BJD
British Journal of Dermatology

Psychometric validation and responder definition of the sleep disturbance numerical rating scale in moderate-to-severe atopic dermatitis*

J. Puelles,[1] F. Fofana,[2] D. Rodriguez,[3] J.I. Silverberg,[4] A. Wollenberg,[5] C. Dias Barbosa[6] M. Vernon,[7] R. Chavda,[1] S. Gabriel[1] and C. Piketty[1]

[1]Galderma, La Tour-de-Peilz, Switzerland
[2]Evidera, Bennekom Born, the Netherlands
[3]Evidera, Seattle, WA, USA
[4]Department of Dermatology, The George Washington University School of Medicine and Health Sciences, Washington, DC, USA
[5]Department of Dermatology and Allergology, Ludwig-Maximilians-Universität, Munich, Germany
[6]Evidera, Ivry-sur-Seine, France
[7]Evidera, Bethesda, MD, USA

Summary

Background Sleep disturbance (SD) is an important part of the burden of atopic dermatitis (AD), but patient-reported outcomes that are easy to understand and interpret in the target population have been lacking. A daily, single-item, self-reported SD 11-point numerical rating scale (NRS) was recently developed to assess SD for patients with moderate-to-severe AD, but its psychometric properties have not yet been described.

Objectives To assess the psychometric properties of the SD NRS in patients with moderate-to-severe AD.

Methods The psychometric properties of the SD NRS were assessed using data from a phase IIb clinical trial in 218 adults with moderate-to-severe AD.

Results Test–retest reliability of the SD NRS was substantial to almost perfect (interclass correlation 0·66–1·00) in participants who had stable SD or stable pruritus scores over 1 week. Baseline correlations were moderate to large ($r > 0·30$) between SD NRS and pruritus or sleep loss scores, but were small ($r = -0·11$ to $0·17$) between SD NRS and EQ-5D-3L index and visual analogue scores, Hospital Anxiety and Depression Scale, Scoring Atopic Dermatitis, and Investigator's Global Assessment. The SD NRS could discriminate groups of participants in the expected direction according to different quality-of-life scores but not according to different clinician-reported disease severity scores. SD NRS scores significantly decreased as sleep loss, itch and quality-of-life scores improved. Analysis of meaningful change suggested a 2–5-point improvement as the initial range of responder definition in the SD NRS score.

Conclusions The SD NRS is a reliable, valid and responsive measure of SD in adults with moderate-to-severe AD.

Correspondence
Carla Dias Barbosa.
Email: carla.dias-barbosa@evidera.com

Accepted for publication
30 September 2021

Funding sources
This research was funded by Galderma. Galderma participated in the design, interpretation, review, decision to publish, and approval of the final manuscript.

Conflicts of interest
J.P., R.C., S.G. and C.P. are employees of Galderma. F.F., D.R., C.D.B. and M.V. are employees of Evidera, which was paid by Galderma for work related to this study.

Data availability
The research data are not shared.

*Plain language summary available online

DOI 10.1111/bjd.20783

What is already known about this topic?

- Sleep disturbance (SD) is a dynamic, multidimensional concept resulting in daytime fatigue and subsequent changes in physical and mental health that vary from day to day.
- SD is an important part of the burden of atopic dermatitis, but ways of effectively and reliably measuring it from the patient perspective have been lacking.
- A self-reported, daily, 11-point SD numerical rating scale (NRS) was recently developed for assessing SD in patients with moderate-to-severe atopic dermatitis, and its content validity was previously established.

What does this study add?

- The study showed that the SD NRS is reliable, valid and responsive and can measure day-to-day fluctuations in SD related to atopic dermatitis.
- The study also established an initial responder definition (i.e. meaningful interpatient change) for the SD NRS score.

What are the clinical implications of this work?

- The SD NRS is a brief, simple, easy-to-interpret and validated patient-reported global measure for the daily assessment of SD related to atopic dermatitis.
- The SD NRS can be used in clinical trials and clinical practice to assess changes in sleep quality in patients with atopic dermatitis.

Although sleep disturbance (SD) is an important part of the burden of atopic dermatitis (AD), few clinical trials of AD treatments have assessed it as a treatment outcome. Objective measures of sleep, especially actigraphy and polysomnography, have been included in some trials,[1–3] but they do not capture the patient's perspective,[4] including how they feel or function in daily life.[5] Patient-reported outcomes (PROs) are available for assessing sleep quality from the patient's perspective.[6,7] Some have been employed as outcome measures in clinical trials of AD treatments,[8] including the Pittsburgh Sleep Quality Index for adults[9] and the PROMIS-SD and PROMIS-SRI (sleep-related impairment) item banks.[10]

SD has been captured for many years using a visual analogue scale (VAS) included in Scoring Atopic Dermatitis (SCORAD), a clinician-reported score for assessing AD signs and symptoms, and more recently it has been captured using a VAS included in the patient-reported version, PO-SCORAD.[11,12] Although these scales have been validated in AD, some have been designed to monitor SD only at specific timepoints or use VAS scales whose design or administration limit their usefulness.[13,14] Others do not allow an SD score to be derived, and none capture day-to-day fluctuations in SD. Further, patients prefer numerical rating scales (NRSs) over VASs and find them easier to use and interpret.[15–18]

Because of these issues, we developed a self-reported 11-point SD NRS for the daily assessment of SD in adults and adolescents with moderate-to-severe AD. In accordance with US Food and Drug Administration guidance on PROs,[5] the SD NRS was first demonstrated to have content validity for assessing SD in moderate-to-severe AD through cognitive interviews.[19] The interviews, which included 20 adults and 10 adolescents with moderate-to-severe AD, moderate-to-severe pruritis, and SD, confirmed that the SD NRS and its anchors were easily understood as intended. Daily morning administration of the SD NRS was recommended to provide accurate recall of the number or duration of night-time awakenings. The interviews also showed that most participants would consider a 1- or 2-point decrease in the SD NRS a meaningful improvement in AD-associated SD.

As a next step in demonstrating that the SD NRS is easy to understand and fit for purpose for assessing SD in patients with AD, the current study examined its other psychometric properties (test–retest reliability, convergent–divergent validity, known-groups validity, and ability to detect change) using data collected in a phase IIb clinical trial in patients with moderate-to-severe AD.[20] In addition, the study used anchor-based approaches, supported by distribution-based approaches and the previously collected qualitative information, to establish preliminary responder definitions for meaningful within-patient change in the SD NRS score for this target population.

Patients and methods

Study design

This post hoc analysis examined the psychometric properties of the SD NRS using data from a multicentre, randomized, placebo-controlled, double-blinded phase IIb clinical trial assessing the efficacy and safety of nemolizumab in patients with moderate-to-severe AD and severe pruritus (NCT03100344).[20] Participants were eligible to participate in the trial if they were ≥ 18 years old, experienced chronic AD (according to the consensus criteria of the American Academy of Dermatology)[21] for ≥ 2 years before study entry, had confirmed moderate-to-severe AD based on clinical assessments, had severe itch on at least three of the last 7 days before study entry, and had a documented history of an inadequate response to topical medications. The trial's primary outcome was the percentage change in Eczema Area and Severity Index (EASI)[22] from baseline at week 24. Secondary outcomes included the SD NRS, Peak Pruritis (PP) NRS,[23] Average Pruritus (AP) NRS,[24] Pruritus Categorical Score (PCS),[25] EQ 5-Dimensions 5-Levels (EQ-5D-3L; using the UK value set),[26] Dermatology Life Quality Index (DLQI),[27] Hospital Anxiety and Depression Scale (HADS),[28] Investigator's Global Assessment (IGA) and SCORAD.[11] A list of outcome measures, recall periods and frequencies of assessments is provided in Table S1 (see Supporting Information).

The SD NRS asks the following question to the participant: 'On a scale of 0–10, with 0 being "no sleep loss related to the symptoms of atopic dermatitis" and 10 being "I did not sleep at all due to the symptoms of atopic dermatitis", how would you rate your sleep last night?'. The SD NRS was completed by participants on an electronic device once daily in the morning throughout the clinical trial. Previous work supported the importance and relevance of SD in participants with moderate-to-severe AD.[19] Furthermore, the SD NRS question and its anchors were found to be easy or very easy to understand in participants with moderate-to-severe AD, and the evidence of its content validity has been previously described.[19]

Analysis of the psychometric properties of the sleep disturbance numerical rating scale

Test–retest reliability was assessed in 'stable' participants, who were defined as having no change or minimal change in the concept measured over 1 week based on various anchors. The primary anchor for assessing test–retest reliability was the SCORAD sleep loss VAS, a component of SCORAD (see Table S1). Other anchors included the PCS (average weekly score), PP NRS (average weekly score), AP NRS (average weekly score) or IGA. For primary analyses, stable participants were defined as having no change or minimal change in the anchor from baseline (test) to week 1 (retest). In exploratory analyses, they were defined as having no change in the anchor from week 15 (test) to week 16 (retest) or from week 23 (test) to week 24 (retest). Paired *t*-tests and intraclass correlation coefficients (ICCs) were computed comparing SD NRS values using a two-way mixed-effects ANOVA model for absolute agreement following Shrout and Fleiss.[29] ICCs were categorized according to Koo and Li[30] as no agreement for < 0, slight agreement for 0·00–0·20, fair agreement for 0·21–0·40, moderate agreement for 0·41–0·60, substantial agreement for 0·61–0·80 and almost perfect agreement for 0·81–1·00.

Construct validity was assessed through convergent–divergent validity and known-groups validity. Convergent–divergent validity was assessed by calculating Spearman rank-order correlation coefficients for the SD NRS average weekly score at baseline vs. the baseline values for all clinical outcome assessments measured in the trial: 5-D Itch sleep item score, SCORAD sleep loss VAS score, PCS average weekly score, PP NRS average weekly score, AP NRS average weekly score, 5-D Itch total score, DLQI total score, EQ-5D-3L index score, EQ-5D-3L VAS score, HADS anxiety score, HADS depression score, SCORAD total score, SCORAD body surface area (BSA; a component of SCORAD; see Table S1) and EASI. A priori hypotheses were that correlations would be stronger between the SD NRS and scales measuring a similar construct (i.e. scales measuring SD: 5-D Itch sleep item, SCORAD sleep loss VAS) than scores from scales measuring dissimilar constructs (i.e. scales not measuring SD).

Known-groups validity of the SD NRS was assessed by comparing the mean SD NRS average weekly scores at baseline between severity groups of participants categorized according to 5-D Itch sleep item score, DLQI total score, IGA score and EASI score. Comparisons were made by two-sample t-test (two groups) or ANOVA (three or more groups) adjusted for multiple comparisons based on the Scheffe method. Raw and refined categories, grouping and explanations are provided in Table S2 (see Supporting Information).

As a first step in assessing the responsiveness of the SD NRS, Spearman rank-order correlation coefficients were calculated between the change from baseline to week 24 in SD NRS weekly scores and the changes from baseline to week 24 in SCORAD sleep loss VAS score, PCS average weekly score, PP NRS average weekly score, NRS average weekly score, DLQI total score, EQ-5D-3L index score, EQ-5D-3L VAS score, HADS anxiety score, HADS depression score, SCORAD total score, BSA score, IGA score and EASI. In a second step, responsiveness was further tested at week 16 (exploratory) and week 24 (primary) using paired *t*-tests and an ANCOVA adjusted to baseline score to compare the change in mean weekly SD NRS score in participants who were categorized as 'improved' or 'not improved' according to change from baseline in clinical outcome assessments having a large correlation with the SD NRS change from baseline (definitions in Table S3; see Supporting Information). Effect-size statistics were calculated for each group of participants as the mean difference between the baseline and week 16 or 24 average weekly score divided by the standard deviation of the baseline average weekly score of the SD NRS. The standardized response mean was also calculated for each group of participants as the mean difference between the baseline average weekly score and week 16 or 24 average weekly score divided by the standard deviation of the change in average weekly score.

Responder definition

As recommend by the US Food and Drug Administration,[31] multiple anchor-based methods were used to establish meaningful within-patient change and to derive responder definition estimates for the SD NRS score. The anchor-based responder definitions were estimated as the mean change from baseline in the SD NRS average weekly score to week 24 based on the following criteria: (i) change from baseline in SCORAD sleep loss VAS ($\geq$ 1-point decrease),[32,33] (ii) PCS average weekly score $\leq$ 1 (rounded value),[34] (iii) change from baseline in PP NRS average weekly score ($\geq$ 4-point decrease)[34] and (iv) change from baseline in DLQI total score ($\geq$ 4-point decrease).[34] As an additional exploratory analysis, the anchor-based responder definitions were estimated based on differences between baseline and week 16.

To help interpretation, the anchor-based methods were supported by distribution-based methods (standard error of measurement and the half- and quarter-standard deviation) and the previous qualitative findings of the smallest improvement in the SD NRS considered satisfactory (1–3-point decrease) and meaningful change (1- or 2-point decrease) in the SD NRS in patients with moderate-to-severe AD.[19] The standard

error of measurement was computed as the standard deviation of an observed score related to its reliability [standard deviation × square root (1 − ICC)], where the ICC was from the SD NRS test–retest reliability in participants defined as stable based on the SCORAD sleep loss VAS.

Statistical considerations

All analyses were conducted on all participants randomized in the phase IIb clinical trial who had SD NRS data at baseline. Because these are post hoc analyses, the sample size was not calculated prospectively. All available data, irrespective of the treatment arm, were included as this is a standard approach when validating a PRO measure using clinical trial data.[23,35] Analyses were performed using SAS version 9.4 (SAS Institute, Cary, NC, USA). According to Cohen's conventions, absolute values of correlations are considered large if ≥ 0·50, moderate if 0·30–0·49 and small if 0·10–0·29.[36]

Results

Participant characteristics

In total, 226 patients were randomized in the phase IIb study. Of these, 218 (96% of the initial study sample) had an SD NRS score at baseline and constituted the sample analysed in this work (Table 1). SD NRS data were available for 175 patients at week 16 and for 154 patients at week 24. The mean (standard deviation) age at baseline was 39·2 (15·2) years, and just over half of the participants (52%) were male. The majority of participants were white (75%) and non-Hispanic (95%). Investigators assessed the global severity as moderate (IGA = 3) for 66% of participants and severe (IGA = 4) for 34%, and most participants (87%) reported having had severe pruritus for at least three of the last 7 days. The mean (standard deviation) baseline SD NRS score was 7·8 (1·6) (Table 1).

Table 1 Participant demographic and clinical characteristics at baseline

Characteristic/measure	N = 218
Age (years)	
Mean (SD)	39·2 (15·2)
Range	18–82
Sex, n (%)	
Male	113 (51·8)
Female	105 (48·2)
Race, n (%)	
White	164 (75·2)
African American/black	26 (11·9)
Asian	24 (11·0)
American Indian/Alaska native	1 (0·5)
Other	3 (1·4)
Ethnicity, n (%)	
Hispanic/Latino	11 (5·0)
Not Hispanic/Latino	207 (95·0)
EASI score, mean (SD)	25·6 (10·9)
IGA score, n (%)[a]	
3 (moderate)	144 (66·1)
4 (severe)	74 (33·9)
Body surface area score, mean (SD)[b]	41·7 (18·6)
SCORAD total score, mean (SD)[b]	66·9 (11·6)
≥ 3 days with severe pruritus in the past 7 days, n (%)	190 (87·2)
SD NRS, mean (SD)	7·8 (1·6)

EASI, Eczema Area and Severity Index; IGA, investigator's Global Assessment; SCORAD, Scoring Atopic Dermatitis; SD NRS, sleep disturbance numerical rating scale. [a]N = 215; [b]N = 217.

Psychometric properties of the sleep disturbance numerical rating scale

Test–retest reliability

ICCs (range 0·66–0·98) indicated substantial to almost perfect agreement from baseline to week 1. Furthermore, ICCs exceeded the recommended threshold of 0·70[37] for three out of the four anchors (Table 2). In exploratory analyses, agreement was almost perfect between week 15 and week 16 and between week 23 and week 24 when using PCS, PP NRS or AP NRS as the anchor (ICC 0·97–1·00) (Table 2).

Convergent and divergent validity

As expected, correlations at baseline were large between the SD NRS average weekly score and scores assessing SD (r = 0·52 for 5-D Itch sleep item and r = 0·58 for SCORAD sleep loss VAS) (Table 3). Correlations were small between the SD NRS average weekly score and scores assessing a dissimilar construct (r = −0·11 to 0·17) except for those assessing itch (r = 0·42–0·84 for 5-D itch, PP NRS, AP NRS and PCS) and quality of life (r = 0·42 for DLQI) (Table 3).

Known-groups validity

The SD NRS was able to discriminate participants in the expected direction according to groups defined by the 5-D Itch sleep item score and DLQI total score at baseline (P < 0·001 for both) (Table 4). Known-groups validity was not supported when using clinician-reported scales such as IGA (P = 0·25) or EASI (P = 0·11).

Responsiveness

As a first step in assessing responsiveness, the change in SD NRS weekly score between baseline and week 24 was compared with the change in the other outcomes over the same period. Correlations were large between the change in SD NRS and the change in AP NRS average weekly score (r = 0·91), PP NRS weekly average score (r = 0·88), SCORAD sleep loss VAS score (r = 0·80), PCS weekly average score (r = 0·75) and SCORAD total score (r = 0·51) (Table 5). Correlations

Table 2 Test–retest reliability of the sleep disturbance numerical rating scale (SD NRS) in stable participants

		SD NRS			t-test		
Anchor	N	Week n, mean (SD)	Week n + 1, mean (SD)	Mean difference	t-value	P-value	ICC[a]
Baseline to week 1							
SCORAD sleep loss VAS score	56	7·81 (1·70)	7·37 (1·91)	−0·11	3·7	< 0·001	0·96
PCS average weekly score	158	7·74 (1·71)	6·84 (1·80)	−0·90	8·97	< 0·001	0·66
PP NRS average weekly score	16	8·19 (1·14)	8·10 (1·22)	−0·09	0·92	0·37	0·95
AP NRS average weekly score	17	7·58 (1·86)	7·58 (1·77)	0·01	−0·06	0·95	0·98
IGA score	146	7·90 (1·56)	6·40 (2·07)	−1·5	9·61	< 0·001	0·36
Week 15–16							
PCS average weekly score	172	2·75 (2·63)	2·70 (2·60)	−0·04	1·01	0·32	0·97
PP NRS average weekly score	44	2·35 (3·04)	2·30 (3·02)	−0·05	1·2	0·24	1.00
AP NRS average weekly score	52	2·38 (2·99)	2·34 (2·97)	−0·04	1·12	0·27	1.00
Week 23–24							
PCS average weekly score	142	2·45 (2·52)	2·48 (2·53)	0·03	−0·63	0·53	0·98
PP NRS average weekly score	33	2·48 (3·26)	2·47 (3·30)	−0·01	0·13	0·90	1.00
AP NRS average weekly score	18	2·61 (2·97)	2·62 (2·98)	0·01	−0·13	0·90	0·99

Stable participants were defined as having no change or minimal change over a week in the indicated anchor. AP NRS, Average Pruritus Numerical Rating Scale; ICC, interclass correlation coefficient; IGA, Investigator's Global Assessment; PCS, pruritus categorical score; PP NRS, Peak Pruritus Numerical Rating Scale; SCORAD, Scoring Atopic Dermatitis; VAS, visual analogue score. [a]ICCs were categorized as no agreement for < 0, slight agreement for 0·00–0·20, fair agreement for 0·21–0·40, moderate agreement for 0·41–0·60, substantial agreement for 0·61–0·80 and almost perfect agreement for 0·81–1·00.[30]

Table 3 Convergent and divergent validity: relationship between the sleep disturbance numerical rating scale (SD NRS) average weekly score and other outcome scores at baseline

		Spearman rank-order correlation	
Scale score	N	r[a]	P-value
5-D Itch sleep item	207	0·52	< 0·001
SCORAD sleep loss VAS	218	0·58	< 0·001
PCS (weekly average)	217	0·47	< 0·001
PP NRS (weekly average)	217	0·84	< 0·001
AP NRS (weekly average)	217	0·81	< 0·001
5-D Itch (total score)	207	0·42	< 0·001
DLQI (total score)	207	0·42	< 0·001
EQ-5D-3L index	207	−0·11	0·116
EQ-5D-3L VAS	207	−0·07	0·310
HADS anxiety	206	0·16	0·024
HADS depression	206	0·12	0·099
SCORAD (total score)	217	0·17	0·010
Body surface area	217	0·08	0·254
Investigator's Global Assessment	218	−0·05	0·435
Eczema Area and Severity Index	218	0·14	0·034

AP NRS, Average Pruritus Numerical Rating Scale; DLQI, Dermatology Life Quality Index; EQ-5D-3L, EuroQol 5-Dimensions 3-Levels; HADS, Hospital Anxiety and Depression Scale; PCS, pruritus categorical score; PP NRS, Peak Pruritus Numerical Rating Scale; SCORAD, Scoring Atopic Dermatitis; VAS, visual analogue score. [a]The correlation was considered small for $|r| < 0{\cdot}30$, moderate for $0{\cdot}30 \leq |r| < 0{\cdot}50$, and large for $|r| \geq 0{\cdot}50$.[45]

were moderate between the change in SD NRS and change in DLQI total score (r = 0·41). Correlations for all other outcome measures were small (range, r = −0·12 to 0·28).

Table 4 Known-groups validity: comparison of mean SD NRS average weekly scores at baseline between participants categorized by 5-D Itch sleep item score, DLQI total score, IGA score and EASI

		SD NRS average weekly score,	Overall F-test	
Scale	N	mean (SD)	Test value	P-value
5-D Itch sleep item			22·9	< 0·001
1 + 2 (never affects sleep + sleep onset delay)	33	6·7 (2·0)		
3 (both sleep onset delay and night awakenings)	174	8·1 (1·4)		
DLQI (total score)			14·5	< 0·001
0–1 (no effect)	0	–		
2–10 (small-to-moderate effect)	47	7·2 (1·8)		
11–20 (very large effect)	100	7·7 (1·6)		
21–30 (extremely large effect)	60	8·7 (1·1)		
IGA			1·35	0·247
3 (moderate)	144	7·9 (1·5)		
4 (severe)	74	7·7 (1·7)		
EASI			2·53	0·113
12–21·0 (moderate AD)	93	7·6 (1·6)		
21·1–72·0 (severe/very severe AD)	125	8·0 (1·6)		

AD, atopic dermatitis; DLQI, Dermatology Life Quality Index; EASI, Eczema Area and Severity Index; IGA, Investigator's Global Assessment; SD NRS, sleep disturbance numerical rating scale.

Table 5 Correlations between the change in sleep disturbance numerical rating scale (SD NRS) weekly scores from baseline and changes in other outcome scores from baseline

	Spearman rank-order correlation					
	Week 24			Week 16		
Outcome scale	N	r[a]	P-value	N	r[a]	P-value
SCORAD sleep loss VAS	154	0·80	< 0·001	167	0·77	< 0·001
PCS (average weekly)	146	0·75	< 0·001	172	0·79	< 0·001
PP NRS (average weekly)	146	0·88	< 0·001	171	0·88	< 0·001
AP NRS (average weekly)	146	0·91	< 0·001	171	0·90	< 0·001
DLQI (total score)	137	0·41	< 0·001	–	–	–
EQ-5D-3L index	137	−0·12	0·157	–	–	–
EQ-5D-3L VAS	137	−0·12	0·152	–	–	–
HADS anxiety	136	0·15	0·072	–	–	–
HADS depression	136	0·15	0·082	–	–	–
SCORAD (total score)	154	0·51	< 0·001	–	–	–
Body surface area	154	0·17	0·041	171	0·19	0·012
Investigator's Global Assessment	154	0·28	< 0·001	171	0·34	< 0·001
Eczema Area and Severity Index	154	0·21	0·009	171	0·29	< 0·001

AP NRS, Average Pruritus Numerical Rating Scale; DLQI, Dermatology Life Quality Index; HADS, Hospital Anxiety and Depression Scale; EQ-5D-3L, EuroQol 5-Dimensions 3-Levels; PCS, pruritus categorical score; PP NRS, Peak Pruritus Numerical Rating Scale; SCORAD, Scoring Atopic Dermatitis; VAS, visual analogue score. [a]The correlation was considered small for $|r| < 0{\cdot}30$, moderate for $0{\cdot}30 \leq |r| < 0{\cdot}50$ and large for $|r| \geq 0{\cdot}50$.[45]

Responsiveness of the SD NRS was further assessed based on clinical outcomes for which correlations were large or moderate ($|r| > 0{\cdot}30$ in Table 5). In all cases, the SD NRS average weekly mean score decreased significantly more between baseline and week 24 in participants classified in the 'improved' group than in participants classified in the 'not improved' group (Table 6). Also, in all cases, effect sizes and standardized response means were larger for patients classified as 'improved' than in those classified as 'not improved' (effect sizes 3·31–4·51 in the 'improved' group and 0·16–2·45 in the 'not improved' group; standardized response mean 2·02–3·11 in the 'improved' group and 0·36–1·46 in the 'not improved' group). Results were similar using week 16 data (Table 6).

Responder definition estimate

At week 24, a responder definition of the SD NRS score was estimated based on the SCORAD sleep loss VAS, PCS, PP NRS and DLQI. Anchor-based responder definition estimates ranged from 5·6 to 6·7 (Table 7). Anchor-based responder definition estimates were similar at week 16, ranging from 5·4 to 6·7. Distribution-based estimates were calculated using week 24 SD NRS data. These estimates included a standard error of measurement of 1·58, a quarter-standard deviation of 0·40 and half-standard deviation of 0·81.

Final responder definitions were based on the anchor-based estimates, with the support of the distribution-based estimates along with previous qualitative findings exploring meaningful change.[19] Taken together, the results suggested that a 2–5-point reduction in the SD NRS represents a meaningful improvement for the target population.

Discussion

A variety of PROs are available to assess SD, especially the PO-SCORAD sleep loss VAS,[11,12] but some of these monitor SD only at specific timepoints or use VAS scales whose design or administration limit their usefulness.[13,14] Others do not allow an SD score to be derived, and none capture day-to-day fluctuations in SD. Further, patients prefer NRSs over VASs and find them easier to use and interpret.[15–18] As SD fluctuates daily in patients with AD,[19] the SD NRS was therefore developed to provide clinicians and patients with a simple, easy-to-interpret alternative to a VAS that can be used to measure daily fluctuations in SD.

This post hoc analysis of data from a phase IIb trial[20] examined the psychometric properties of the SD NRS and established an initial responder definition for meaningful change in patients with moderate-to-severe AD. The results provided strong support that the SD NRS is reliable, valid and responsive in patients with moderate-to-severe AD. They also extend evidence for the content validity of the SD NRS based on cognitive interviews in patients with moderate-to-severe AD.[19] The SD NRS had very good test–retest reliability, with ICCs in almost all cases above the recommended threshold of 0·70,[37] indicating that it stably measures the SD concept over time.

The study also confirmed that the SD NRS measured the targeted concept of SD associated with AD. As hypothesized, correlations between the SD NRS and sleep-related measures (SCORAD sleep loss VAS and 5-D Itch sleep item) were large. However, the correlations were not exact, indicating that, although the SD NRS measures SD, the three PRO measures are not interchangeable. This is likely due to

Table 6 Responsiveness: association between change in sleep disturbance numerical rating scale (SD NRS) average weekly scores from baseline and changes in other measures from baseline in participants who improved vs. participants who did not improve[a]

		SD NRS, mean (SD)						Overall F-test	
Anchor	N	Baseline	Follow-up	Mean change	Effect size[b]	Standardized response mean[c]	P-value (paired t-test)	Test value	P-value
Week 16									
SCORAD sleep loss VAS								29·3	< 0·001
Improved (change < 0)	145	7·8 (1·6)	2·2 (2·3)	−5·5 (2·6)	3·41	2·17	< 0·001		
Not improved (change ≥ 0)	9	7·8 (1·6)	6·6 (2·0)	−0·9 (1·2)	0·56	0·74	0·057		
PCS (average weekly)								156	< 0·001
Improved (change ≤ −1)	112	7·9 (1·4)	1·7 (1·7)	−6·3 (2·0)	4·54	3·22	< 0·001		
Not improved (change −1 to < 1)	34	7·2 (2·2)	5·4 (2·5)	−1·7 (1·7)	0·77	0·98	< 0·001		
PP NRS (average weekly) score								80·4	< 0·001
Improved (change ≤ −1)	130	7·8 (1·5)	2·0 (2·0)	−5·8 (2·3)	3·81	2·55	< 0·001		
Not improved (change −1 to < 1)	16	7·4 (2·4)	6·6 (2·5)	−0·6 (1·3)	0·25	0·44	0·10		
AP NRS (average weekly)									
Improved (change ≤ −1)	133	7·8 (1·5)	2·1 (2·1)	−5·7 (2·4)	3·79	2·41	< 0·001		
Not improved (change −1 to < 1)	13	7·2 (2·6)	6·5 (2·7)	−0·5 (1·4)	0·18	0·35	0·23		
DLQI (total score)								6·98	0·009
Improved (change < 0)	126	7·8 (1·6)	2·3 (2·3)	−5·5 (2·6)	3·4	2·11	< 0·001		
Not improved (change ≥ 0)	11	8·0 (1·4)	4·6 (3·2)	−3·3 (3·5)	2·33	0·93	0·012		
Week 24									
SCORAD sleep loss VAS								33·6	< 0·001
Improved (change < 0)	157	7·8 (1·6)	2·5 (2·4)	−5·3 (2·6)	3·38	2·06	< 0·001		
Not improved (change = 0)	10	7·5 (2·5)	6·7 (2·5)	−0·5 (1·0)	0·22	0·54	0·12		
PCS (average weekly)								179	< 0·001
Improved (change ≤ −1)	128	7·9 (1·4)	1·6 (1·7)	−6·2 (2·0)	4·51	3·11	< 0·001		
Not improved (change −1 to < 1)	44	7·4 (2·2)	5·7 (2·4)	−1·7 (1·8)	0·74	0·91	< 0·001		
PP NRS (average weekly)								90·9	< 0·001
Improved (change ≤ −1)	152	7·8 (1·5)	2·2 (2·1)	−5·6 (2·3)	3·81	2·42	< 0·001		
Not improved (change −1 to < 1)	19	7·1 (2·6)	6·7 (2·7)	−0·4 (1·2)	0·17	0·37	0·13		
AP NRS (average weekly)								91·5	< 0·001
Improved (change ≤ −1)	152	7·9 (1·4)	2·2 (2·1)	−5·6 (2·3)	3·97	2·42	< 0·001		
Not improved (change −1 to < 1)	19	6·7 (2·7)	6·3 (2·8)	−0·4 (1·2)	0·16	0·36	0·14		
IGA								12·3	< 0·001
Improved (change < 0)	116	7·7 (1·7)	2·2 (2·5)	−5·5 (2·7)	3·31	2·02	< 0·001		
Not improved (change ≥ 0)	55	7·9 (1·6)	4·0 (2·7)	−3·9 (2·7)	2·45	1·46	< 0·001		

AP NRS, Average Pruritus Numerical Rating Scale; DLQI, Dermatology Life Quality Index; IGA, Investigator's Global Assessment; PCS, pruritus categorical score; PP NRS, Peak Pruritus Numerical Rating Scale; SCORAD, Scoring Atopic Dermatitis; VAS, visual analogue score. [a]Definitions of 'worsened' and 'improved' for responsiveness testing are provided in Table S3 (see Supporting Information). [b]Mean difference between the baseline and week 24 average weekly score divided by the standard deviation of the baseline average weekly score of the SD NRS. [c]Mean difference between the baseline and week 24 average weekly score divided by the standard deviation of the change from baseline average weekly score of the SD NRS.

differences in formats (NRS vs. VAS) and recall periods (24 h for SD NRS, 3 days for SCORAD sleep loss VAS, and 2 weeks for 5-D Itch sleep item). Correlations with scales measuring concepts other than SD, such as HADS scores, EQ-5D scores and clinician-reported disease severity scores (SCORAD, BSA, IGA and EASI), were weak, confirming that the SD NRS can complement standard clinician-reported disease severity and other disease impacts (e.g. health status and emotional impact) when assessing the benefit of AD treatments.[4] Of the non-SD PROs, correlations were strongest between the SD NRS and the itch severity scales (PP NRS and AP NRS), confirming that SD is a proximal impact of the primary symptom of AD.[38–40]

The SD NRS was able to discriminate between severity groups of participants according to PRO sleep and quality-of-life scales, but it was not able to discriminate between severity groups of participants according to clinician-reported measures (IGA and EASI for AD severity). This is not surprising because AD severity reported by clinicians does not necessarily translate to symptom severity or impact as perceived and reported by patients; indeed, correlations were low between baseline SD NRS and clinician-reported scales. This should not diminish the ability of the SD NRS to discriminate among severity groups of participants.

The study further showed that the SD NRS was responsive to change. Correlations were large between the changes in SD

Table 7 Responder definitions for the change in sleep disturbance numerical rating scale between baseline and week 24

Definition type	Study	Definition	Meaningful or detectable change threshold estimate
Anchor based	Phase IIb	≥ 1-point decrease in SCORAD sleep loss VAS	
		Week 24	5·6
		Week 16	5·4
	Phase IIb	≥ 4-point decrease in PCS average weekly score	
		Week 24	6·4
		Week 16	6·3
	Phase IIb	≥ 4-point decrease in PCS average weekly score	
		Week 24	6·7
		Week 16	6·7
	Phase IIb	≥ 4-point decrease in DLQI total score	
		Week 24	5·7
		Week 16	ND
Distribution based	Phase IIb	Quarter-standard deviation	0·40
	Phase IIb	Half-standard deviation	0·81
		Standard error of measurement	1·58
Qualitative	Qualitative research[a]	Smallest change considered satisfactory	3
		Meaningful improvement	2

DLQI, Dermatology Life Quality Index; ND, not determined; PCS, pruritus categorical score; SCORAD, Scoring Atopic Dermatitis; VAS, visual analogue score. [a]Dias-Barbosa *et al.*[19]

NRS score and changes in itch, sleep loss and disease severity scores, and they were moderate between the changes in SD NRS score and changes in quality-of-life scores. Further, the SD NRS score changes were able to significantly differentiate between participants whose pruritus, quality of life and disease severity improved and those whose did not. Even though both groups of participants experienced improvements in SD, as indicated by a decrease in their SD NRS score, decreases in the SD NRS were larger in the 'improved' group than in the 'not improved' group.

This study allowed a preliminary responder definition of SD NRS score for meaningful change (2–5 points) to be established in patients with moderate-to-severe AD. This is consistent with responder definitions for other patient-reported 11-point itch NRSs in participants with moderate-to-severe plaque psoriasis (4 points)[41] and moderate-to-severe AD (2–4 points).[23] The relatively wide range in SD NRS for the responder definition was due to the trial including participants with severe itch and severe SD at baseline, which allows substantial room for change.

The current study had some limitations. Firstly, the ability of the SD NRS to assess the less severe spectrum of SD was not examined in the current analysis. However, based on qualitative findings,[19] the SD NRS is expected to perform as well among participants with less severe SD. Secondly, over the course of the trial, approximately 30% of the SD NRS data were missing after 24 weeks of follow-up. Nonetheless, the initial analysis sample of 218 participants, the final sample size (n = 154) and the subject-to-item ratio were adequate to allow the SD NRS data to be analysed at baseline and at follow-up, for the psychometric properties to be confirmed, and for an initial responder definition to be established.[42,43] A third limitation was that, of the various anchors included in the current analysis, a few did not cover the 7-day recall period covered by the SD NRS weekly score. For example, the SCORAD sleep loss VAS has a 3-day recall period, and the IGA does not have a recall period. The slight differences in the recall periods of the SD NRS and these two anchors should not jeopardize the stability assessment because the test–retest analyses were conducted using a short time window, which should prevent substantial changes in the concepts measured by these different anchors.

In conclusion, the current quantitative analysis suggests that the SD NRS is a reliable, fit-for-purpose and well-defined measure of the overall severity of SD specifically adapted to patients with moderate-to-severe AD. The SD NRS should provide clinicians with an alternative to VAS because it is simple and easy to interpret and can be used to assess day-to-day changes in SD related to AD. Along with the core outcome measures recommended by the Harmonising Outcome Measures for Eczema initiative,[44] the SD NRS can be used to assess impact in AD clinical trials. Finally, the current study confirmed a close link between SD and pruritus in patients with AD, and it highlighted the importance of adequately measuring SD as an outcome of AD treatment.

Acknowledgments

Medical writing was provided by Phillip Leventhal, PhD and paid for by Galderma.

References

1 Chang YS, Lin MH, Lee JH *et al.* Melatonin supplementation for children with atopic dermatitis and sleep disturbance: a randomized clinical trial. *JAMA Pediatr* 2016; **170**:35–42.

2 Hon KL, Lam MC, Leung TF *et al.* Assessing itch in children with atopic dermatitis treated with tacrolimus: objective versus subjective assessment. *Adv Ther* 2007; **24**:23–8.
3 Ruzicka T, Mihara R. Anti-interleukin-31 receptor A antibody for atopic dermatitis. *N Engl J Med* 2017; **376**:2093.
4 Barrett A, Hahn-Pedersen J, Kragh N *et al.* Patient-reported outcome measures in atopic dermatitis and chronic hand eczema in adults. *Patient* 2019; **12**:445–59.
5 Food and Drug Administration. Guidance for Industry on Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims. *Fed Regist* 2009; **74**:65132–33.
6 Ji X, Liu J. Subjective sleep measures for adolescents: a systematic review. *Child Care Health Dev* 2016; **42**:825–39.
7 Lewandowski AS, Toliver-Sokol M, Palermo TM. Evidence-based review of subjective pediatric sleep measures. *J Pediatr Psychol* 2011; **36**:780–93.
8 Lei D, Yousaf M, Janmohamed SR *et al.* Validation of four single-item patient-reported assessments of sleep in adult atopic dermatitis patients. *Ann Allergy Asthma Immunol* 2020; **124**:261–6.
9 Jeon C, Yan D, Nakamura M *et al.* Frequency and management of sleep disturbance in adults with atopic dermatitis: a systematic review. *Dermatol Ther (Heidelb)* 2017; **7**:349–64.
10 Lei DK, Yousaf M, Janmohamed SR *et al.* Validation of Patient-Reported Outcomes Information System Sleep Disturbance and Sleep-Related Impairment in adults with atopic dermatitis. *Br J Dermatol* 2020; **183**:875–82.
11 European Task Force on Atopic Dermatitis. Severity scoring of atopic dermatitis: the SCORAD index. Consensus Report of the European Task Force on Atopic Dermatitis. *Dermatology* 1993; **186**:23–31.
12 Stalder JF, Barbarot S, Wollenberg A *et al.* Patient-Oriented SCORAD (PO-SCORAD): a new self-assessment scale in atopic dermatitis validated in Europe. *Allergy* 2011; **66**:1114–21.
13 Peters ML, Patijn J, Lame I. Pain assessment in younger and older pain patients: psychometric properties and patient preference of five commonly used measures of pain intensity. *Pain Med* 2007; **8**:601–10.
14 Herr KA, Mobily PR. Comparison of selected pain assessment tools for use with the elderly. *Appl Nurs Res* 1993; **6**:39–46.
15 Hjermstad MJ, Fayers PM, Haugen DF *et al.* Studies comparing numerical rating scales, verbal rating scales, and visual analogue scales for assessment of pain intensity in adults: a systematic literature review. *J Pain Symptom Manage* 2011; **41**:1073–93.
16 Ismail AK, Abdul Ghafar MA, Shamsuddin NS *et al.* The assessment of acute pain in pre-hospital care using verbal numerical rating and visual analogue scales. *J Emerg Med* 2015; **49**:287–93.
17 Larroy C. Comparing visual-analog and numeric scales for assessing menstrual pain. *Behav Med* 2002; **27**:179–81.
18 Mohan H, Ryan J, Whelan B *et al.* The end of the line? The visual analogue scale and verbal numerical rating scale as pain assessment tools in the emergency department. *Emerg Med J* 2010; **27**:372–5.
19 Dias-Barbosa C, Matos R, Vernon M *et al.* Content validity of a sleep numerical rating scale and a sleep diary in adults and adolescents with moderate-to-severe atopic dermatitis. *J Patient Rep Outcomes* 2020; **4**:100.
20 Silverberg JI, Pinter A, Pulka G *et al.* Phase 2B randomized study of nemolizumab in adults with moderate-to-severe atopic dermatitis and severe pruritus. *J Allergy Clin Immunol* 2020; **145**:173–82.
21 Eichenfield LF, Tom WL, Berger TG *et al.* Guidelines of care for the management of atopic dermatitis: section 2. Management and treatment of atopic dermatitis with topical therapies. *J Am Acad Dermatol* 2014; **71**:116–32.
22 Vocks E, Plotz SG, Ring J. The Dyshidrotic Eczema Area and Severity Index – a score developed for the assessment of dyshidrotic eczema. *Dermatology* 1999; **198**:265–9.
23 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**:761–9.
24 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.
25 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.
26 EuroQol Research Foundation. *EQ-5D-3L User Guide*. Rotterdam: EuroQoL Research Foundation, 2018.
27 Finlay AY, Khan GK. Dermatology Life Quality Index (DLQI) – a simple practical measure for routine clinical use. *Clin Exp Dermatol* 1994; **19**:210–16.
28 Zigmond AS, Snaith RP. The hospital anxiety and depression scale. *Acta Psychiatr Scand* 1983; **67**:361–70.
29 Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. *Psychol Bull* 1979; **86**:420–8.
30 Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *J Chiropr Med* 2016; **15**:155–63.
31 US Food and Drug Administration. *Patient-focused drug development: collecting comprehensive and representative input*. Rockville, MD: US FDA. Available at: https://www.fda.gov/regulatory-information/search-fda-guidance-documents/patient-focused-drug-development-collecting-comprehensive-and-representative-input (last accessed 18 October 2021).
32 Oranje AP, Glazenburg EJ, Wolkerstorfer A *et al.* Practical issues on interpretation of scoring atopic dermatitis: the SCORAD index, objective SCORAD and the three-item severity score. *Br J Dermatol* 2007; **157**:645–8.
33 Zisapel N, Nir T. Determination of the minimal clinically significant difference on a patient visual analog sleep quality scale. *J Sleep Res* 2003; **12**:291–8.
34 Basra MK, Salek MS, Camilleri L *et al.* Determining the minimal clinically important difference and responsiveness of the Dermatology Life Quality Index (DLQI): further data. *Dermatology* 2015; **230**:27–33.
35 Kawata AK, Revicki DA, Thakkar R *et al.* Flushing ASsessment Tool (FAST): psychometric properties of a new measure assessing flushing symptoms and clinical impact of niacin therapy. *Clin Drug Investig* 2009; **29**:215–29.
36 Cohen J. A power primer. *Psychol Bull* 1992; **112**:155–9.
37 Nunnally JC, Bernstein IH. *The Assessment of Reliability. Psychometric Theory*. New York: McGraw-Hill, 1994.
38 Huet F, Faffa MS, Poizeau F *et al.* Characteristics of pruritus in relation to self-assessed severity of atopic dermatitis. *Acta Derm Venereol* 2019; **99**:279–83.
39 Kaaz K, Szepietowski JC, Matusiak L. Influence of itch and pain on sleep quality in atopic dermatitis and psoriasis. *Acta Derm Venereol* 2019; **99**:175–80.
40 Silverberg JI, Kantor RW, Dalal P *et al.* A comprehensive conceptual model of the experience of chronic itch in adults. *Am J Clin Dermatol* 2018; **19**:759–69.
41 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.

42 Anthoine E, Moret L, Regnault A *et al.* Sample size used to validate a scale: a review of publications on newly-developed patient reported outcomes measures. *Health Qual Life Outcomes* 2014; **12**:176.
43 Frost MH, Reeve BB, Liepa AM *et al.* What is sufficient evidence for the reliability and validity of patient-reported outcome measures? *Value Health* 2007; **10** (Suppl. 2):S94–105.
44 Schmitt J, Apfelbacher C, Spuls PI *et al.* The Harmonizing Outcome Measures for Eczema (HOME) roadmap: a methodological framework to develop core sets of outcome measurements in dermatology. *J Invest Dermatol* 2015; **135**:24–30.
45 Cohen J. *Statistical Power Analysis for the Behavioral Sciences*, 2nd edn. Hillsdale, NJ: Lawrence Erlbaum Associates Inc., 1988.

Supporting Information

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

Table S1 Outcome measures, recall periods, scales, and frequencies of assessments.

Table S2 Raw and refined categories, grouping, and explanations for known-groups validity testing.

Table S3 Definitions of 'worsened' and 'improved' for responsiveness testing.