported outcome; DLQI, Dermatology Life Quality Index; EASI, Eczema Area and Severity Index; IGA, Investigator's Global Assessment; NRS, Numerical Rating Scale; PCS, Pruritus Categorical Scale; PGADS, Patient Global Assessment of Disease Status; PRO, patient-reported outcome; SCORAD, Scoring Atopic Dermatitis; VAS, visual analogue scale.

Table 2 Patient- and clinician-reported outcome measures used in Peak Pruritus (worst itch) Numerical Rating Scale (NRS) validation assessments: sensitivity to change

Outcome measure	Type of outcome measure	Response scale	Recall period	Analysis
SCORAD itch VAS	PRO	Range: 0–10	Last 3 days	Correlation analysis between the change in Peak Pruritus NRS (baseline to week 16) and the change in each outcome measure confirming a priori hypotheses and effect-size estimates of change
DLQI itch item	PRO	Four-point scale: 0 (not at all) to 3 (very much)	Past week	
PCS	PRO	Four-point scale: 0 (absence of pruritus) to 3 (severe pruritus)	24 h	
EASI	ClinRO	Range: 0–72 points	Current	
IGA	ClinRO	Five-point scale: 0 (clear) to 4 (severe)	Current	

ClinRO, clinician-reported outcome; DLQI, Dermatology Life Quality Index; EASI, Eczema Area and Severity Index; IGA, Investigator's Global Assessment; PCS, Pruritus Categorical Scale; PRO, patient-reported outcome; SCORAD, Scoring Atopic Dermatitis; VAS, visual analogue scale.

(test) and week 16 (retest) scores and a two-way mixed-effects ANOVA with absolute agreement for single measures.[27–29] These two time points at the end of the treatment period were selected because the underlying condition and intensity of symptoms were relatively stable, on average, between these 2 weeks, as demonstrated by weekly trends of disease severity data from the relevant trial.[19,20]

Construct validity of the Peak Pruritus NRS was evaluated using correlation analyses of Peak Pruritus NRS with PCS, DLQI itch item (item 1 question: 'Over the last week, how itchy, sore, painful or stinging has your skin been?'), SCORAD itch VAS, EASI and IGA. The aim of these correlation analyses was to show a stronger relationship of Peak Pruritus NRS with measures addressing similar concepts (PCS, DLQI itch item and SCORAD itch VAS) than with measures addressing more disparate concepts (EASI and IGA). Absolute values of correlations $\geq$ 0·50 are considered strong, correlations of 0·30–0·49 are moderate and correlations of 0·10–0·29 are weak.[30]

To assess the discriminating ability of the Peak Pruritus NRS at week 16, known-groups validity was evaluated using ANOVA comparing PCS levels (absent, mild, moderate or severe pruritus), DLQI bands (no impact, minimal impact, moderate impact, very large impact, extremely large impact) and PGADS (excellent, very good, good, fair, poor).

Sensitivity of the Peak Pruritus NRS to change, or responsiveness, was evaluated using correlation analyses of Peak Pruritus NRS score change from baseline to week 16 with those for PCS, DLQI itch item, SCORAD itch VAS, EASI and IGA. We anticipated stronger associations between the change in Peak Pruritus NRS score and changes in measures addressing similar concepts, compared with the associations between the change in Peak Pruritus NRS score and those in measures addressing more disparate concepts. Also, two effect-size estimates were computed to show the magnitude of change. The effect-size estimates of change were expressed in units of standard deviation (SD) at baseline, obtained by dividing the mean change from baseline at week 16 in Peak Pruritus NRS score by the SD of the baseline score. Standardized response means were calculated as the mean change from baseline at week 16 in Peak Pruritus NRS score divided by the SD of the mean change score. Effect-size estimates above the threshold of 0·8 are considered to be large effect sizes.[30]

Empirically deriving a threshold of meaningful change for the Peak Pruritus Numerical Rating Scale

Analyses were conducted using the trial data to define a clinically meaningful within-person change, or response,[11–13,18,31] by using both distribution- and anchor-based methods.[30] The distribution-based method used one-half SD of the average Peak Pruritus NRS at baseline. The anchors included both PRO and ClinRO assessments. The PRO-based anchor was $\geq$ 1-point improvement in PCS at week 16. The ClinRO-based anchors were EASI score at week 16 according to EASI 50–74 (50–74% improvement), EASI 75–90 or EASI 90–100; and IGA score of 0 or 1 at week 16, or improvement of $\geq$ 2 points from baseline to week 16.

Results

Content validation

In June 2014, qualitative researchers conducted in-depth interviews with 14 patients with AD in Detroit, Michigan, U.S.A. (n = 6) and Tampa, Florida, U.S.A. (n = 8). The mean age of participants was 40·1 years (range 19–71) and 64% were female (Table 3). Half were Black or African American (n = 4, 29%) or Hispanic/Latino (n = 3, 21%), and approximately one-third (n = 5, 36%) had no more than a high school education.

All interview participants endorsed itch as a symptom of their AD. Most participants (n = 9, 64%) stated that once an episode of AD had begun, their pruritus was relatively constant (i.e. continuously present throughout the day), whereas the remaining participants (n = 5, 36%) described their pruritus as intermittent (i.e. comes and goes throughout the day). When asked about the impact of itch, most participants (n = 13, 93%) reported at least one meaningful consequence such as feeling embarrassed or self-conscious about itching or scratching in front of others (n = 4, 29%), problems with concentration (n = 3, 21%), sleep interruption (n = 3, 21%)

Table 3 Patient characteristics

Characteristic	Content validation	Psychometric evaluation	
	Concept elicitation and cognitive interviews (N = 14)	Exploratory analysis using phase IIb data (n = 379)	Confirmatory analysis using pooled phase III data (n = 1379)
Sex, female, n (%)	9 (64.3)	145 (38·3)	581 (42·1)
Age (years), mean ± SD	40·1 ± 15·2	37·0 ± 12·2	38·3 ± 14·3
Race, n (%)[a]			
White	7 (50)	257 (67·8)	939 (68·9)
Black or African American	4 (28.3)	33 (8·7)	94 (6·9)
Asian	0	82 (21·6)	300 (22·0)
Other	3 (21.4)	7 (1·8)	29 (2·1)
Ethnicity, n (%)[b]			
Hispanic or Latino	3 (21.4)	14 (3·7)	52 (3·9)
Region, n (%)			
Americas	14 (100)	166 (43·8)	632 (45·8)
Asia Pacific	0	58 (15·3)	205 (14·9)
Eastern Europe	0	73 (19·3)	183 (13·3)
Western Europe	0	82 (21·6)	359 (26·0)
Body mass index (kg m^{-2}), mean ± SD[c]	–	26·2 ± 6·1	26·5 ± 5·7
Duration of AD, mean ± SD[d]	–	28·0 ± 13·6	28·1 ± 15·0

AD, atopic dermatitis. [a]Pooled SOLO 1 and 2 data based on 1362 patients. [b]Pooled SOLO 1 and 2 data based on 1348 patients. [c]Phase IIb data based on 377 patients and pooled SOLO 1 and 2 data on 1377 patients. [d]Pooled SOLO 1 and 2 data based on 1367 patients.

and scratching to the point of bleeding (n = 3, 21%). All participants reported experiencing their AD overall according to fluctuations between worsening and improvement of itch.

Participants were presented with two NRSs, one measuring 'worst itch' (the Peak Pruritus NRS) and one measuring 'average itch' in the previous 24 h. Participants indicated that both the NRSs were relevant, clear and easy to answer, and comprehensive in their assessment of itch severity. Comparing the concept of 'worst' itch, as measured by the Peak Pruritus NRS, with that of 'average itch intensity' in the past 24 h, most participants reported that 'worst' itch was easier to remember and to rate (n = 11, 79%) and was more important to improve with AD treatment (n = 8, 57%). The participants' interpretation of the Peak Pruritus NRS was also more precise and consistent than 'average itch' across participants. All participants (100%) reported that it would be easy to recall their pruritus every day for 4 months in order to provide ratings of severity. The majority of participants reported that they preferred to respond via phone (n = 8, 57%) than by web system or handheld device, or using a pen and paper. Therefore, the Peak Pruritus NRS was determined to be a content-valid instrument for measuring worst itch in the past 24 h.

Psychometric assessment

Patient characteristics

The study design and patient characteristics from the dupilumab phase IIb and phase III monotherapy trials (SOLO 1 and SOLO 2) have been reported previously.[19,20] Patient characteristics were balanced across the phase IIb and pooled phase III trials (Table 3). Missing values for Peak Pruritus NRS were low at baseline and week 16 in the phase IIb trial (2·6% and 13·7%, respectively) and in the pooled phase III analysis (0·4% and 7·0%, respectively).

Test–retest reliability

With respect to test–retest reliability, ICCs for week 15 and week 16 test–retest measures of Peak Pruritus NRS were strong in both the exploratory phase IIb (0·95) and confirmatory phase III (0·96) analyses. The values were above the generally recommended threshold of 0·70 for multi-item scales (Table 4).[29,32]

Construct validity

Construct validity was conducted using correlations of Peak Pruritus NRS scores at baseline with those of similar constructs (the PRO measures PCS, DLQI itch item and SCORAD itch VAS) and dissimilar constructs (the ClinRO measures EASI and IGA). Large positive correlations were found for the NRS and measures of similar constructs, with Pearson correlation coefficients ranging from 0·61 to 0·77 (Table 4). Correlations with measures of dissimilar constructs were expectedly weak, with Pearson correlation coefficients ranging from 0·09 to 0·24 (Table 4).

Known-groups validity

All known-groups comparisons were in the anticipated direction and statistically significant. Patients reporting 'absent' or 'mild' itch vs. 'severe' itch had significantly lower Peak

Table 4 Peak Pruritus Numerical Rating Scale (NRS): validity and reliability in dupilumab clinical trials

Measurement property, outcome	Preliminary analysis using phase IIb data (n = 379)	Confirmatory analysis using pooled phase III data (n = 1379)
Test–retest reliability: Peak Pruritus NRS scores at week 15 (test) and week 16 (retest)		
	Intraclass correlation coefficient	
	Construct validity: Pearson correlation coefficient with baseline Peak Pruritus scores, Pearson r (n)	
PRO: Average Pruritus NRS[a]	1·00 (369)	–
PRO: PCS[a,b]	0·75 (369)	0·66 (1374)
PRO: DLQI Itch item[a,b]	0·67 (369)	0·61 (1373)
PRO: SCORAD Itch VAS[a]	0·77 (369)	0·72 (1363)
ClinRO: EASI[c]	0·09 (369)	0·21 (1373)
ClinRO: IGA[b,c]	0·17 (369)	0·24 (1373)
Known-groups validity at week 16, mean ± SD (n)		
PCS[b]		
Absent (0)	0·18 ± 0·3 (22)[d]	0·38 ± 0·7 (101)[e]
Mild (1)	2·84 ± 1·5 (172)[d]	2·99 ± 1·6 (625)[e]
Moderate (2)	5·52 ± 1·6 (96)[d]	5·63 ± 1·7 (417)[e]
Severe (3)	7·20 ± 1·8* (37)[d]	7·71 ± 1·7* (139)[e]
Total DLQI[b]		
No impact (0–1)	1·84 ± 1·4 (79)[f]	2·06 ± 1·8 (332)[g]
Small impact (2–5)	3·44 ± 1·8 (100)[f]	3·81 ± 2·1 (409)[g]
Moderate impact (6–10)	4·62 ± 2·2 (69)[f]	4·97 ± 2·0 (258)[g]
Very large impact (11–20)	5·78 ± 1·9 (64)[f]	6·11 ± 2·0 (218)[g]
Extremely large impact (21–30)	7·63 ± 2·0* (14)[f]	7·51 ± 1·9* (58)[g]
PGADS		
Poor (1)	5·97 ± 2·3* (52)[h]	6·60 ± 2·2* (140)[i]
Fair (2)	4·97 ± 2·3 (67)[h]	5·50 ± 2·1 (306)[i]
Good (3)	3·97 ± 2·0 (91)[h]	4·25 ± 2·0 (364)[i]
Very good (4)	2·59 ± 1·7 (72)[h]	2·88 ± 1·9 (300)[i]
Excellent (5)	2·10 ± 1·9 (44)[h]	1·61 ± 1·6 (161)[i]

ClinRO, clinician-reported outcome; DLQI, Dermatology Life Quality Index; EASI, Eczema Area and Severity Index; IGA, Investigator's Global Assessment; PCS, Pruritus Categorical Scale; PGADS, Patient Global Assessment of Disease Status; PRO, patient-reported outcome; SCORAD, Scoring Atopic Dermatitis; VAS, visual analogue scale. [a]Expected $r \geq 0.50$. [b]Polyserial correlation was used for the exploratory analysis. [c]Expected $r < 0.30$. [d]Omnibus $F(3,323) = 165.44$, $P < 0.0001$; all pairwise comparisons $P < 0.0001$ (Tukey–Kramer adjustment for multiplicity); [e]Omnibus $F(3,1278) = 676.98$, $P < 0.0001$. All pairwise comparisons $P < 0.0001$ (Tukey–Kramer adjustment for multiplicity); [f]Omnibus $F(4,321) = 59.91$, $P < 0.0001$. All pairwise comparisons $P < 0.01$ (Tukey–Kramer adjustment for multiplicity); [g]Omnibus $F(4,1270) = 205.16$, $P < 0.0001$. All pairwise comparisons $P < 0.0001$ (Tukey–Kramer adjustment for multiplicity); [h]Omnibus $F(4,321) = 34.04$, $P < 0.0001$. All pairwise comparisons $P < 0.73$ (Tukey–Kramer adjustment for multiplicity); [i]Omnibus $F(4,1266) = 182.39$, $P < 0.0001$. All pairwise comparisons $P < 0.0001$ (Tukey–Kramer adjustment for multiplicity). *$P < 0.0001$ for very severe vs. mild or absent; extremely large impact vs. no impact; or poor vs. excellent.

Pruritus NRS scores at week 16 ($P < 0·0001$). Similarly, patients who reported 'no impact' vs. 'extremely large impact' on the DLQI or 'excellent' vs. 'poor' on the PGADS had significantly lower Peak Pruritus NRS scores ($P < 0·0001$ for all comparisons) (Table 4).

Sensitivity to change

To assess responsiveness or sensitivity to change, changes from baseline in Peak Pruritus NRS scores were correlated with changes from baseline in other PRO (PCS, DLQI itch item and SCORAD itch VAS) and ClinRO (EASI, IGA) measures. Pearson correlation coefficients were positive and strong for correlations with PROs (correlation coefficients ranged from 0·64 to 0·77), as anticipated, and were positive and predominantly moderate for correlations with ClinRO measures (correlation coefficients ranged from 0·46 to 0·50) (Table 5). Improvements in itch reported by patients using Peak Pruritus NRS were consistent with those reported using other PROs for itch and with improvements in signs of the disease reported by EASI and IGA. Effect-size estimates expressed in SD baseline units were −1·4 in the exploratory phase IIb analysis and −1·8 in the confirmatory pooled phase III analysis. The standardized response means were −1·1 and −1·3, respectively (Table 5).

Threshold of meaningful change

The response threshold estimates based on the PCS, EASI and IGA, as well as the distribution-based estimate are shown in Table 6. Response estimates based on clinician-reported anchors (EASI and IGA) ranged between 2·2 and 4·2, with the

Table 5 Peak Pruritus Numerical Rating Scale (NRS): sensitivity to change in dupilumab clinical trials

Measurement property, outcome	Preliminary analysis using phase IIb data (n = 379)	Confirmatory analysis using pooled phase III data (n = 1379)
Sensitivity to change: Pearson correlation coefficients with change from baseline at week 16 in Peak Pruritus NRS scores, Pearson r (n)		
PRO: PCS[a,b]	0·71 (321)	0·72 (1280)
PRO: DLQI Itch item[a,b]	0·66 (320)	0·64 (1273)
PRO: SCORAD Itch VAS[b]	0·77 (320)	0·73 (1259)
ClinRO: EASI[c]	0·50 (321)	0·46 (1273)
ClinRO: IGA[a,c]	0·50 (321)	0·46 (1273)
Sensitivity to change: effect-size estimates of change from baseline at week 16 in Peak Pruritus NRS scores		
Effect-size estimate change in baseline SD units	−1·4	−1·8
Standardized response mean	−1·1	−1·3

ClinRO, clinician-reported outcome; DLQI, Dermatology Life Quality Index; EASI, Eczema Area and Severity Index; IGA, Investigator's Global Assessment; PCS, Pruritus Categorical Scale; PRO, patient-reported outcome; SCORAD, Scoring Atopic Dermatitis; VAS, visual analogue scale. [a]Polyserial correlation was used for the exploratory analysis. [b]Expected r ≥ 0.50. [c]Expected r ≥ 0.30.

highest estimates based on the most stringent clinical criteria (EASI 90–100 and IGA 0 or 1). The patient-reported anchor (PCS) gave a response estimate of 2·6 and a much lower distribution-based estimate of 0·76.

Discussion

Several PROs have been developed for assessing itch,[22,33–35] but none of these instruments has been fully validated for use in randomized controlled trials of patients with moderate-to-severe AD. Therefore, the Peak Pruritus NRS was developed to assess patient-reported peak pruritus or 'worst itch' over the past 24 h in this population. Cognitive interviews were conducted to select the most relevant and appropriate measure between 'average' and 'worst' itch items. Although participants were able to understand and complete both the average and Peak Pruritus NRS items using the 24-h recall period, the participants' interpretation of the Peak Pruritus NRS was more precise and consistent across participants. The concept of worst (peak) itch was also easier to rate than average itch, and hence the Peak Pruritus NRS was retained over the Average Pruritus NRS.

Table 6 Peak Pruritus Numerical Rating Scale (NRS) thresholds of meaningful change estimates[a]

Method	Peak Pruritus NRS threshold of meaningful change estimate
Anchor-based estimates: mean change from baseline to week 16 in Peak Pruritus NRS for anchor group	
PCS improvement ≥ 1 point	2·6
EASI 50–74	2·2
EASI 75–89	3·2
EASI 90–100	4·2
IGA score of 0 or 1	4·1
IGA improvement ≥ 2 points	3·9
Distribution-based estimate	
One-half SD at baseline in Peak Pruritus NRS	0·76

EASI, Eczema Area and Severity Index; IGA, Investigator's Global Assessment; NRS, Numerical Rating Scale; PCS, Pruritus Categorical Scale. [a]Data from the phase IIb clinical trial of dupilumab were used.[19]

In the qualitative interview study, we have offered evidence validating itch as a relevant concept for measuring symptom severity and demonstrating the Peak Pruritus NRS to be a clear and comprehensive tool to assess itch severity in patients with moderate-to-severe AD. In the quantitative analysis of data from the three clinical trials, we have shown that the Peak Pruritus NRS is psychometrically valid in patients with moderate-to-severe AD.

The test–retest reliability was very good, with ICCs ranging from 0·95 to 0·96, thus demonstrating the stability of Peak Pruritus NRS scores when the disease was hypothesized to be stable. These values are similar, if not superior, to the ICCs obtained in validation analyses of PROs for assessing current itch intensity in patients with chronic itch on an NRS, a VAS or a verbal rating scale (ICCs 0·74–0·87),[22,34] and the ICC of 0·91 obtained for the dynamic pruritus score.[35]

The construct validity assessments have demonstrated that Peak Pruritus NRS scores accurately capture the intensity of worst itch. The Peak Pruritus NRS was also found to discriminate predictably between groups of extreme bands on the pruritus-specific PCS, and on PROs not specifically developed to assess AD (PGADS and DLQI).

Effect-size estimates indicating the overall magnitude of change in the Peak Pruritus NRS were above Cohen's 0·80 threshold.[30] As anticipated, changes in Peak Pruritus NRS scores were moderately to strongly correlated with changes in both PROs (SCORAD itch VAS, DLQI itch item and PCS) and ClinROs (EASI and SCORAD) and support the responsiveness of the Peak Pruritus NRS to change. Correlations by active treatment vs. placebo (results not reported) confirmed the conclusion drawn from analyses of pooled treatment arms.

The findings from the anchor-based response analysis suggest that the most appropriate definition of a response on the

Peak Pruritus NRS is in the range of ≥ 2–4 points. This response threshold is similar to the 2–3 points previously determined to be the minimal clinically important difference for similar single-item PROs in patients with chronic itch, which have assessed either the worst itch over the past 3 days or average itch intensity over the past 24 h on an NRS.[35,36] Similarly, ≥ 4-point improvement on the Itch NRS was recently determined as clinically meaningful in patients with plaque psoriasis.[16] Thresholds of meaningful change can be calculated using anchor-based or distribution-based methods; estimates from both methods are reported herein. This study agrees with previous findings that distribution-based methods tend to report a lower threshold than anchor-based approaches,[37] and illustrates how the choice of anchor can affect the threshold calculation. The PCS may be the most appropriate anchor for the Peak Pruritus NRS (2·6), because the changes in scores between the two measures correlated strongly, whereas changes in the EASI and IGA scores were only moderately correlated with the change in Peak Pruritus NRS score.

There were several limitations to our analysis. Although there were a variety of ethnicities in the adult sample interviewed, children and parents or carers of children with eczema were not included; this validation applies only to adults. Data were from an international trial, and questionnaires were administered only to the subset of patients who fluently spoke a language into which the assessment tool had been translated. The daily completion of the Peak Pruritus NRS by patients is recommended by the FDA to aid interpretation of symptom data without requiring longer recall by patients.[18] Taking the average score of the Peak Pruritus NRS over a 7-day period accounts for the variation in itch that may be seen between days. However, in taking the average rather than all daily data points, variability in the dataset is reduced, and this may artificially increase the correlations with other measures. Although both SCORAD and DLQI scores have been extensively validated in several studies,[23] the individual DLQI itch and SCORAD itch VAS items have not been validated separately.

The key strength of this body of work lies in the fact that it comprehensively summarizes the content validity and assessment of psychometric properties of a single-item PRO measure to assess worst itch intensity in patients with moderate-to-severe AD. Validation of the tool using several clinical trial databases allowed us to leverage a large sample size and to perform correlation analysis of the Peak Pruritus NRS by using multiple patient- and clinician-reported outcomes. The response threshold was estimated using both anchor-based and distribution-based methods. The low correlation at baseline between Peak Pruritus NRS and EASI and IGA scores suggests that the Peak Pruritus NRS score is not a redundant measure for assessing the severity of AD and stands independently of the clinical assessment of AD signs. The HOME initiative is seeking a suitable scale for itch intensity, and the Peak Pruritus NRS may be a useful contender for consideration.[11] The analyses presented here provide strong evidence that the Peak Pruritus NRS is a well-defined, reliable, fit-for-purpose measure for evaluating the intensity of worst itch in clinical trials among patients with moderate-to-severe AD.

Acknowledgments

Medical writing and editorial assistance was provided by Sven Holm, PhD, and Juliet Bell, PhD, of Excerpta Medica, funded by Sanofi and Regeneron Pharmaceuticals, Inc.

References

1 Bieber T. Atopic dermatitis. *N Engl J Med* 2008; **358**:1483–94.
2 Murray CJ, Richards MA, Newton JN *et al.* UK health performance: findings of the Global Burden of Disease Study 2010. *Lancet* 2013; **381**:997–1020.
3 Wittkowski A, Richards HL, Griffiths CE, Main CJ. The impact of psychological and clinical factors on quality of life in individuals with atopic dermatitis. *J Psychosom Res* 2004; **57**:195–200.
4 Murota H, Takahashi A, Nishioka M *et al.* Showering reduces atopic dermatitis in elementary school students. *Eur J Dermatol* 2010; **20**:410–11.
5 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**:491–8.
6 Sutton EL. Psychiatric disorders and sleep issues. *Med Clin North Am* 2014; **98**:1123–43.
7 Zachariae R, Zachariae C, Ibsen HH *et al.* Psychological symptoms and quality of life of dermatology outpatients and hospitalized dermatology patients. *Acta Derm Venereol* 2004; **84**:205–12.
8 O'Neil JL, Chan YH, Rapp SR, Yosipovitch G. Differences in itch characteristics between psoriasis and atopic dermatitis patients: results of a web-based questionnaire. *Acta Derm Venereol* 2011; **91**:537–40.
9 Ständer S, Augustin M, Reich C *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**:509–14.
10 Harmonising Outcome Measures for Eczema (HOME). HOME Meetings and Progress on COS. Available at: http://www.homeforeczema.org/meetings-and-events/next-home-meeting.aspx (last accessed 18 March 2019).
11 Chalmers JR, Simpson E, Apfelbacker CJ *et al.* Report from the fourth international consensus meeting to harmonize core outcome measures for atopic eczema/dermatitis clinical trials (HOME initiative). *Br J Dermatol* 2016; **175**:69–79.
12 Wyrwich KW, Norquist JM, Lenderking WR, Acaster S. Methods for interpreting change over time in patient-reported outcome measures. *Qual Life Res* 2013; **22**:475–83.
13 McLeod LD, Coon CD, Martin SA *et al.* Interpreting patient-reported outcome results: US FDA guidance and emerging methods. *Expert Rev Pharmacoecon Outcomes Res* 2011; **11**:163–9.
14 Yosipovitch G, Reaney M, Mastey V et al. Validation of the peak pruritus numerical rating scale: results from clinical studies of dupilumab in adult patients with moderate-to-severe atopic dermatitis. Presented at the 75th *Annual Meeting of the American Academy of Dermatology*, Orlando, FL, U.S.A., 20–24 March 2017.
15 Haydek CG, Love E, Mollanazar NK *et al.* Validation and banding of the ItchyQuant: a self-report itch severity scale. *J Invest Dermatol* 2017; **137**:57–61.
16 Kimball AB, Naegeli AN, Edson-Heredia E *et al.* Psychometric properties of the Itch Numeric Rating Scale in patients with moderate-to-severe plaque psoriasis. *Br J Dermatol* 2016; **175**:157–62.

17 Naegeli AN, Flood E, Tucker J *et al.* The Worst Itch Numeric Rating Scale for patients with moderate to severe plaque psoriasis or psoriatic arthritis. *Int J Dermatol* 2015; **54**:715–22.
18 US Food and Drug Administration. Patient-reported outcome measures: use in medical product development to support labeling claims. Available at: http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM193282.pdf (last accessed 18 March 2019).
19 Thaçi D, Simpson EL, Beck LA *et al.* Efficacy and safety of dupilumab in adults with moderate-to-severe atopic dermatitis inadequately controlled by topical treatments: a randomized, placebo-controlled, dose-ranging phase 2b trial. *Lancet* 2016; **387**:40–52.
20 Simpson EL, Bieber T, Guttman-Yassky E *et al.* Two phase 3 trials of dupilumab versus placebo in atopic dermatitis. *N Engl J Med* 2016; **375**:2335–48.
21 Kaufmann R, Bieber T, Helgesen AL *et al.* Onset of pruritus relief with pimecrolimus cream 1% in adult patients with atopic dermatitis: a randomized trial. *Allergy* 2006; **61**:375–81.
22 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**:502–7.
23 Rehal B, Armstrong A. 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**:e17520.
24 Hanifin JM, Thurston M, Omoto M *et al.* The Eczema Area and Severity Index (EASI): assessment of reliability in atopic dermatitis. EASI Evaluator Group. *Exp Dermatol* 2001; **10**:11–18.
25 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.
26 Barbier N, Paul C, Luger T *et al.* Validation of the Eczema Area and Severity Index for atopic dermatitis in a cohort of 1550 patients from the pimecrolimus cream 1% randomized controlled clinical trials programme. *Br J Dermatol* 2004; **150**:96–102.
27 McGraw K, Wong SP. Forming inferences about some intraclass correlation coefficients. *Psychol Methods* 1996; **1**:30–46.
28 Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. *Psychol Bull* 1979; **86**:420–8.
29 Nunnally JC, Bernstein IH. *Psychometric Theory*, 3rd edn. New York, NY: McGraw-Hill Inc, 1994.
30 Cohen J. A power primer. *Psychol Bull* 1992; **112**:155–9.
31 Coon CD, Cappelleri JC. Interpreting change in scores on patient-reported outcome instruments. *Ther Innov Regul Sci* 2016; **50**:22–9.
32 Terwee CB, Bot SD, de Boer MR *et al.* Quality criteria were proposed for measurement properties of health status questionnaires. *J Clin Epidemiol* 2007; **60**:34–42.
33 Elman S, Hynan LS, Gabriel V *et al.* The 5-D itch scale: a new measure of pruritus. *Br J Dermatol* 2010; **162**:587–93.
34 Reich A, Heisig M, Phan NQ *et al.* Visual analogue scale: evaluation of the instrument for the assessment of pruritus. *Acta Derm Venereol* 2012; **92**:497–501.
35 Ständer S, Blome C, Anastasiadou Z *et al.* Dynamic pruritus score: evaluation of the validity and reliability of a new instrument to assess the course of pruritus. *Acta Derm Venereol* 2017; **97**:230–4.
36 Reich A, Riepe C, Anastasiadou Z *et al.* Itch assessment with visual analog scale and numerical rating scale: determination of minimal clinically important difference in chronic itch. *Acta Derm Venereol* 2016; **96**:978–80.
37 Wright A, Hannon J, Hegedus EJ, Kavchak AE. Clinimetrics corner: a closer look at the minimal clinically important difference (MCID). *J Man Manip Ther* 2012; **20**:160–6.

Supporting Information

Additional Supporting Information may be found in the online version of this article at the publisher's website:

Video S1 Author video.