Patient-reported outcome (PRO) instruments for disease severity and quality of life in patients with atopic dermatitis: a systematic review of English and Chinese literature

Aihua Li[1], Minlu Zhang[1], Yating Yang[2], Joshua Zhang[1], Xiaoping Xie[1], Xiaoxian Li[2], Hui Zhang[3]

[1]Sanofi (China) Investment, Shanghai, China; [2]Shanghai Jusure Health Technology, Shanghai, China; [3]Department of Dermatology, Xinhua Hospital Affiliated to Shanghai Jiaotong University School of Medicine, Shanghai, China

Contributions: (I) Conception and design: A Li, H Zhang; (II) Administrative support: M Zhang, X Li, X Xie; (III) Provision of study materials or patients: A Li, M Zhang, Y Yang, J Zhang; (IV) Collection and assembly of data: Y Yang, A Li, X Li, J Zhang; (V) Data analysis and interpretation: Y Yang, A Li, H Zhang; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

Correspondence to: Dr. Hui Zhang. Xinhua Hospital Affiliated to Shanghai Jiaotong University School of Medicine, Shanghai 200092, China. Email: zhanghui@xinhuamed.com.cn.

Background: Many patient-reported outcome (PRO) on disease severity quality of life (QOL) have been developed for atopic dermatitis (AD) patients. However, none of them on the reliability and validity of the instruments was sufficient for their application in clinical studies. The objective of this study is to identify and assess the quality of recently developed PROs for disease severity and QOL in English and Chinese in AD patients.

Methods: We conducted a systematic review of PROs for disease severity and QOL for AD from PubMed, Web of Science, PsycINFO and ERIC (English literatures), and CNKI and Wanfang Data (Chinese literatures) from September 2010 to December 2021 with string including "atopic dermatitis" and "scaling". All studies were screened by 2 reviewers. After being removed duplications, the studies developed the instruments for the AD patients, were reported by patients, and assessing the disease severity or QOL were included.

Results: Twenty-six instruments were retrieved. Three single-item Numeric Rating Scale (NRS) and 8 multidimensional instruments assessing disease severity and 15 assessing QOL were found to be originally developed in English (n=23) or Chinese (n=3). After full assessment on the reliability and validity, 3 NRS and 9 multidimensional instruments were recommended. The 3 NRS were Peak Pruritus/Itch NRS, Skin Pain NRS and Sleep Disturbance (SD) NRS. The multidimensional instruments for disease severity included the Patient-Oriented Eczema Measure (POEM), the patient oriented-SCORAD (PO-SCORAD), and Atopic Dermatitis Symptom Score (ADSS), and the instruments for QOLs included Infant's Dermatology Quality of Life Index (IDQOL), Children's Dermatology Life Quality Index (CDLQI), Atopic Dermatitis Control Tool (ADCT), PROMIS® Itch Questionnaire Mood and Sleep (PIQ-MS), PROMIS-Sleep Disturbance (PROMIS-SD), and PROMIS-Sleep-Related Impairment (PROMIS-SRI) for QOL. However, none of the Chinese PROs either originally developed or adapted were fully validated.

Discussion: Single-item NRS is a complement to multidimensional PROs in assessing the disease severity of AD. Quality of these instruments vary greatly and only a few instruments that meet the Consensus-based Standards for the Selection of Health Measurement Instruments (COSMIN) standards are recommended. Therefore, standardization of PROs is essential for developing new instruments, and for adapting a PRO in other populations with different culture and languages.

Keywords: Atopic dermatitis (AD); patient report outcome (PRO); quality of life (QOL)

Introduction

Atopic dermatitis (AD) is a common major pruritic skin condition that leads to a substantial burden on patients, their families, and society (1-3). AD affects ~20% of children and 1–10% of adults with an increasing prevalence worldwide (4). AD most often begins in infancy or early childhood, with ~90% of cases appearing within the first 5 years of live (5). The course of AD can be long-lasting, relapsing, and often significantly affects the quality of life (QOL) of patients and their families. In the past a few years, increasing attention has been given to AD, with the development of various treatments and therapies, including biological therapies (6). However, the outcome measurements for assessing AD are usually based on clinical signs, as currently few satisfactory objective marker of disease activity is adequate and reliable to be used as a golden standard (7). Patient-reported outcome (PRO) is able to reflect any status of a patient's health condition that immediately provided by the patient. And PRO avoids the secondary changes and misunderstanding of the patient's responses by any third party (8). Therefore, the disease severity related symptoms that are directly reported by patients are essential for assessing the efficacy of the treatment in patients with AD.

The Food and Drug Administration (FDA) Guidelines for Industry recommend the primary endpoint of AD treatment success to be based on the Investigator's Global Assessment (IGA) score difference, which is a clinical assessment scale that depends on physical symptoms assessed by physicians (9). This clinician-rated scale is intended to be objective measures of highly visible symptoms (e.g., redness, flaking, bleeding from scratching) and measurable functional impairment. Unfortunately, this approach has limited patient input on treatment outcomes in AD. Furthermore, the IGA was defined by a particular study sponsor in a particular context, resulting in variation in IGA versions and IGA has not been adequately validated until a standardized vIGA is published (10). Importantly, AD often causes constant, intensive itching, and impaired psychosocial and working functioning (1). Psychiatric comorbidities, including depression, anxiety, and suicidal ideation, are more common in AD patients than general population, even among patients with clinically mild or moderate conditions (11). Therefore, despite the physical burden of AD, healthcare providers may underestimate the psychological effect of the disease (1). On the other hand, key symptoms and impacts of AD, such as pruritus,

sleep disturbance, and interference with activities, are more difficult or impossible for clinicians to assess. Furthermore, the meaningfulness of clinical improvement can only be assessed by patients (12). Unfortunately, PROs have been used in some forms in only a small portion of clinical trials (13). As the principle of patient-centered care is becoming increasingly recognized and valued, instruments that directly assess the impact of AD on the QOL of patients are needed to determine the effectiveness of treatments as stated in the 21st century Cures Act (14). Therefore, PROs, which collects disease-related information directly from the patient without any interpretation, are an important complement to the clinician-reported outcomes and are increasingly expected and considered for the evaluation of treatment outcomes (1). Meanwhile, unlike a survey questionnaire, PROs are conducted by experts and their content covers every aspect of the patient's experience, including symptom burden, mood, physical function, QOL, and distress (15). Using the information extracted from the PROs, healthcare providers can provide more patient-focused and specific therapies, which meet the requirement of the Patient-Focused Drug Development guideline issued by the US Food and Drug Administration in June, 2020 (16).

The rapid development of new treatments for AD, such as alefacept, dupilumab, and upadacitinib, has also highlighted the need for a generalizable scale to assess the severity and progress of AD to evaluate and compare treatment effectiveness (17-19). At present, the Eczema Area and Severity Index (EASI) and IGA instruments are the primary endpoints in most clinical studies of AD, but neither was included in this review because they are completed by healthcare providers. Increasingly, more clinical trials are utilizing PROs as an endpoint for testing drug efficacy and safety. There are a variety of instruments aiming at quantifying AD outcomes. A patient-reported outcome measures (PROM) is a questionnaire used to elicit information directly from patients, covering the measures of symptoms, activity limitations, health status, health-related quality of life (HRQOL), QOL, etc. (20). As physiological and psychological burdens often occur concurrently for patients with AD, both disease severity and QOL measures are fundamental to patient evaluation and care (21). Therefore, this study focused on two aspects of PRO: disease severity and QOL.

A previous systematic review published in 2011 identified a total of 20 disease severity scales and 14 QOL instruments in English used in randomized controlled trials (RCTs) of AD treatment from 1985 to 2010 (22). In

Table 1 Search strategy

#	Search strategy	No. of records
1	(Atopic Dermatit*) OR (Atopic Neurodermatit*)	16,206 (PubMed); 9,922 (WoS); 167 (PsycINFO); 167 (ERIC); 112,111 (CNKI); 15,104 (Wanfang)
2	(Animals) OR (canine OR dog OR dogs OR cat OR cats)	1,928,982 (PubMed); 611,382 (WoS); 58,559 (PsycINFO); 58,559 (ERIC); 4,455,372 (CNKI); 690,783 (Wanfang)
3	1 NOT 2	9,893 (PubMed); 9,034 (WoS); 162 (PsycINFO); 162 (ERIC); 36,602 (CNKI); 14,584 (Wanfang)
4	(questionnaire*) OR (scal*) OR (assessment) OR (indicato*) OR (measur*) OR (scor*)	3,838,315 (PubMed); 5,936,191 (WoS); 627,265 (PsycINFO); 627,265 (ERIC); 2,596,416 (CNKI); 2,581,210 (Wanfang)
5	3 AND 4	4,372 (PubMed); 3,297 (WoS); 69 (PsycINFO); 69 (ERIC); 1,332 (CNKI); 1,293 (Wanfang)

Database(s): PubMed, WoS, PsycINFO and ERIC (English); CNKI and Wanfang Data (Chinese). Publication period: September 2010 to 31 December 2021.

a more recent review published in 2016, 62 disease severity measures and 28 QOL instruments in English were identified and analyzed (23). However, in such a rapid developing field, there are numerous new instruments and follow-up studies assessing the validity and reliability of existing instruments, but none has delivered a complete and precise evaluation of the quality of the instruments. Therefore, the identification of new instruments and reassessment of the quality of the existing ones are crucial for choosing the appropriate instrument for use in clinical studies and practice.

There are significantly fewer AD assessment instruments in Chinese and they often lack validation and reliability testing after translation. In China, the prevalence of AD has increased from 0.7% (age 6–20 years) in 2000 to 8.3% in Shanghai in 2012 (age 3–6 years) (24). The incidence of AD in outpatients has also dramatically increased from 2.3% in 2008 to 7.8% in 2016 (24,25). In addition to changes in environmental factors and lifestyle, the increased AD incidence rate is largely contributed to a correction in diagnostic methods, which historically have caused overdiagnosed eczema and underdiagnosed AD (21). The lack of an accurate gold standard to differentiate the two diseases further challenges comparability between AD trials in China. The treatment guideline of AD in China has been a combination of topical corticosteroid (TCS) and traditional Chinese medicine (TCM) (26). TCM has been used especially to treat children aged 0–12 years in clinical practice, mostly aiming to reduce the use for TCS (27). As 31–36% of TCM users are combining it with TCS, unique instruments that offers insights to the prescription of both TCM and

TCS would be a good reference for clinicians and also good candidates for further studies (28).

We aimed to assess the quality of existing PROs in English and Chinese in patients with AD which measure the disease severity and QOL instruments by systematically reviewing the instrument development and validation literature published between September 2010 and December 2021. Specifically, we sought to (I) evaluate the measurement properties of the outcome measurements of commonly used instruments in both Chinese and English; (II) identify the gaps in instrument translation, adaptation and validation; and (III) prioritize future validation studies of instruments to assess disease severity and QOL of AD. We present the following article in accordance with the PRISMA reporting checklist (available at https://atm.amegroups.com/article/view/10.21037/atm-22-3164/rc).

Methods

Search strategy and eligibility criteria

A comprehensive systematic literature search was carried out in PubMed, Web of Science, PsycINFO and ERIC (for literature in English), and in CNKI and Wanfang Data (for literature in Chinese). In order to capture recently used instruments, the search was limited to studies published from September 2010 to December 31, 2021. A specific search string including search terms of "atopic dermatitis" and "scaling" was developed (*Table 1*). Included studies were full-text papers with human subjects and with the aim of developing or validating an instrument to measure symptoms of AD and the QOL of patients with AD. Instruments that had not been validated were ineligible.

Study selection

Inclusion criteria

(I)   AD or eczema was included in the target population of the instrument.
(II) Instruments should be patient-reported outcomes or contain at least one domain that is self-reported.

Exclusion criteria

(I)   Patient was not an accessor of the instrument (i.e., instrument was not self-reported).
(II) Instrument was not developed for measurement purposes.
(III) No psychometric validation was available for the instrument or the instrument had poor psychometric properties.
(IV) Instruments not for assessing the disease severity or the QOL of patients with AD.

Two reviewers (YY and XL) screened the titles and abstracts, and then assessed the full-text for eligibility.

Data extraction

All instruments that met the inclusion criteria were extracted and two types of information were recorded: basic information and instrument properties.

For the basic information, name, symptoms assessed, target population, assessor, number of items and components, rating method, score algorithm and available translation were listed. For the instrument properties, internal consistency, reliability measurement error, content validity, construct validity, cross-cultural validity, and responsiveness were recorded.

Assessment of measurement properties of instruments of AD

The measurement properties assessed in this study were selected based on the recommendations of the Consensus-based Standards for the Selection of Health Measurement Instruments (COSMIN) group, which included reliability (internal consistency, reliability and measurement error), validity (content and construct validity) and responsiveness (29) (*Table 2*).

The instruments assessed were then placed in one of the three recommendation categories based on the sufficiency of their measurement properties as suggested by the COSMIN group (29). An instrument was placed in category A if there was evidence for sufficient content validity (any level) and at least low-quality evidence of sufficient internal consistency. An instrument was placed in category C if there was high-quality evidence of an insufficient measurement property. If an instrument could not be categorized as A or C, it was placed in category B. Instruments categorized as 'A' are recommended for use, and results obtained with these instruments can be seen as trustworthy. Instruments categorized as 'B' have the potential to be recommended for use, but require further validation. Instruments categorized as 'C' are not recommended for use. If only PROMs categorized as 'B' are found in a review, the one with the best evidence for content validity is the one to be provisionally recommended for use, until further evidence is found (29).

Results

Study characteristics

After filtering out all database a total number of 10,432 studies were evaluated, of which 9,349 were non-duplicate records according to the PRISMA statement (*Figure 1*) (30). After screening, 38 studies in English and 2 in Chinese were included in this systematic review. The 40 articles covered 26 instruments, which comprised 11 different instruments for disease severity and 15 for QOL that met the inclusion and exclusion criteria. All studies used at least 1 disease severity scale. Only 13 (37%) studies used QOL instruments and 9 (25%) studies used ≥1 QOL measurement. We also found 1 article for 1 instrument by manually searching the relevant reference.

Content of the English instruments identified

A total of 11 disease severity scales, including 3 single-item (unidimensional) Numeric Rating Scales (NRS) and 8 multidimensional scales, were utilized in the included studies. The most commonly used disease severity instrument in English was Peak Pruritus/Itch NRS which was used in more than 10 clinical studies (14,31,32). The second most common disease severity instrument was the Scoring Atopic Dermatitis Index (SCORAD), which was used in 7 studies (33). The next two commonly used disease severity instruments were the Patient-Oriented Eczema Measure (POEM) and the Objective SCORAD, both of which were used in 5 studies (33,34). These were closely followed by the Three Item Severity score (n=4) (35,36).

Table 2 Methods and definitions in evaluating the measurement properties

Measurement property	Measurement property name	Criteria for adequate rating (+)	Criteria for intermediate rating (?)	Criteria for inadequate rating (–)
Reliability	Internal consistency	Cronbach's α calculated per dimension AND Cronbach's α 0.70–0.95	Unclear whether the instrument is unidimensional OR doubtful design or method	Cronbach's α not calculated per dimension despite being a unidimensional instrument OR Cronbach's α <0.70 or >0.95
	Reliability	Pearson's R >0.80 OR weighted κ >0.60 OR coefficient of variation <20% OR ANOVA <10%	Reliability not evaluated OR (Pearson's R 0.60–0.80 OR weighted κ 0.40–0.60 OR coefficient of variation 20–30% OR ANOVA 10–20%)	Pearson's R <0.60 OR weighted κ <0.40 OR coefficient of variation >30% OR ANOVA >20%
	Measurement error	SEM, SDC or limits of agreement was calculated OR both positive and negative PA was calculated	SEM, SDC or LoA can be calculated from the given data OR PA was calculated	SEM was calculated based on Cronbach's α or SD from another population AND PA was not calculated
Validity	Content validity	Professionals OR patients were involved in item selection AND professionals AND patients considered >90% of items to be relevant, comprehensive and understandable	Professionals OR patients were involved in item selection AND professionals AND patients considered 70–89% of items to be relevant, comprehensive and understandable	NEITHER professionals NOR patients were involved in item selection OR professionals OR patients considered <70% of items to be relevant, comprehensive and understandable
	Construct validity	Factor analysis performed with adequate sample size (>7-fold the number of items AND >100) AND two different instruments that aims to measure signs of AD show high correlation (correlation coefficient >0.70)	No factor analysis OR factor analysis performed with intermediate sample size (>5-fold the number of items AND <100) OR two different instruments that measure signs of AD show correlation (correlation coefficient 0.60–0.69)	Factor analysis performed with inadequate sample size (<5-fold the number of items) OR two different instruments that measure signs of AD do not show correlation (correlation coefficient <0.50)
	Cross-cultural validity	Instrument functions in the same way in different translated versions across different samples of respondents	No translated versions	Instrument does not function in the same way in different translated versions across different samples of respondents
Responsiveness	The correlation between changes from baseline in a PRO score with changes from baseline in other PROs or outcomes	Moderate to strong correlation	Weak correlation	No correlation
Meaningful change estimation	Threshold	ROC curve was plotted OR (MIC defined AND MIC > SDC)	MIC undefined	ROC curve was not plotted AND (MIC defined AND MIC ≤ SDC)

+, sufficient; –, insufficient; ?, indeterminate. PRO, patient-reported outcome; ANOVA, analysis of variance; AD, atopic dermatitis; SEM, standard error of measurement; SDC, smallest detectable change; PA, percentage agreement; ROC, receiver operating characteristic; MIC, minimal important change.
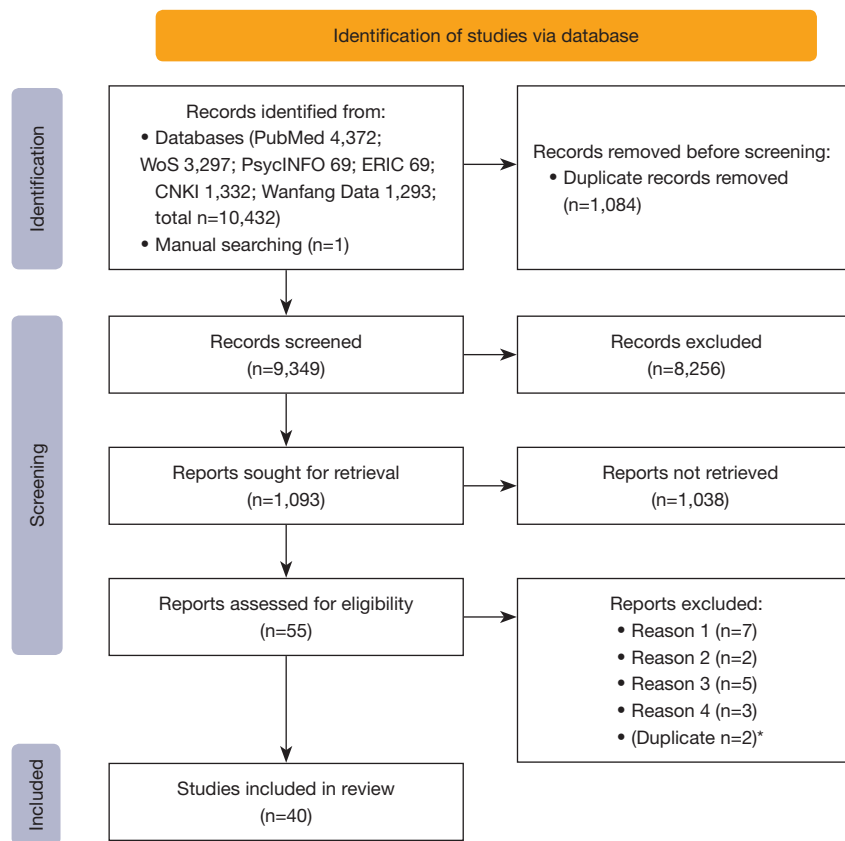
Figure 1 Flow chart of the search procedure adapted from the 2020 PRISMA statement. Reason 1 refers to 'Instruments were not patient reported outcomes or contain at least one domain that is self-reported'; Reason 2 refers to 'Instrument was not developed for measurement purposes'; Reason 3 refers to 'No psychometric validation was available for the instrument or the instrument had poor psychometric properties'; Reason 4 refers to 'Instruments not for assessing the QOL or the disease severity or the QOL of patients with AD'. *, duplicate refers to overlapped studies across reason 1 to 4. QOL, quality of life; AD, atopic dermatitis.

No discernable trend was found in the use of the top 5 instruments and their publication years.

Twelve QOL instruments in the included studies were analysed. The most frequently used English QOL instrument was the Dermatology Life Quality Index (DLQI), which was used in 6 of the included studies (37). Following that were the Children's Dermatology Life Quality Index (CDLQI), Infant's Dermatology Quality of Life Index (IDQOL) and Patient-Reported Outcomes Measurement Information System (PROMIS®) Itch Questionnaire (PIQ), all of which were used in 3 studies (19,38,39).

Intensive itch, skin pain and related sleeping disturbance are highly prevalent symptoms in the patients with AD that impact both physical and mental functioning. Three single-item NRS (Peak Pruritus/Itch NRS, Skin Pain NRS, and Sleep Disturbance NRS) are developed and validated to

measure a specific aspect of AD severity that are meaningful to patients (31,32,40). The multidimensional disease severity instruments included 3–10 clinical signs. Among them erythema was the most frequently included item (mentioned in 7/8 disease severity instruments), followed by oedema (6/8) and lichenification (5/8). *Table 3* presents the characteristics of the included instruments.

Content of the Chinese instruments identified

The 2 included studies in Chinese reported 13 different instruments to assess the clinical signs of AD. Of 13 instruments, 3 were originally developed in Chinese for disease severity in patients with AD, and 10 were Chinese versions of the Quality of Life Index for Atopic Dermatitis (QoLIAD), IDQOL. None of the self-developed original

Table 3 Characteristics of included instruments

Instrument	Symptoms assessed	Target population	Assessor	No. of components (c) and/or items (i)	Rating method	Scoring algorithm	Available translations
Peak Pruritus/ Itch NRS	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?	Moderate-to-severe AD	Patient	1 (i)	11-point NRS	Daily scores are averaged over 1-week interval	NA
Skin Pain NRS	The patients were asked to select a number from 0 ("no pain") to 10 ("worst pain imaginable") that best described the worst level of skin pain in the past 24 hours	Moderate-to-severe AD	Patient	1 (i)	11-point NRS	Daily scores are averaged over 1-week interval	NA
DS NRS	On a scale of 0–10, with 0 being "no sleep loss related to the symptoms of AD" and 10 being "I did not sleep at all due to the symptoms of AD", how would you rate your sleep last night?	Moderate-to-severe AD	Patient	1 (i)	11-point NRS	Daily scores are averaged over 1-week interval	NA
POEM	Frequency of pruritus, sleep disturbance, bleeding, weeping or oozing, cracking, flaking, dryness or roughness	Eczema	Patient or parent/caregiver	7 (i)	5-point VAS	Sum score (range, 0–28)	38 (including simplified and traditional Chinese)
TIS	Scoring of erythema, oedema and excoriation	AD	Patient	3 (i)	3-point NRS	Sum score (range, 0–9)	NA
PO-SCORAD	Affected surface area; severity of dryness, redness, swelling, crusting/oozing, scratching, thickening; severity of itching, sleep disturbance	AD	Patient or parent/caregiver	3 (c), 9 (i)	Component 1: shading and describing, physician estimates %; component 2: 4-point NRS; component 3: VAS; 100-mm VAS	NA	36 (including simplified and traditional Chinese)
ADIS	Pruritus (severity, timing); sleep disturbance	AD	Patient	2 (c), 8 (i)	10-point NRS and VRS	NA	NA
ADerm-SS	Itch during sleep hours, itch during awake hours, skin pain; intensity of skin cracking, pain caused by skin cracking, dry skin, skin flaking rash, skin thickening, bleeding, skin oozing	AD	Patient	2 (c), 11 (i)	11-point NRS	Sum score	NA
ADSS	Subjective symptoms (itching, sleep disturbance); objective signs (erythema, dryness, oozing, edema)	AD	Patient or parent/caregiver	6 (i)	VAS	Sum score (range, 0–24)	NA

Table 3 (continued)

Table 3 (*continued*)

Instrument	Symptoms assessed	Target population	Assessor	No. of components (c) and/or items (i)	Rating method	Scoring algorithm	Available translations
Rajka-Langeland severity score	Eczema extent, course, and intensity (sleeplessness)	AD	Patient	3 (c),	3-point VRS; 11-point NRS; Sum score VAS	Sum score	NA
ZRADSQ	AD symptoms (itching aggravated at night, itching, dry skin, itching accompanied by pain, burning skin, insomnia), heat (thirst, mouth dryness, constipation, dark urine), mood (fidgeting, irascibility)	AD	Patient or parent/caregiver	15 (i)	4-point VRS, 10 cm VAS	Sum score	Chinese, English
DLQI	Symptoms and feelings, daily activities, leisure, work and school, personal relationships, treatment	Patients with pruritus	Patient	10 (i)	4-point VRS	Sum score (range, 0–30)	124 (including simplified and traditional Chinese)
IDQOL	Itching and scratching, mood, time to sleep, sleep disturbances, disturbed playing, disturbed family activities, problems during meal times, problems from treatment, dressing problems, problems at bath time	AD	Parent/caregiver	10 (i)	3-point NRS	Sum score (range, 0–30)	33 (including simplified and traditional Chinese)
CDLQI	Symptoms and feelings, leisure, school or holidays, leisure, personal relationships, sleep, treatments	AD	Patient with the help of parent/caregiver	10 (i)	3-point NRS	Sum score (range, 0–30)	113 (including simplified and traditional Chinese)
PIQ-MS	General concerns, mood and sleep, clothing and physical activity, scratching behaviours	Patients with pruritus	Patient	4 (c), 63 (i)	4-point VRS	Sum score	NA
5-D itch	Duration, degree, direction, disability (sleep, leisure/social, housework/errands, work/school), distribution	Patients with pruritus	Patient	5 (c), 9 (i)	5-point VRS, check-box	Sum score (range, 5–25)	21 (including traditional Chinese)
ItchyQoL	Symptoms, functional limitations, emotions	Patients with pruritus	Patient	3 (c), 22 (i)	5-point VRS	Sum score (range, 22–110)	11
SF-12	Limitations in physical activities, social activities, and usual role activities; bodily pain, general mental health, vitality, general health perceptions	General patients	Patient	12 (i)	VRS	Sum score	141 (including simplified and traditional Chinese)

Table 3 (*continued*)

Table 3 (*continued*)

Instrument	Symptoms assessed	Target population	Assessor	No. of components (c) and/or items (i)	Rating method	Scoring algorithm	Available translations
DFI	Housework, food, sleep, leisure, time shopping, expenditure, tiredness, distress, relationships, treatment	Eczema, AD	Parent	10 (i)	4-point VRS	Sum score (range, 0–30)	30 (including Chinese)
ABS	Family life, budget & work, daily life, treatment	AD	Parent	14 (i)	6-point VRS	Sum score	NA
ABS-A	Daily life, economic constraints, care & management of disease, work and stress	AD	Patient	18 (i)	6-point VRS	Sum score	NA
ADerm-IS	Sleep impact of AD (difficulty falling asleep, effect on sleep, bothersomeness of waking up at night); daily impacts of AD (limitation in household activities, physical activities, and social activities; difficulty concentrating, feeling self-conscious, feeling embarrassed, feeling sad)	AD	Patient	2 (c), 10 (i)	11-point NRS	Sum score	NA
ESS	Erythema; edema, induration, or papules; excoriation; oozing, weeping, or crusting; scaling; and lichenification	AD	Patient	6 (i)	3-point VRS, 6-point NRS	Proportional score (range, 0–6) multiply by sum score of clinical signs (range, 0–3); total range, 0–108	NA
PROMIS SD	Sleep disturbance	General patients	Patient	8 (i)	NRS	Sum score	NA
PROMIS SRI	Sleep-related impairment	General patients	Patient	8 (i)	NRS	Sum score	NA
ADCT	Symptom severity, itch, bother, impacts on AD sleep, daily activities, mood and emotions	AD	Patient	6 (c), 6 (i)	5-point VRS	Sum score	50 (including simplified Mandarin)

AD, atopic dermatitis; VAS, Visual Analogur Scale; NRS, numeric rating scale; VRS, visual rating scale; NA, not applicable.

instruments had been properly tested for reliability, validity or responsiveness.

It's worth noting that, the Quality of Life Scale for Chronic Eczema Patients-Prior Test Version (EQOLS) was the only instrument based on TCM features. Although it was initially tested for test-retest reliability, construct validity and criterion-related validity, no other studies providing reliable assessment of this instrument were found. The quality of this instrument was therefore not rated in this study.

Psychometric properties of the instruments identified

We summarized the psychometric properties of the 26 included instruments for AD and the recommendations were made based on the COSMIN checklist in *Table 4*. Three single-item NRS were reliable, valid, and responsive with a meaningful threshold (31,32,40). They were recommended to be in category A. Among the 23 multidimensional instruments, 9 (39.1%) were category A (34,35,38,39,41-57), 6 (26.1%) category B (36,48,58-62), and 8 (34.8%) were category C (37,39,46,51,54,63-78). None of the included multidimensional instruments has been proved to be sufficient in all of the assessed measurement properties. There was evidence for all of the assessed instruments for sufficient content validity in patients with AD. The most common measurement property considered to be insufficient was measurement error (found in 3 instruments), following by internal consistency and responsiveness (each found in 2 instruments).

Although further validation was needed for the assessed multidimensional instruments, it is still possible for the instruments in category A to be recommended, namely POEM, patient-oriented SCORAD (PO-SCORAD), Atopic Dermatitis Symptom Score (ADSS), IDQOL, CDLQI, PROMIS® Itch Questionnaire Mood and Sleep (PIQ-MS), PROMIS-Sleep Disturbance (PROMIS-SD) and PROMIS-Sleep-Related Impairment (PROMIS-SRI). Among those, POEM and PO-SCORAD were the most valid and reliable instruments to assess the disease severity of AD, and IDQOL and CDQOL were the most valid and reliable QOL instruments.

Both POEM and PO-SCORAD had adequate content validity (34,46,64) and were highly correlated (r=0.75–0.79) (43), so both are adequate in terms of construct validity. POEM and PO-SCORAD showed evidence for sufficient internal consistency, with the Cronbach α

being 0.86–0.88 and 0.84, respectively (34,46). POEM showed evidence for sufficient responsiveness, as the area under the receiver operating characteristic (ROC) curve was 0.67 and the minimal clinically important difference (MCID) was 3.4 (44). Neither instrument was examined for reliability or measurement error (42).

The IDQOL and CDLQI showed high-quality evidence for adequate content validity and construct validity, and both were considered to be internally consistent, because the Cronbach α of IDQOL was 0.89 and that of CDLQI was 0.83–0.87 (49,52). Both instruments had adequate test-retest reliability, with the Spearman's rank order correlation coefficient ranging from 0.74 to 0.97 for IDQOL and from 0.73 to 0.92 for CDQOL (49,53).

The quality of evidence for sufficient content validity, reliability and responsiveness was low for 4 instruments (ADSS, PIQ-MS, PROMIS-SD and PROMIS-SRI). Although these instruments were still placed in category A and therefore recommended, their measurement properties were all assessed in only one study per instrument.

Seven instruments [Objective SCORAD, TIS, Atopic Dermatitis Itch Scale (ADIS), Zheng-Related Atopic Dermatitis Symptom Questionnaire (ZRADSQ), Atopic Dermatitis Symptom Scale (ADerm-SS), Atopic Dermatitis Impact Scale (ADerm-IS) and 5-dimensions itch scale (5-D itch)] were adequate in some measurement properties but their overall performance was unclear. It is still possible that these instruments could be recommended when more validation studies are available.

Nine instruments [SCORAD, Rajka-Langeland severity score, DLQI, Itchy Quality of Life (ItchyQoL), Short-Form 12 items (SF-12), Dermatitis Family Impact questionnaire (DFI), Atopic dermatitis Burden Scale (ABS), Atopic Dermatitis Burden Scale for Adults (ABS-A) and Epworth Sleepiness Scale (79)] had inadequate quality in at least one of the assessed measurement properties and therefore are not recommended for use in clinical settings. Note that in a study conducted by Liu *et al.* in 2016 (66), the Chinese translated version of DLQI showed poor fit to the Rasch model in Chinese patients, indicating insufficient structural validity for the translated instrument.

Description of the recommended instruments

The Peak Pruritus/Itch NRS is a single self-reported item designed to measure peak pruritus, or worst itch, over the past 24 hours based on the following question: "on a scale of 0 to 10, with 0 being "no itch" and 10 being "worst itch

Table 4 Measure properties of the instruments included and degrees of recommendation

Instrument	Internal consistency	Reliability	Measurement error	Content validity	Construct validity	Responsiveness	Recommendation*
Peak Pruritus/Itch NRS	NA	+ (Yosipovitch et al., 2019)	+ (Yosipovitch et al., 2019)	+ (Yosipovitch et al., 2019; Newton et al. 2019)	+ (Yosipovitch et al., 2019)	+ (Yosipovitch et al., 2019)	A
Skin Pain NRS	NA	+ (Silverberg et al., 2021)	+ (Silverberg et al., 2021)	+ (Newton et al. 2019)	+ (Silverberg et al., 2021)	+ (Silverberg et al., 2021)	A
DS NRS	NA	+ (Puelles, et al., 2022)	+ (Puelles, et al., 2022)	+ (Dias-Barbosa C et al., 2020)	+ (Puelles, et al., 2022)	+ (Puelles, et al., 2022)	A
POEM	+ (Charman et al., 2004; Gerbens et al., 2017; Silverberg et al., 2020)	? (Charman et al., 2004; Gerbens et al., 2017)	? (Charman et al., 2004)	+ (Charman et al., 2004)	+ (Charman et al., 2004; Coutanceau & Stalder, 2014)	+ (Schram et al., 2012)	A
TIS	–	? (Wolkerstorfer et al., 1999)		+ (Wolkerstorfer et al., 1999)	+ (Charman et al., 2005)		B
PO-SCORAD	+ (Stalder et al., 2011; Coutanceau & Stalder, 2014; Silverberg et al., 2020)	? (Stalder et al., 2011; Silverberg et al., 2020)		+ (Stalder et al., 2011)	+ (Stalder et al., 2011; Silverberg et al., 2020)		A
ADIS				+ (Martin et al., 2020)	+ (Martin et al., 2020)		B
ADerm-SS				+ (Foley et al., 2019)			B
ADSS	+ (Lee et al., 2018)	+ (Lee et al., 2018)	+ (Lee et al., 2018)	+ (Lee et al., 2018)	+ (Lee et al., 2018)	+ (Lee et al., 2018)	A
Rajka-Langeland severity score	– (Gånemo et al., 2016)	+ (Gånemo et al., 2016; Silverberg et al., 2020a)		+ (Rajka & Langeland, 1989; Gånemo et al., 2016)	+ (Silverberg et al., 2020a)	+ (Silverberg et al., 2020a)	C
ZRADSQ	+ (Wu et al., 2013)	? (Wu et al., 2013)	? (Wu et al., 2013)	? (Wu et al., 2013)	? (Wu et al., 2013)	? (Wu et al., 2013)	B
DLQI	+ (Wang et al., 2004)	+ (Finlay & Khan, 1994)	+ (Finlay & Khan, 1994)	+ (Finlay & Khan, 1994)	– (Finlay & Khan, 1994; Liu et al., 2016)	+ (Shikiar et al., 2005; Wang et al., 2004)	C
IDQOL	+ (Neri et al., 2012)	+ (Lewis-Jones et al., 2001; van Valburg et al., 2011; Neri et al., 2012)		+ (Gabes & Apfelbacher, 2021)	+ (Lewis-Jones et al., 2001; Neri et al., 2012; Wu et al., 2013)		A

Table 4 *(continued)*

Table 4 (*continued*)

Instrument	Internal consistency	Reliability	Measurement error	Content validity	Construct validity	Responsiveness	Recommendation*
CDLQI	+ (Ramírez-Anaya et al., 2010; Neri et al., 2012; Salek et al., 2013)	+ (Ramírez-Anaya et al., 2010; Salek et al., 2013)		+ (Lewis-Jones & Finlay, 1995; Gabes & Apfelbacher, 2021)	+ (Lewis-Jones & Finlay, 1995; Ramírez-Anaya et al., 2010)		A
PIQ-MS	+ (Lei et al., 2020a)	+ (Lei et al., 2020)		+ (Silverberg et al., 2020b)	+ (Lei et al., 2020)	+ (Lei et al., 2020)	A
5-D itch	? (Lin et al., 2019)	+ (Elman et al., 2010)		+ (Elman et al., 2010)	+ (Elman et al., 2010)		B
ItchyQoL	– (Patel et al., 2019)			+ (Patel et al., 2019)	+ (Patel et al., 2019)	+ (Patel et al., 2019)	C
SF-12	+ (Huo et al., 2018)	+ (Cheak-Zamora et al., 2009)		+ (Cheak-Zamora et al., 2009)	+ (Cheak-Zamora et al., 2009)	– (Cheak-Zamora et al., 2009)	C
DFI	+ (Al Robaee, 2010)	+ (Lewis-Jones et al., 2001)	– (Lewis-Jones et al., 2001)	+ (Lewis-Jones et al., 2001)	+ (Lewis-Jones et al., 2001)	+ (Jiráková et al., 2012)	C
ABS	+ (Méni et al., 2013)		– (Méni et al., 2013)	+ (Méni et al., 2013)	+ (Méni et al., 2013)		C
ABS-A	+ (Taieb et al., 2015)		– (Taieb et al., 2015)	+ (Gabes & Apfelbacher, 2021)	? (Gabes & Apfelbacher, 2021)		C
ADerm-IS				+ (Foley et al., 2019)			B
ESS	+ (Manzar et al., 2019; Lei et al., 2020b)	? (Lei et al., 2020b)	+ (Crook et al., 2019)	+ (Johns et al., 1991)	+ (Johns et al., 1991)	– (Sanford et al., 2006)	C
PROMIS SD	+ (Lei et al., 2020a)			+ (Li et al., 2018)	+ (Li et al., 2018; Lei et al., 2020a)	+ (Lei et al., 2020a)	A
PROMIS SRI	+ (Lei et al., 2020a)			+ (Li et al., 2018)	+ (Li et al., 2018; Lei et al., 2020a)	+ (Lei et al., 2020a)	A
ADCT	+ (Simpson et al., 2019)	+ (Simpson et al., 2019)		+ (Pariser et al., 2020)	+ (Simpson et al., 2019)	+ (Simpson et al., 2019)	A

*, category of recommendation: A, recommended; B, will recommended with more evidence; C, not recommended. +, sufficient; –, insufficient; ?, indeterminate. NA, not applicable.

imaginable", how would you rate your itch at the worst moment during the previous 24 hours?" (31,32,35). The single-item Skin Pain NRS assesses self-reported severity of worst skin pain each day. The patients are asked to select a number from 0 ("no pain") to 10 ("worst pain imaginable") that best described the worst level of skin pain in the past 24 hours (32,35). The current versions of Peak Pruritus/ Itch NRS and Skin Pain NRS are easy for moderate-to-severe AD patients to understand and answer compared to earlier versions of assessing average itch (rather than worst itch) or other types of pain (rather than skin pain), or other point scale (rather than 11-point scale), or longer recall (rather than the past 24 hours). Similarly, Sleep Disturbance (SD) NRS asks the AD patients: On a scale of 0-10, with 0 being "no sleep loss related to the symptoms of AD" and 10 being "I did not sleep at all due to the symptoms of AD", how would you rate your sleep last night (32,41)? All the three NRS are validated in moderate-to-severe AD patients and can be completed on an electronic device once daily throughout the clinical trial. The most appropriate thresholds for a minimally important changes for Peak Pruritus/Itch NRS, Skin Pain NRS and SD NRS are also determined, which are 2–4 points, 4 points and 2–5 points, respectively.

The Atopic Dermatitis Control Tool (ADCT©) is a newly developed 6-item PRO tool for disease severity with acceptable reality and validity properties (57). It enables patients to assess their long-term AD control situation, including AD symptom assessment, and QOL and bodily function effects of AD (e.g., itch, sleep, daily activities and emotions). ADCT is the first tool for assessing itch, sleep and impact on daily activity due to the AD in a single instrument. In a real-word study that focused on the effectiveness of dupilumab in patients with AD, the reality, validity and responsiveness of ADCT was assessed. In terms of validity, ADCT showed significant convergent validity with the DLQI and pain NRS. The ability of the ADCT to detect change was also affirmed. In terms of reality, Cronbach's α ranged from 0.9 to 0.95, and the item-total correlations ranged from 0.68 to 0.81, which are acceptable (56).

POEM is a valid and internal consistent patient-oriented instrument designed explicitly to measure the disease severity of AD, published in 2004 (34). Based on the frequency of occurrence during the preceding week, 7 items (dryness, itching, flaking, cracking, sleep loss, bleeding, and weeping) are assessed using a 5-point scale. POEM has adequate responsiveness and has been translated and validated in multiple languages. It can be completed by most patients in less than 2 minutes while also providing a more comprehensive view of disease severity than measurement of itch and/or sleep disturbance alone (34). However, the reliability and measurement error for POEM still require further investigation. POEM is commonly used in clinical trials and everyday practice as a subjective measurement to describe the disease outcome. Often together with physician-assessed objective outcome measurements such as EASI, POEM has been used by thousands of patients among all age groups in clinical trials of drug development for AD (79,80,81).

PO-SCORAD is a patient-oriented instrument derived from SCORAD, published in 2010 (52). It assesses 3 components of AD: the affected body surface area (BSA), the severity of clinical signs, and other clinical symptoms. The affected BSA is calculated as a percentage of each defined body area and reported as the sum of all areas. It is validated and internally consistent. One advantage of PO-SCORAD is that it provides visual explanations that are understandable by patients regardless of their age, which therefore improves the accuracy of the instrument (45). Also, PO-SCORAD measures both subjectively and objectively, which may minimize the bias caused by any misunderstanding between patients and physicians (45). The disadvantages of PO-SCORAD include lack of evidence supporting its reliability, and assessment of the measurement error and responsiveness. PO-SCORAD is also widely used in clinical trials for drug development and therapy evaluation in various populations including Chinese (82-84).

IDQOL is an instrument designed to assess the QOL in children (<4 years old) with AD from the parental view, and was published in 2001 (39). It has two parts: dermatitis severity and life quality index. It is been proved to have adequate internal consistency. IDQOL can be easily used by parents with or without another assessment of clinical severity. Although adequate, more research needs to be done to further evaluate the content validity and responsiveness to increase confidence in using IDQOL in clinical settings. IDQOL and CDLQI are often used together to evaluate the effects of treatment in children in clinical trials (85,86). These instruments were also used in several studies investigating the effect of AD on QOL of the children and their caregivers in various languages and populations (87-89).

CDLQI is an instrument designed to measure the effect of skin disease on children's QOL, published in 1995 (38). Unlike the abovementioned instruments, CDLQI is a generic instrument for skin and connective tissue diseases.

There is also a DLQI version for adults and a family version (FDLQI). Similar to those instruments, the CDLQI is a 10-item questionnaire that assesses 6 different aspects (symptoms and feelings, leisure, school or holidays, personal relationships, sleep, treatment) that may affect a child's QOL. It has cartoon illustrations based on the theme of every question, aiming to be more user-friendly for younger children. CDLQI has 131 translated versions and has been verified for reliability, interpretability and cross-cultural validity multiple times, so it can be easily adapted to different culture groups using different languages. However, as a generic instrument, the wording of the questions in CDLQI may lack precision for the effects of AD.

Discussion

This systematic review identified, summarized, and assessed the measurement properties of 26 different instruments used in the literature since 2010 to assess the disease severity and QOL of AD. Three single-item NRS (Peak Pruritus/Itch, Skin Pain and SD) demonstrate good reliability, validity and responsiveness and can measure day-to-day fluctuations in a specific aspect of disease severity related to AD (31,32,40). With determined minimal important changes, these single-item NRS are easy-to-interpret in clinical trials and are comparative among studies. However, AD is a characteristic of a variety of symptoms. Single-item NRS is an important complement to multidimensional scales as an outcome in clinical studies. Among multidimensional scales, only 4 instruments, namely POEM, PO-SCORAD, IDQOL and CDLQI, had adequate content and construct validity, internal consistency and reliability, but unclear measurement error, responsiveness, interpretability and feasibility. These are recommended to be used in clinical settings according to the COSMIN guidelines. Nine of the assessed instruments reported insufficiency in at least 1 measurement property and therefore are not recommended until future studies are available.

Nowadays, the diagnosis of AD is mostly based on clinical criteria, namely the historical features, morphology and distribution of skin lesions, and associated clinical signs (90,91). One of the earliest and most recognized sets of diagnostic criteria, the Rudzki criteria, has been used and validated in clinical practice for more than 40 years (92). With systematic modifications intended to provide a tool for researchers who are not dermatologists, the UK Working Party diagnostics were developed and are widely used and assessed. Both of these diagnostic schemes have

been validated in a wide range of age groups, languages and populations (93). However, neither disease severity nor QOL measurement scales are commonly used, or even recommended, for routine clinical practice, as suggested by Eichenfield *et al.* in 2014 (94).

However, in current clinical trials and drug development procedures, disease severity and QOL measurement scales are getting increasing attention and are often used as endpoint measures (90). Some disease severity scales, such as EASI, have been used extensively as the primary or secondary endpoint for large clinical drug trials (e.g., dupilumab, nemolizumab and tezepelumab) (18,95). In those studies, most of the patients had a 50% reduction in the EASI score after 12 weeks of treatment (18,95). Other endpoints used in clinical trials include the SCORAD and 5-D itch scales (56,96).

Recognizing the lack of generality, uniformity and accessibility of AD measurement scales, the development, validation and standardization of the outcome methodology requires father attention for a golden standard to be established. Few studies have covered different AD patient populations, especially with respect to cross-cultural equivalence, age groups and sex. In terms of content uniformity, a systematic review conducted by Schmitt *et al.* in 2007 (97) demonstrated substantial heterogeneity in the domains included in the different outcomes, the items used to measure the domains, the relative weights of the domains in the summary score, the scales used to measure the items, and the person performing the assessment. This phenomenon could lead to unfeasible comparison between clinical trials, and therefore misunderstanding and biased results when used in clinical settings. The accessibility and acceptance of AD measurement scales, especially for young patients, also requires attention. Although many measures were originally developed to be plain text-based for paper-and-pencil administration, some of them (e.g., CDLQI) do have illustrations to aid in understanding and could be easily adapted to electronic format. Moreover, electronic adaptation of existing measure scales may bring less administrative burden, higher patient acceptance rate, less secondary data entry errors, and more accurate and complete data (98). Migration of existing validated AD measurement scales to electronic platforms may help to improve the quality and accessibility of future clinical investigations, considering that some AD scales may involve assessment of affected BSA, severity of dryness, redness, swelling, crusting and oozing. It can be difficult for the patients to understand text descriptions, so if the electronic

version of these instruments includes illustrations, it will help patients to determine the progression of their disease more accurately and help researchers make judgments about the effectiveness of the drugs more precisely.

This study has several strengths and limitations. We used a highly sensitive search strategy for study collection instead of regular search strings. We also took advantage of 6 databases in English and Chinese, minimizing relevant information from being missed. Furthermore, we applied the validated COSMIN checklist methodology to rate the studies' quality and classified them in a systematic way based on predefined criteria.

In conclusion, as the treatment of AD is developing rapidly, the requirement for instruments to measure both disease severity and QOL for patients with AD is urgent. Therefore, in this review, we systematically analyzed 23 instruments utilized in the clinical trials from 2010 to 2021 (8 for disease severity, 15 for QOL). Of them, 9 instruments with significant reality and validity properties were recommended for further application. We also searched for the Chinese adapted versions of these 23 instruments and for instruments originally developed in Chinese that satisfied the inclusion criteria. However, there were too few Chinese AD instruments and their validation was inadequate. Therefore, further studies are needed to develop more original Chinese instruments for patients with AD.

Acknowledgments

Footnote

Reporting Checklist: The authors have completed the PRISMA reporting checklist. Available at https://atm.amegroups.com/article/view/10.21037/atm-22-3164/rc

Conflicts of Interest: All authors have completed the ICMJE uniform disclosure form (available at https://atm.amegroups.com/article/view/10.21037/atm-22-3164/coif). AL, MZ, JZ, and XX are from Sanofi (China) Investment. YY and XL are from Shanghai Jusure Health Technology. The other author has no conflicts of interest to declare.

Ethical Statement: The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

References

1. Carroll CL, Balkrishnan R, Feldman SR, et al. The burden of atopic dermatitis: impact on the patient, family, and society. Pediatr Dermatol 2005;22:192-9.

2. Lewis-Jones S. Quality of life and childhood atopic dermatitis: the misery of living with childhood eczema. Int J Clin Pract 2006;60:984-92.

3. Silverberg JI, Lei D, Yousaf M, et al. Measurement properties of the product of investigator's global assessment and body surface area in children and adults with atopic dermatitis. J Eur Acad Dermatol Venereol 2021;35:180-7.

4. Megna M, Napolitano M, Patruno C, et al. Systemic Treatment of Adult Atopic Dermatitis: A Review. Dermatol Ther (Heidelb) 2017;7:1-23.

5. Chamlin SL, Frieden IJ, Williams ML, et al. Effects of atopic dermatitis on young American children and their families. Pediatrics 2004;114:607-11.

6. Fabbrocini G, Napolitano M, Megna M, et al. Treatment of Atopic Dermatitis with Biologic Drugs. Dermatol Ther (Heidelb) 2018;8:527-38.

7. Schmitt J, Langan S, Deckert S, et al. Assessment of clinical signs of atopic dermatitis: a systematic review and recommendation. J Allergy Clin Immunol 2013;132:1337-47.

8. U.S. Department of Health and Human Services FDA Center for Drug Evaluation and Research; U.S. Department of Health and Human Services FDA Center for Biologics Evaluation and Research; U.S. Department of Health and Human Services FDA Center for Devices and Radiological Health. Guidance for industry: patient-reported outcome measures: use in medical product development to support labeling claims: draft guidance. Health Qual Life Outcomes 2006;4:79.

9. Food and Drug Administration. Draft guidance on pimecrolimus; 2012.

10. Simpson E, Bissonnette R, Eichenfield LF, et al. The

Validated Investigator Global Assessment for Atopic Dermatitis (vIGA-AD): The development and reliability testing of a novel clinical outcome measurement instrument for the severity of atopic dermatitis. J Am Acad Dermatol 2020;83:839-46.

11. Thyssen JP, Andersen YMF, Zhang H, et al. Incidence of pediatric atopic dermatitis following thymectomy: A Danish register study. Allergy 2018;73:1741-3.

12. Townshend AP, Chen CM, Williams HC. How prominent are patient-reported outcomes in clinical trials of dermatological treatments? Br J Dermatol 2008;159:1152-9.

13. Copley-Merriman C, Zelt S, Clark M, et al. Impact of Measuring Patient-Reported Outcomes in Dermatology Drug Development. Patient 2017;10:203-13.

14. 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.

15. LeBlanc TW, Abernethy AP. Patient-reported outcomes in cancer care - hearing the patient voice at greater volume. Nat Rev Clin Oncol 2017;14:763-72.

16. Food and Drug Administration. Patient-Focused Drug Development: Collecting Comprehensive and Representative Input; 2020.

17. Sawangjit R, Dilokthornsakul P, Lloyd-Lavery A, et al. Systemic treatments for eczema: a network meta-analysis. Cochrane Database Syst Rev 2020;9:CD013206.

18. Beck LA, Thaçi D, Hamilton JD, et al. Dupilumab treatment in adults with moderate-to-severe atopic dermatitis. N Engl J Med 2014;371:130-9.

19. Guttman-Yassky E, Thaçi D, Pangan AL, et al. Upadacitinib in adults with moderate to severe atopic dermatitis: 16-week results from a randomized, placebo-controlled trial. J Allergy Clin Immunol 2020;145:877-84.

20. Twiss J, Meads DM, Preston EP, et al. Can we rely on the Dermatology Life Quality Index as a measure of the impact of psoriasis or atopic dermatitis? J Invest Dermatol 2012;132:76-84.

21. Liu P, Zhao Y, Mu ZL, et al. Clinical Features of Adult/Adolescent Atopic Dermatitis and Chinese Criteria for Atopic Dermatitis. Chin Med J (Engl) 2016;129:757-62.

22. 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:e17520.

23. Hill MK, Kheirandish Pishkenari A, Braunberger TL, et al. Recent trends in disease severity and quality of life instruments for patients with atopic dermatitis: A

systematic review. J Am Acad Dermatol 2016;75:906-17.

24. Wang X, Li LF, Zhao DY, et al. Prevalence and Clinical Features of Atopic Dermatitis in China. Biomed Res Int 2016;2016:2568301.

25. Li LF, Liu G, Wang J. Prognosis of unclassified eczema: a follow-up study. Arch Dermatol 2008;144:160-4.

26. Lopez Carrera YI, Al Hammadi A, Huang YH, et al. Epidemiology, Diagnosis, and Treatment of Atopic Dermatitis in the Developing Countries of Asia, Africa, Latin America, and the Middle East: A Review. Dermatol Ther (Heidelb) 2019;9:685-705.

27. Chen HY, Lin YH, Wu JC, et al. Use of traditional Chinese medicine reduces exposure to corticosteroid among atopic dermatitis children: a 1-year follow-up cohort study. J Ethnopharmacol 2015;159:189-96.

28. Chen YC, Lin YH, Hu S, et al. Characteristics of traditional Chinese medicine users and prescription analysis for pediatric atopic dermatitis: a population-based study. BMC Complement Altern Med 2016;16:173.

29. Mokkink LB, Prinsen CA, Bouter LM, et al. The COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN) and how to select an outcome measurement instrument. Braz J Phys Ther 2016;20:105-13.

30. Moher D, Shamseer L, Clarke M, et al. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. Syst Rev 2015;4:1.

31. 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.

32. Silverberg JI, DeLozier A, Sun L, et al. Psychometric properties of the itch numeric rating scale, skin pain numeric rating scale, and atopic dermatitis sleep scale in adult patients with moderate-to-severe atopic dermatitis. Health Qual Life Outcomes 2021;19:247.

33. Severity scoring of atopic dermatitis: the SCORAD index. Consensus Report of the European Task Force on Atopic Dermatitis. Dermatology 1993;186:23-31.

34. Charman CR, Venn AJ, Williams HC. The patient-oriented eczema measure: development and initial validation of a new tool for measuring atopic eczema severity from the patients' perspective. Arch Dermatol 2004;140:1513-9.

35. Newton L, Randall JA, Hunter T, et al. A qualitative study exploring the health-related quality of life and symptomatic experiences of adults and adolescents with ulcerative colitis. J Patient Rep Outcomes 2019;3:66.

36. Wolkerstorfer A, de Waard van der Spek FB, Glazenburg EJ, et al. Scoring the severity of atopic dermatitis: three item severity score as a rough system for daily practice and as a pre-screening tool for studies. Acta Derm Venereol 1999;79:356-9.

37. Finlay AY, Khan GK. Dermatology Life Quality Index (DLQI)--a simple practical measure for routine clinical use. Clin Exp Dermatol 1994;19:210-6.

38. Lewis-Jones MS, Finlay AY. The Children's Dermatology Life Quality Index (CDLQI): initial validation and practical use. Br J Dermatol 1995;132:942-9.

39. Lewis-Jones MS, Finlay AY, Dykes PJ. The Infants' Dermatitis Quality of Life Index. Br J Dermatol 2001;144:104-10.

40. Puelles J, Fofana F, Rodriguez D, et al. Psychometric validation and responder definition of the sleep disturbance numerical rating scale in moderate-to-severe atopic dermatitis. Br J Dermatol 2022;186:285-94.

41. Dias-Barbosa C, Matos R, Vernon M, et al. Correction to: 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;4:107.

42. Gerbens LA, Prinsen CA, Chalmers JR, et al. Evaluation of the measurement properties of symptom measurement instruments for atopic eczema: a systematic review. Allergy 2017;72:146-63.

43. Coutanceau C, Stalder JF. Analysis of correlations between patient-oriented SCORAD (PO-SCORAD) and other assessment scores of atopic dermatitis severity and quality of life. Dermatology 2014;229:248-55.

44. Schram ME, Spuls PI, Leeflang MM, et al. EASI, (objective) SCORAD and POEM for atopic eczema: responsiveness and minimal clinically important difference. Allergy 2012;67:99-106.

45. 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.

46. Silverberg JI, Lei D, Yousaf M, et al. Comparison of Patient-Oriented Eczema Measure and Patient-Oriented Scoring Atopic Dermatitis vs Eczema Area and Severity Index and other measures of atopic dermatitis: A validation study. Ann Allergy Asthma Immunol 2020;125:78-83.

47. Lee JY, Kim M, Yang HK, et al. Reliability and validity of the Atopic Dermatitis Symptom Score (ADSS). Pediatr Allergy Immunol 2018;29:290-5.

48. Wu D, Huang C, Mo X, et al. Development and initial validation of a Traditional Chinese Medicine symptom-specific outcome measure: a Zheng-related atopic dermatitis symptom questionnaire (ZRADSQ). Health Qual Life Outcomes 2013;11:212.

49. Neri E, Agostini F, Gremigni P, et al. Italian validation of the Childhood Atopic Dermatitis Impact Scale: a contribution to its clinical application. J Invest Dermatol 2012;132:2534-43.

50. van Valburg RW, Willemsen MG, Dirven-Meijer PC, et al. Quality of life measurement and its relationship to disease severity in children with atopic dermatitis in general practice. Acta Derm Venereol 2011;91:147-51.

51. Gabes M, Apfelbacher C; quality of life working group of the Harmonising Outcome Measures for Eczema (HOME) initiative. IDQoL, CDLQI and the 45-item CADIS received a sufficient content validity rating during the HOME VII meeting in Japan: a group discussion study. J Eur Acad Dermatol Venereol 2021;35:458-63.

52. Ramírez-Anaya M, Macías ME, Velázquez-González E. Validation of a Mexican Spanish version of the Children's Dermatology Life Quality Index. Pediatr Dermatol 2010;27:143-7.

53. Salek MS, Jung S, Brincat-Ruffini LA, et al. Clinical experience and psychometric properties of the Children's Dermatology Life Quality Index (CDLQI), 1995-2012. Br J Dermatol 2013;169:734-59.

54. Lei DK, Yousaf M, Janmohamed SR, et al. Measurement Properties of 4 Patient-Reported Outcome Measures to Assess Sleep Disturbance in Adults With Atopic Dermatitis. Dermatitis 2020;31:321-7.

55. Li JC, Fishbein A, Singam V, et al. Sleep Disturbance and Sleep-Related Impairment in Adults With Atopic Dermatitis: A Cross-sectional Study. Dermatitis 2018;29:270-7.

56. Simpson E, Eckert L, Gadkari A, et al. Validation of the Atopic Dermatitis Control Tool (ADCT©) using a longitudinal survey of biologic-treated patients with atopic dermatitis. BMC Dermatol 2019;19:15.

57. Pariser DM, Simpson EL, Gadkari A, et al. Evaluating patient-perceived control of atopic dermatitis: design, validation, and scoring of the Atopic Dermatitis Control Tool (ADCT). Curr Med Res Opin 2020;36:367-76.

58. Charman CR, Venn AJ, Williams H. Measuring atopic eczema severity visually: which variables are most important to patients? Arch Dermatol 2005;141:1146-51; discussion 1151.

59. Martin SA, Brown TM, Fehnel S, et al. The atopic dermatitis itch scale: development of a new measure to assess pruritus in patients with atopic dermatitis. J

Dermatolog Treat 2020;31:484-90.

60. Foley C, Tundia N, Simpson E, et al. Development and content validity of new patient-reported outcome questionnaires to assess the signs and symptoms and impact of atopic dermatitis: the Atopic Dermatitis Symptom Scale (ADerm-SS) and the Atopic Dermatitis Impact Scale (ADerm-IS). Curr Med Res Opin 2019;35:1139-48.

61. Lin JN, Chiang DL, Chen TY. A Psychometric Evaluation of the Chinese Version of the 5D Itch Scale in Taiwanese Elderly. Int J Appl Sci Technol 2019;9:1-9.

62. 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.

63. Gånemo A, Svensson Å, Svedman C, et al. Usefulness of Rajka & Langeland Eczema Severity Score in Clinical Practice. Acta Derm Venereol 2016;96:521-4.

64. Rajka G, Langeland T. Grading of the severity of atopic dermatitis. Acta Derm Venereol Suppl (Stockh) 1989;144:13-4.

65. Wang XL, Zhao TE, Zhang XQ. Assessment on the reliability and validity of the Dermatology Life Quality Index in Chinese version. Zhonghua Liu Xing Bing Xue Za Zhi 2004;25:791-3.

66. Liu Y, Li T, An J, et al. Rasch analysis holds no brief for the use of the Dermatology Life Quality Index (DLQI) in Chinese neurodermatitis patients. Health Qual Life Outcomes 2016;14:17.

67. Shikiar R, Harding G, Leahy M, et al. Minimal important difference (MID) of the Dermatology Life Quality Index (DLQI): results from patients with chronic idiopathic urticaria. Health Qual Life Outcomes 2005;3:36.

68. Patel KR, Singam V, Vakharia PP, et al. Measurement properties of three assessments of burden used in atopic dermatitis in adults. Br J Dermatol 2019;180:1083-9.

69. Huo T, Guo Y, Shenkman E, et al. Assessing the reliability of the short form 12 (SF-12) health survey in adults with mental health conditions: a report from the wellness incentive and navigation (WIN) study. Health Qual Life Outcomes 2018;16:34.

70. Cheak-Zamora NC, Wyrwich KW, McBride TD. Reliability and validity of the SF-12v2 in the medical expenditure panel survey. Qual Life Res 2009;18:727-35.

71. Al Robaee AA. Reliability and validity of the Arabic version of "dermatitis family impact" questionnaire in children with atopic dermatitis. Int J Dermatol 2010;49:1063-7.

72. Jiráková A, Vojáčková N, Göpfertová D, et al. A comparative study of the impairment of quality of life in Czech children with atopic dermatitis of different age groups and their families. Int J Dermatol 2012;51:688-92.

73. Méni C, Bodemer C, Toulon A, et al. Atopic dermatitis burden scale: creation of a specific burden questionnaire for families. J Eur Acad Dermatol Venereol 2013;27:1426-32.

74. Taïeb A, Boralevi F, Seneschal J, et al. Atopic Dermatitis Burden Scale for Adults: Development and Validation of a New Assessment Tool. Acta Derm Venereol 2015;95:700-5.

75. Manzar MD, Salahuddin M, Alamri M, et al. Psychometric properties of the Epworth sleepiness scale in Ethiopian university students. Health Qual Life Outcomes 2019;17:30.

76. Crook S, Sievi NA, Bloch KE, et al. Minimum important difference of the Epworth Sleepiness Scale in obstructive sleep apnoea: estimation from three randomised controlled trials. Thorax 2019;74:390-6.

77. Johns MW. A new method for measuring daytime sleepiness: the Epworth sleepiness scale. Sleep 1991;14:540-5.

78. Sanford SD, Lichstein KL, Durrence HH, et al. The influence of age, gender, ethnicity, and insomnia on Epworth sleepiness scores: a normative US population. Sleep Med 2006;7:319-26.

79. Schlessinger J, Shepard JS, Gower R, et al. Safety, Effectiveness, and Pharmacokinetics of Crisaborole in Infants Aged 3 to <24 Months with Mild-to-Moderate Atopic Dermatitis: A Phase IV Open-Label Study (CrisADe CARE 1). Am J Clin Dermatol 2020;21:275-84.

80. Blauvelt A, de Bruin-Weller M, Gooderham M, et al. Long-term management of moderate-to-severe atopic dermatitis with dupilumab and concomitant topical corticosteroids (LIBERTY AD CHRONOS): a 1-year, randomised, double-blinded, placebo-controlled, phase 3 trial. Lancet 2017;389:2287-303.

81. Paller AS, Stein Gold L, Soung J, et al. Efficacy and patient-reported outcomes from a phase 2b, randomized clinical trial of tapinarof cream for the treatment of adolescents and adults with atopic dermatitis. J Am Acad Dermatol 2021;84:632-8.

82. Tan HY, Zhang AL, Xue CC, et al. Evaluation of the efficacy and safety of a Chinese herbal formula (RCM-106) for atopic dermatitis: study protocol for a randomised, double-blind, placebo-controlled trial in children. BMJ Open 2013;3:e003906.

83. Mengeaud V, Phulpin C, Bacquey A, et al. An innovative oat-based sterile emollient cream in the maintenance therapy of childhood atopic dermatitis. Pediatr Dermatol 2015;32:208-15.

84. Väkevä L, Niemelä S, Lauha M, et al. Narrowband ultraviolet B phototherapy improves quality of life of psoriasis and atopic dermatitis patients up to 3 months: Results from an observational multicenter study. Photodermatol Photoimmunol Photomed 2019;35:332-8.

85. Msika P, De Belilovsky C, Piccardi N, et al. New emollient with topical corticosteroid-sparing effect in treatment of childhood atopic dermatitis: SCORAD and quality of life improvement. Pediatr Dermatol 2008;25:606-12.

86. Kubota Y, Yoneda K, Nakai K, et al. Effect of sequential applications of topical tacrolimus and topical corticosteroids in the treatment of pediatric atopic dermatitis: an open-label pilot study. J Am Acad Dermatol 2009;60:212-7.

87. Alzolibani AA. Impact of atopic dermatitis on the quality of life of Saudi children. Saudi Med J 2014;35:391-6.

88. Xu X, van Galen LS, Koh MJA, et al. Factors influencing quality of life in children with atopic dermatitis and their caregivers: a cross-sectional study. Sci Rep 2019;9:15990.

89. Maksimovic N, Zaric M, Reljic V, et al. Factors associated with improvement of quality of life among parents of children with atopic dermatitis: 1-year prospective cohort study. J Eur Acad Dermatol Venereol 2020;34:325-32.

90. Eichenfield LF, Tom WL, Chamlin SL, et al. Guidelines of care for the management of atopic dermatitis: section 1. Diagnosis and assessment of atopic dermatitis. J Am Acad Dermatol 2014;70:338-51.

91. 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.

92. Rudzki E. Significance of skin dryness in atopic dermatitis. Pol Tyg Lek 1994;49:352-3.

93. Gu H, Chen XS, Chen K, et al. Evaluation of diagnostic criteria for atopic dermatitis: validity of the criteria of Williams et al. in a hospital-based setting. Br J Dermatol 2001;145:428-33.

94. Eichenfield LF, Totri C. Optimizing outcomes for paediatric atopic dermatitis. Br J Dermatol 2014;170 Suppl 1:31-7.

95. Ruzicka T, Hanifin JM, Furue M, et al. Anti-Interleukin-31 Receptor A Antibody for Atopic Dermatitis. N Engl J Med 2017;376:826-35.

96. Oldhoff JM, Darsow U, Werfel T, et al. Anti-IL-5 recombinant humanized monoclonal antibody (mepolizumab) for the treatment of atopic dermatitis. Allergy 2005;60:693-6.

97. Schmitt J, Langan S, Williams HC, et al. What are the best outcome measurements for atopic eczema? A systematic review. J Allergy Clin Immunol 2007;120:1389-98.

98. Coons SJ, Gwaltney CJ, Hays RD, et al. Recommendations on evidence needed to support measurement equivalence between electronic and paper-based patient-reported outcome (PRO) measures: ISPOR ePRO Good Research Practices Task Force report. Value Health 2009;12:419-29.

(English Language Editor: K. Brown)