



Assessing the severity of atopic dermatitis in clinical trials and practice

Rishi Chopra, MS^a, Jonathan I. Silverberg, MD, PhD, MPH^{b,c,*}

^a*Department of Dermatology, Northwestern University Feinberg School of Medicine, Chicago, Illinois, USA*

^b*Department of Dermatology, Preventive Medicine and Medical Social Sciences, Northwestern University Feinberg School of Medicine, Chicago, Illinois, USA*

^c*Northwestern Medicine Multidisciplinary Eczema Center, Chicago, Illinois, USA*

Abstract There is a tremendous need for accurate and reproducible scoring systems for the grading of skin disease to further the development of research and standards of care. There are presently greater than 60 measures that have been used to assess the severity of atopic dermatitis. These assessments vary considerably with respect to content, scale, instructions, validity, and concordance. This contribution reviews the available scoring systems of atopic dermatitis signs based on their design and merit in specific settings. These scores assess lesional intensity and/or extent, symptoms, disease course, and epidermal function. Scoring atopic dermatitis, Investigator Global Assessment, and Eczema Area and Severity Index are the most commonly used assessments of atopic dermatitis signs. Eczema Area and Severity Index has emerged as the preferred outcome measure of atopic dermatitis signs for use in clinical trials. Unfortunately, Eczema Area and Severity Index can be cumbersome in clinical practice. Itch intensity (visual analog or numeric rating scales) and Patient-Oriented Eczema Measure have emerged as the preferred patient-reported outcome in clinical trials. Clinicians' gestalt global assessment of severity, Patient-Oriented Eczema Measure, and intensity of itch may be feasible for clinical practice.

© 2018 Elsevier Inc. All rights reserved.

Introduction

Atopic dermatitis (AD) is a chronic, inflammatory skin disease with heterogeneous manifestations and intense itch. There is a wide spectrum of lesional morphology, including acute oozing and crusting, subacute lesions with dryness and scaling, chronic lesions with lichenification and/or prurigo nodules, and erythema, excoriations, and dryness occurring at all stages of disease. AD lesions may vary from being mild and barely perceptible to severe and profoundly inflamed. AD

lesions may be limited to the flexural areas or cover the entire body. Severity and intrusiveness of itch and localization of lesions to certain anatomic distributions may have a particularly negative impact on quality of life (QOL). All of these aspects are important considerations for assessing the severity of AD in both clinical trials and practice.

Presently, there exists no gold standard for evaluating the severity of AD.¹ There are also no reliable laboratory tests or biomarkers to assess the severity of AD. As a result, clinicians must rely upon clinical assessments of disease parameters, which can be subjective and difficult to standardize.² Standardized measures have been developed to quantify the clinical disease burden at baseline and the effectiveness of a

* Corresponding author.

E-mail address: JonathanISilverberg@gmail.com (J.I. Silverberg).

treatment regimen.^{2, 3} The first objective scoring systems, Rajka & Langeland (R-L) and the Simple Scoring System, were developed in 1989.⁴ Since then, many additional instruments were developed and there are now greater than 60 different named and unnamed disease severity measures in the published literature⁴; however, the validity, reliability, sensitivity to change, and acceptability have not been adequately elucidated in many of these studies.¹ A 2013 systematic review found that only 16 such instruments have undergone validation in at least one study.⁵ The use of inadequately validated instruments and heterogeneity of instruments may impede comparison between different clinical trials using different instruments.

Given the increasing number of outcome measures developed for AD, the Harmonization of Outcome Measures in Eczema (HOME) international consensus group was assembled to promote the standardization of use of AD outcome measures in clinical trials internationally. HOME performed multiple systematic reviews to examine the validity of currently available measures. Rigorous validation of a severity measure requires the assessment of an array of measurement properties: (1) truth (ie, content validity and construct validity), (2) discrimination (ie, internal consistency, interobserver reliability, intraobserver reliability, responsiveness/sensitivity to change, floor and ceiling effects, and interpretability), and (3) feasibility (ie, acceptability, ease of use, and time to perform and interpret).^{6–8} Despite considerable progress in harmonizing outcome measures, the landscape of severity measures remains muddled.

Given the recent explosion of research and development of novel treatments for AD, it is imperative for clinicians to understand the various severity assessments used in past and future clinical trials. As more evidence emerges about the validity, strengths and weaknesses of different measures, some measures may even emerge as being feasible and relevant for clinical practice. This review describes, compares, and contrasts the most commonly used AD severity assessments systems and highlights their role in both clinical trials and in clinical practice.

Recent trends of use of AD outcome measures

A systematic review of randomized controlled trials (RCT) performed in 2016 demonstrated that the most commonly used severity measures were SCORAD, Visual Analogue Scale (VAS)-pruritus, IGA, and EASI (Table 1).⁴ Three of the top four instruments—SCORAD, IGA, and EASI— assess signs, with SCORAD also assessing symptoms of pruritus and sleep loss.⁹ The signs of AD most frequently evaluated across all outcome measures were erythema, lichenification, edema (induration/papulation), exudation (oozing/crusting/weeping), excoriation, and dryness.⁵

A systematic review investigated 20 named outcome measures for AD and found that they fit into five broad domains: clinical signs, extent of disease, symptoms, disease course,

and epidermal function¹; however, when an international Delphi panel comprised of consumers, experts, regulatory agencies, and journal editors were asked to recommend the most important outcome domains for use in clinical trials, only one of the 13 domains selected was unanimously chosen to be included in the core outcome set for AD studies, ie, “clinical signs” (objective outcomes that are assessed by clinician or investigator).¹⁰

Scoring AD

SCORAD was inaugurated in 1993 by the European Task Force on AD, and it is the most commonly used instrument in RCT for AD.^{4, 11} SCORAD incorporates clinician assessment of lesional severity and extent, as well as patient-reported symptoms of pruritus and sleep loss.^{4, 12} Six signs of AD are assessed, including erythema, edema/papulation, oozing/crusting, excoriations, xerosis, and lichenification. Each sign is scored as: 0 (not present), 1 (barely perceptible), 2 (clearly perceptible), or 3 (very prominent). A representative lesion is selected for scoring, ie, neither the best nor the worst; xerosis is only scored for nonlesional skin. Lesional extent is assessed via an estimation of body surface area (BSA), using the “rule of nines” or palmar method and ranges from 0% to 100%. SCORAD relies on the assumption that extent and severity are linearly correlated, but this relationship only exists in persons with less than or equal to 30% BSA.¹³ Symptoms of itch and sleep loss over the past 3 days are assessed on scales of 0 to 10 using 10-cm VAS. Total SCORAD is calculated as follows: $BSA/5 + 7(\text{sum of intensity scores})/2 + \text{subjective symptom scores}$ and range from 0 to 103.¹¹ That is, lesional severity comprises 61.2%; extent and symptoms each comprise only 19.4% of the total score.

Most trials (92%) have not reported the individual sign and symptom components of SCORAD.¹⁴ Ideally, when SCORAD is used, the sign and symptom components should also be reported to distinguish between the objective and subjective components of AD severity.⁵ Twenty-six validation studies have been performed on SCORAD, and they have found good convergent and divergent construct validity, internal consistency, responsiveness, adequate interobserver reliability, and no floor or ceiling effects.^{4, 5, 9} Unfortunately, there is interobserver variability in evaluating lichenification and disease extent.¹²

Objective SCORAD

Objective SCORAD (oSCORAD) includes the assessments for lesional severity and extent but not symptoms. oSCORAD ranges from 0 to 83, unless disfiguring lesions on the hands or face or functionally limiting lesions are present. In this scenario, 10 bonus points are added and oSCORAD ranges from 0 to 93.¹⁵ oSCORAD is internally

Table 1 Description of severity assessments, extent of validation, and frequency of use in trials.^{1, 4, 5, 22, 37}

Severity assessment	Year of validation	Signs/symptoms	Content		Intensity/extent	How extent is measured	Validation studies ^a	Frequency of use in trials ⁴
			Signs	Symptoms				
SCORAD	1993	Both	erythema, edema/induration/population, excoriation, oozing/crusting/weeping/exudation, lichenification, dryness, bleeding	Pruritus, sleep loss	Both	Rule of 9 s	26 ^{b5}	87
oSCORAD	1993	Signs	erythema, edema/induration/population, excoriation, oozing/crusting/weeping/exudation, lichenification, dryness, bleeding	None	Both	Rule of 9 s	26 ^{b5}	8
EASI	1998	Signs	erythema, edema/induration/population, excoriation, lichenification	None	Both	4 body sites scale	9 ⁵	28
Global assessments (IGA, PGA, etc.)	Unknown	Signs ^c	Erythema, population/edema, oozing/weeping, crusting, excoriation, scaling, lichenification (combinations of signs varied from study to study) ²²	None	Intensity	N/A	None	31
PO-SCORAD	Not listed	Both	erythema, edema/induration/population, excoriation, oozing/crusting/weeping/exudation, lichenification, dryness, scaling, cracking/fissuring	Pruritus, sleep loss	Both	Shading of topographic silhouette	2 ⁵	1
SA-EASI	2002	Both	VAS scoring of erythema, edema/induration/population, excoriation, dryness,	Sleep loss	Both	4 body sites scale	4 ⁵	0
SASSAD	1996	Signs	Erythema, oozing/crusting/weeping/exudation, excoriation, lichenification, drying, cracking/fissuring,	None	Both	6 body sites	5 ⁵	2
POEM	2004	Both	oozing/crusting/weeping/exudation, dryness, cracking/fissuring, flaking, bleeding,	Pruritus, sleep loss	Frequency of signs*	N/A	2 ⁵	5
BCSS	1995	Both	Presence/absence of eczema	Pruritus	Both	5 body sites	1 ⁵	0
ADAM	1999		erythema, excoriation, lichenification, scaling,		Both	6 body areas	2 ⁵	0
ADASI	1991	Both	erythema, edema/induration/population, oozing/crusting/weeping/exudation, lichenification, scaling	Pruritus	Both	Color coding	1 ⁵	1
ADQ	2008	Both	erythema, oozing/crusting/weeping/exudation, dryness, scaling, cracking/fissuring,	Pruritus	Both	7 body parts	1 ⁵	0
OSAAD	2003	Signs	Erythema, oozing/crusting/weeping/exudation, excoriation, lichenification	None	Both	Computer-assisted body surface area	3 ⁵	0
SSS	1989	Both	erythema, edema/induration/population, oozing/crusting/weeping/exudation, excoriation, lichenification, scaling, vesicles, depigmentation,	Pruritus, sleep loss	Both	10 topographic sites	2 ⁵	0
TIS	1999	Signs	Erythema, edema/induration/population, excoriation,	None	Intensity	N/A	7 ⁵	2
W-AZS	2005	Both	erythema, edema/induration/population, oozing/crusting/weeping/exudation, scaling, vesicles, depigmentation	Pruritus, sleep loss	Intensity	N/A	0 ⁵	0
ADSI	1998	Both	Erythema, exudation, pruritus, excoriation and lichenification	Pruritus	Intensity	Target lesion	0 ¹	2
NESS	2000	Both		Sleep	Both	> 20 body sites	5 ¹	0

Table 1 (continued)

Severity assessment	Year of validation	Signs/symptoms	Content	Intensity/extent	How extent is measured	Validation studies ^a	Frequency of use in trials ⁴
			Signs	Symptoms			
R-L	1989	Both	Clinical course and clinical intensity; in each category five questions for the parents	Sleep loss	Extent	Rule of nines	3 ¹
SIS	1992	Both	Clinical course and clinical intensity, assessed by three questions	Pruritus	Both	Target lesion	0 ¹
TBSA	1994	Signs	Itching, erythema and lichenification	None	Signs	6 regions	0 ¹
BSA	N/A	Signs	Erythema, infiltration, vesicles/papules, dryness/scaling, cracking/fissuring and excoriation/crusting	N/A	Extent	Rule of 9 s	0
VAS-itch	N/A	Symptoms	N/A	Pruritus	Intensity	N/A	Not assessed
VAS sleep disturbance	N/A	Symptoms	N/A	Sleep loss	Intensity	N/A	Not assessed
4-point scale for itching	N/A	Symptoms	N/A	Pruritus	Intensity	N/A	Not assessed
Pruritus NRS	N/A	Symptoms	N/A	Pruritus	Intensity	N/A	Not assessed
Pruritus Categorical Response Scale	N/A	Symptoms	N/A	Pruritus	Intensity	N/A	Not assessed

ADAM, assessment measure for atopic dermatitis; ADQ, atopic dermatitis quickscore; ADSI, atopic dermatitis severity index; BCSS, basic clinical scoring system; BSA, body surface area; EASI, eczema area and severity index; IGA, investigator's global assessment; N/A, not assessed; NESS, Nottingham eczema severity score; OSAAD, objective severity assessment of atopic dermatitis; oSCORAD, objective SCORAD; PGA, physician's global assessment; POEM, patient-oriented eczema measure; PO-SCORAD, patient-oriented SCORAD; R-L, Rajka and Langeland; SA-EASI, self-assessed EASI; SASSAD, six-area, six-sign atopic dermatitis severity; SCORAD, SCORing atopic dermatitis; SIS, skin intensity score; SSS, simple scoring system; TBSA, six-area, total body severity assessment; TIS, three item severity score; W-ASZ, new scoring system for assessment of the extent and severity of skin inflammation index in atopic dermatitis patients.

^a Validation studies are according to a 2013 and 2007 systematic review by Schmitt et al⁵ and Schmitt et al,¹ respectively.

^b SCORAD and oSCORAD considered a single instrument with respect to the number of validation studies

^c Twenty-two out of 23 global assessments assessed signs. One assessed both signs and symptoms.

consistent with good intrarater reliability, but interobserver reliability is still unclear.^{5, 16}

Severity strata for oSCORAD were proposed based on author consensus (<15 = mild; 15-40 = moderate; >40 = severe) without any formal testing.¹⁵ These strata were extended to SCORAD by adding 10 points to the oSCORAD thresholds (<25 = mild; 25-50 = moderate; >50 = severe), again without any formal testing.¹¹ Recently, an interpretability study¹⁷ identified different severity bands for oSCORAD and SCORAD in adults with AD, using an anchor-based approach (oSCORAD: 0.0-7.9 = clear, 8.0-23.9 = mild, 24.0-37.9 = moderate and 38.0-83 = severe; SCORAD: 0.0-9.9 = clear, 10.0-28.9 = mild, 29.0-48.9 = moderate and 49.0-103.0 = severe).

Eczema Area and Severity Index

EASI was developed as a modification of the well-established Psoriasis Area and Severity Index (PASI)

using psoriasis.⁹ EASI assesses both lesional intensity and extent. Four signs—erythema, papulation/edema, excoriation, and lichenification—are assessed using a scale of 0 to 3 similar to SCORAD. Unlike SCORAD, EASI assesses the average lesional intensity within four body regions—head and neck, arms, trunk, and legs. Lesional extent is evaluated by estimating the surface area involved of those four body regions (1-9%, 10-29%, 30-49%, 50-69%, 70-89%, and 90-100%).¹⁸ Lesional intensity is multiplied by the surface area involved in that region and summed across regions, yielding a total score ranging from 0 to 72.⁴

Nine validation studies of EASI were performed and demonstrated validity, internal consistency, adequate intraobserver reliability, intermediate interobserver reliability, and adequate responsiveness.⁵ Due to moderate interobserver reliability, it is recommended that the same clinician perform both baseline and follow-up assessments¹⁶; however, there is a wider range of EASI scores for severe versus mild or moderate disease and there are floor and ceiling

effects. Changes in score correlate with larger clinical changes at low versus high EASI values. These differences have been suggested to be due to the absence of itch, impact on QOL, and involvement of high visibility areas in scoring.¹⁹

This contribution has observed that EASI is poorly responsive in the mild disease range due to the broad range of extent at the lowest range. That is, the presence of a 1-cm lesion affecting less than 1% and numerous lesions affecting 9% of a body region would be scored similarly. EASI may not be an ideal measure for studies assessing patients with mild AD owing to potentially insufficient responsiveness.

Severity strata were previously defined for EASI using an anchor-based approach (0.0 = clear, 0.1-1 = almost clear; 1.1-7.0 = mild; 7.1-21 = moderate; 21.1-50.0 = severe; and 50.1-72.0 = very severe).¹⁹ We recently performed an interpretability study and found similar, but slightly different severity strata in adults with AD (0.0 = clear, 1.0-5.9 = mild, 6.0-22.9 = moderate and 23.0-72.0 = severe).¹⁷

The modified EASI (mEASI) is a variant of EASI that is identical but includes an assessment of itch by the patient. Itch is included in scoring, because it is considered a primary symptom of AD.²⁰ Scores range from 0.0 to 90.0 and severity strata are defined as follows: 0.0-0.9 = clear, 1-8.9 = mild, 9-29.9 = moderate and 30.0-90.0 = severe.¹⁷ Unlike EASI, mEASI has never been validated in the literature.⁵

Global assessments

Global assessments are quick and simple “snapshots” of disease severity. There are several different approaches used for global assessment of AD. Assessments may be static (severity at a fixed point of time) or dynamic (comparison of current severity to baseline); however, static assessments are preferred, because they are considered more accurate and less subjective.²¹ There is extensive variability with respect to the use of global assessments. One approach is a gestalt assessment of overall disease severity by a clinician. In this scenario, no explicit scoring instructions are required to perform an evaluation, and classically, only signs are assessed. While this approach has been used as the standard by which other instruments are validated or anchored, it has not been formally validated. We expect this to perform quite well and show good intraobserver and interobserver reliability; nevertheless, this approach may be influenced by the clinical experience of the observer and the most severe cases previously encountered.

There are a number of global assessments that have been used with specific instructions on how to assess the intensity of a representative or target lesion. Although IGA and Physician Global Assessment are the two most common names for global assessments, there are 18 other names for these global scales; moreover, there are 23 unique scoring systems ranging from 4 to 7 points, which vary with respect to size, content, instructions, and analysis. The six-point scale is the most

common and is typically defined as follows: 0 = clear; 1 = almost clear; 2 = mild; 3 = moderate; 4 = severe; 5 = very severe. The most commonly assessed content includes: erythema, papulation/edema, and oozing/weeping. Less commonly, crusting, excoriation, scaling, and lichenification are assessed. Despite there being no studies assessing the validity or responsiveness of IGA,^{9,22-24} the United States Food and Drug Administration generally mandates that IGA be used for RCT in AD. Global assessments may have less interobserver reliability than other measures of AD signs and should be performed by the same clinician from visit to visit.²²

Six Area Six Sign of AD

The Six Area Six Sign of AD (SASSAD) severity score was inaugurated in 1996 and consists of the assessment of six signs (erythema, exudation, excoriation, xerosis, cracking, and lichenification) at six different areas (head and neck, buttocks/trunk, arms, legs, hands, and feet). Each sign is graded on a scale of 0 to 3 at each of the six areas, with a total range of scores from 0 to 108. Although the six areas are not similar in size, they are more deemed similar in importance to the patient. This scoring system simplifies assessments of extent by avoiding estimations of BSA. SASSAD has been used for the study of both adults²⁵⁻²⁷ and children.²⁸ Five validation studies have been performed for SASSAD and found adequate construct validity, floor or ceiling effects, acceptability, intermediate content validity and interobserver reliability. No studies have been performed on internal consistency, intraobserver reliability, interpretability, or sensitivity to change.⁵ No severity strata have been established for SASSAD.

Atopic Dermatitis Severity Index

The AD Severity Index (ADSI) was first described and used in 1998 in a study evaluating the effectiveness of topical ascomycin, a macrolactam.²⁹ ADSI grades pruritus, erythema, exudation, excoriation, and lichenification of a target lesion on a scale of 0 to 3, with total scores ranging from 0 to 15.²⁹ To date, there have been no studies validating its use; however, it was employed recently in a RCT to evaluate the safety and efficacy of crisaborole topical ointment. We recently performed an interpretability study and proposed severity strata (0.0-1.9 = clear, 2.0-5.9 = mild, 6.0-8.9 = moderate and 9.0-15.0 = severe), which may help clarify its use. In that study, ADSI demonstrated reasonable correlation with global AD severity; however, it did not correlate as well as measures including lesional extent, including SCORAD, oSCORAD, EASI, and mEASI.¹⁷

Rajka & Langeland

R-L was the one of the earliest developed AD severity instruments. Scoring is comprised of intensity (sleep disturbance due to itch), extent (“rule of nines”), and disease course (number of months the patient was affected by eczema in the previous year). The assessment of disease course is unique to R-L in comparison to other commonly used severity measures. Scores range from 3 to 9 and severity strata are as follows: mild = 3.0–4.0; moderate = 4.5–7.5; severe = 8.0–9.0. This instrument has adequate content validity, very good interobserver and intraobserver reliability, adequate divergent construct validity, sensitivity to change, and time consumption.³⁰

Patient-reported outcomes (PRO)

There is a profound symptom burden of AD, including pruritus, pain, sleep disturbance, and mental health symptoms. These aspects of disease may not be appreciated by objective assessments of AD signs. PRO and QOL measures are better at capturing the subjective burden of AD.^{9, 31, 32} In addition to communication of disease impact on QOL from patient to provider, these patient assessments allow for monitoring of flares and disease activity between clinical encounters; however, PRO are subjective and can be influenced by coping strategies, comorbidities, and impact on QOL.¹

According to a systematic review of the reporting of symptoms in AD trials since 2000,³³ 30 different composite instruments were used to evaluate both signs and symptoms of AD. The most frequently used instruments were SCORAD (49%), modified EASI (2.4%), ADSI (1.4%), and POEM (1.7%). Itch, sleep loss, and skin dryness/roughness were the most commonly assessed symptoms. The most commonly used scales to measure individual symptoms were VAS, Nonverbal Rating Scale (NRS), and Verbal Rating Scale (VRS).

Differences between assessments

Each outcome measure includes different domains, signs and symptoms, weighting of domains, and assessor.¹ This section will describe the fundamental differences between different outcome measures.

Lesional intensity versus extent

Clinicians and patients alike agree that both intensity and extent of lesions are “very important” for the assessment of AD severity¹; however, there is considerable variability in how these are measured and weighted, stemming from a fundamental debate over which of these matters more for AD severity. BSA assesses extent but not intensity. IGA assesses

intensity but not extent. EASI assesses both, but heavily weights scores by extent (~50%).^{2, 11} oSCORAD and SCORAD are mostly driven by intensity (~60%), with a smaller weighting for extent (~20%). Contributions of disease intensity vary between outcome measures from 33%³⁴ to 100%³⁵ and extent vary from 19%¹¹ to 100%.³⁶ There is a nonlinear relationship between lesional intensity and extent.¹³ We recently performed an interpretability study to determine severity strata for multiple measure of AD signs, and found that measures including extent (BSA, EASI, SCORAD) correlated better with overall disease severity than those with intensity but no extent (ADSI).¹⁷

EASI versus SCORAD

Erythema, edema/papulation, excoriation, and lichenification are assessed in both EASI and SCORAD; however, SCORAD also assesses oozing/crusting and xerosis of nonlesional skin.^{2, 11, 12} Previous studies found strong correlations between EASI and oSCORAD.^{5, 37} We recently showed a nonlinear relationship between EASI and oSCORAD.³⁸ In particular, EASI scores less than or equal to 5 were unable to distinguish between fairly broad ranges of oSCORAD, when moderate to severe xerosis was present or when moderate to severe lesions were localized to the face, eyelid, neck, flexural areas, hands, and feet. For EASI scores greater than 5, there was a linear relationship with oSCORAD. There is a complex relationship between EASI and oSCORAD and limited discriminative ability of low EASI scores compared with SCORAD.³⁸ Unlike SCORAD which only evaluates the intensity of itch, mEASI weights severity of itch according to the extent of affected BSA.²⁰

Limitations of global assessments

Comparisons between global assessments and other severity measures are prevalent in the literature, as global assessments often serve as a proxy gold standard for validation or anchoring.¹ Items noticeably absent from IGA evaluations, include disease extent, involvement of high visibility or functionally significant areas, or symptoms, all of which are important aspects of AD severity.^{13, 39} Global assessments often do not assess excoriations which most often correlate with itch and disease severity according to patients.²² IGA demonstrates reduced responsiveness and may be influenced by the response to previous treatments, patient compliance, and the doctor-patient relationship.^{1, 9} For all these reasons, IGA is not an ideal measure for AD.¹

Concordance of AD signs and symptoms

The concordance between instruments that assess signs versus symptoms varies considerably between studies. The

objective domains of SCORAD and R-L are well correlated to the subjective domains of the QOL measures Infant Dermatitis Quality of Life Index, Children's Dermatology Life Quality Index and Dermatology Life Quality Index (DLQI).⁴⁰ SCORAD and DLQI are not significantly correlated at baseline, but are moderately correlated with respect to response to treatment.⁴¹ There is a low correlation between patient-reported and physician-assessed disease severity as demonstrated by a lack of correlation of oSCORAD, EASI, SASSAD, and TIS with POEM, DLQI, and Skindex-29, with SASSAD and Skindex-29 being the only measures with moderate correlation.⁴² Together, these studies highlight the need for assessments of both AD signs and symptoms to completely capture the broad manifestations of AD.

Concordance of clinician and patient assessments

Patient-oriented SCORAD (PO-SCORAD) and self-administered EASI (SA-EASI) were developed as PRO analogs of SCORAD and EASI. PO-SCORAD utilizes an illustrated tutorial for patients to assess the severity and extent of their disease. SA-EASI allows the caregiver to assess extent via a body silhouette that estimates BSA and intensity of redness, thickness, dryness, number of scratches, and itchiness of the child's average AD lesion via five 100-mm VAS. Although these measures are analogous to their parent clinician assessments, they differ significantly in their review, content, and measurement properties. These differences preclude comparisons between instruments and prevent these objective and subjective scores from being used interchangeably.⁵

Measures used for clinical trial inclusion criteria

Although the HOME group is working toward standardized outcome measures in RCT, no such impetus exists to standardize inclusion criteria. Variability in inclusion criteria may result in the recruitment of patients not representative of the intended target population for severity, and thus, may limit the interpretability of trial results.

A recent systematic review of RCT of AD found that 26 different assessments and 35 unique combinations of two or more assessments were used to define AD severity for inclusion criteria. Global assessments (21.2%), BSA (21.2%), SCORAD (20.7%), and R-L (8.0%) were used most commonly. Additionally, there was considerable variability with respect to the definitions of the minimum and maximum thresholds for severity strata of individual instruments. This heterogeneity in trial methodology limits the ability to assess concordance between the targeted population for severity and the actual severity of recruited subjects. Additionally, PRO and symptoms were rarely used as inclusion criteria.

Basing recruitment solely on AD signs may not fully capture the severity of AD and its patient burden.

In AD trials that defined a target population based on severity, only 62.7% used the same severity instrument for inclusion and for baseline assessment. There may be discordant assessments of severity for inclusion and baseline using different measures. For example, a patient with a severe, localized lesion may be considered severe using IGA but only mild or moderate using SCORAD or EASI. The inconsistent use of severity measures in these three components of a trial—inclusion, baseline, and efficacy outcomes—may result in patients with baseline severity that do not match the target population, differences in populations between studies, and inaccurate interpretations of study outcomes. These measures should be consistently applied across studies to ensure capture of the desired targeted population and the accurate comparison of interventions in patients of a particular disease severity.⁴³

Severity assessments in clinical trial outcomes

A 2003 systematic review revealed that only 27% of AD RCT used a severity measure that had been published previously.⁴⁴ Over the last 15 years, there has been considerable standardization of instruments used and more consistent assessment of specific domains. Yet, significant heterogeneity remains. Before a set of instruments can be recommended for trials, the most important outcome domains for AD need to be defined.

The international HOME collaboration was founded in 2010, and serves as a multidisciplinary evidence-based initiative to standardize outcomes research in AD clinical trials.³⁹ The collaboration arrived at the consensus that symptoms, clinical signs assessed by the physician, the assessment of long-term control of flares, and health-related QOL should be included in the core outcome set.^{8, 10} At a minimum, these four core outcome domains should be assessed in all trials. Three signs—excoriations, erythema, and edema /papulation—were selected as core signs, because they are independent predictors of patient-assessed disease severity.^{23, 44} Additionally, lichenification was also added, because it assesses the chronicity of disease and is more prominent in patients of color.¹⁴ Only oSCORAD and EASI were recommended for assessing signs in clinical trials as they assess the four essential signs and are adequately validated.⁵ EASI was preferred over oSCORAD as a primary end point, because it lacks redundancy by only assessing the four fundamental signs, assesses the extent of signs at multiple sites rather than exclusively a representative lesion, and distributes greater weight to disease extent than oSCORAD.¹⁴ Severity bands have recently been defined for EASI, increasing its interpretability and providing a global assessment conversion.¹⁹

EASI demonstrated greater interrater and intrarater reliability than oSCORAD, IGA, or POEM in assessing erythema in patients of color, and thus, has broader applicability in clinical trials internationally.⁴⁵

POEM was recommended as the preferred assessment of AD symptoms. No QOL instruments are currently recommended for use; however, Quality of Life Index of Atopic Dermatitis has the potential to be recommended pending further validation studies. Interestingly, the most commonly used QOL measure, English DLQI, is not recommended due to unknown content validity, redundancy of items on the scale, absence of cultural validity studies, and unknown responsiveness.⁴⁶ As for selecting measures that assess control over flares, the group could not decide by consensus whether long-term stability should be captured by another instrument specific to flares or by routine measurement with the same scoring systems used to assess signs, symptoms, and QOL. Further research and discussion are required for long-term control measures.⁴⁷

Severity assessments for clinical practice

Ideally, the same outcome domains should be assessed in a patient visit to promote the effective translation of results from clinical trials to general practice. The objectives in both settings are different, and the same assessments may be inappropriate due to variations in core outcome domains. For an instrument to be useful to the general dermatologist, it should have demonstrated validity in clinical practice, require minimal training, be time-efficient, and seamlessly integrate into day-to-day practice. EASI requires extensive training to improve intrarater and interrater reliability, has no demonstrated validity or feasibility in the clinical setting, and is likely a poor measure in the clinical practice setting.

The most critical aspect to implementing an instrument in routine care is the time it takes to perform.⁴⁸ To be “acceptable” or “adequate” for use in clinical practice, a severity score should not take longer than 3 minutes or 3 to 5 minutes, respectively.¹ SCORAD and EASI took 3⁴⁸ to 10¹¹ minutes and 2⁴² to 6¹⁹ minutes to perform across different studies, respectively. Completion time of these assessments is user-dependent and varies based on experience and training. Most studies that specifically addressed completion time for an untrained dermatologist reported that it took longer than 3 minutes.^{11,15,49} SCORAD and EASI are most likely neither “acceptable” nor “adequate” for routine clinical practice.

A patient’s assessment of symptoms may be a more accurate reflection of patient-burden and clinically relevant.⁹ Assessment of itch and sleeplessness alone are insufficient to fully measure disease severity according to HOME¹⁴; however, due to the abundance of AD-related symptoms, it was not possible for HOME to develop a comprehensive short list of important symptoms that need to be assessed by an instrument.

Dryness; red, inflamed, irritated skin; and, to a lesser extent, pain and soreness; were all considered essential.⁴⁷

POEM is the only PRO sufficiently validated to be recommended by the HOME collaboration. POEM is a self-reported eczema-specific tool that assesses frequency of signs and symptoms important to patients, including itch, sleep loss, bleeding, weeping/oozing, skin cracking, flaking, and xerosis over the previous week. It is simple, patient-centered, easily interpreted, and takes most patients 1 to 2 minutes to complete. It was developed in the primary and secondary care settings, supporting its use in both community and hospital settings. Patients may complete questionnaires before seeing their dermatologist either at home or in the waiting room,⁵⁰ although POEM’s structural validity, cross-cultural validity, reliability, measurement error, and lack of assessment of intensity of symptoms raise major concerns about its use in clinical practice.⁴⁷ The most feasible measures in clinical practice are a gestalt clinician’s global assessment of mild, moderate or severe; however, the validity of this approach has yet to be established. NRS-itch is also a simple PRO measure that can be used to measure itch intensity, but may not characterize the full severity of AD.

Conclusions

There are many factors that contribute to the complexity of assessing AD severity, with a myriad of outcome measures, ranging from varying content, scales, and concordance to degree of validation. For this reason, severity measures for AD should be well defined and validated. The HOME group selected EASI as the preferred outcome measure for RCT in 2014; however, EASI is not feasible for clinical practice. A gestalt physician’s global assessment is feasible for clinical practice, but demonstration of validity is lacking. NRS-itch and POEM are feasible for clinical practice, but they may not fully capture the patient burden of AD. Future studies should be able to determine the validity and feasibility of different outcome measures in clinical practice.

References

1. 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;120:1389-1398.
2. Hanifin JM, Thurston M, Omoto M, et al. The eczema area and severity index (EASI): Assessment of reliability in atopic dermatitis. *EASI Evaluator Group. Exp Dermatol* 2001;10:11-18.
3. 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.
4. 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;75:906-917.
5. 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;132:1337-1347.

6. Mokkink LB, Terwee CB, Patrick DL, et al. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. *J Clin Epidemiol* 2010;63:737-745.
7. Boers M, Brooks P, Strand CV, et al. The OMERACT filter for outcome measures in rheumatology. *J Rheumatol* 1998;25:198-199.
8. Schmitt J, Spuls P, Boers M, et al. Towards global consensus on outcome measures for atopic eczema research: Results of the HOME II meeting. *Allergy* 2012;67:1111-1117.
9. 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, <https://doi.org/10.1371/journal.pone.0017520>.
10. Schmitt J, Langan S, Stamm T, et al. Harmonizing Outcome Measurements in Eczema (HOME) Delphi panel. Core outcome domains for controlled trials and clinical recordkeeping in eczema: International multiperspective Delphi consensus process. *J Invest Dermatol* 2011;131: 623-630.
11. Consensus Report of the European Task Force on Atopic Dermatitis. Severity scoring of atopic dermatitis: The SCORAD index. *Dermatology* 1993;186:23-31.
12. Kunz B, Oranje AP, Labreze L, et al. Clinical validation and guidelines for the SCORAD index: Consensus report of the European Task Force on Atopic Dermatitis. *Dermatology* 1997;195:10-19.
13. Charman CR, Venn AJ, Williams H. Measuring atopic eczema severity visually: Which variables are most important to patients? *Arch Dermatol* 2005;141:1146-1151. discussion 1151.
14. Chalmers JR, Schmitt J, Apfelbacher C, et al. Report from the third international consensus meeting to harmonise core outcome measures for atopic eczema/dermatitis clinical trials (HOME). *Br J Dermatol* 2014;171: 1318-1325.
15. Oranje AP. Practical issues on interpretation of scoring atopic dermatitis: SCORAD Index, objective SCORAD, patient-oriented SCORAD and Three-Item Severity score. *Curr Probl Dermatol* 2011;41:149-155.
16. Schmitt J, Spuls PI, Thomas KS, et al. The Harmonising Outcome Measures for Eczema (HOME) statement to assess clinical signs of atopic eczema in trials. *J Allergy Clin Immunol* 2014;134:800-807.
17. Chopra R, Vakharia P, Sacotte R, et al. Severity strata for Eczema Area and Severity Index (EASI), modified EASI, Scoring Atopic Dermatitis (SCORAD), objective SCORAD, Atopic Dermatitis Severity Index, and body surface area in adolescents and adults with atopic dermatitis. *Br J Dermatol* 2017;177:1316-1321.
18. Abrams BB. Atopic dermatitis: Elements in clinical study design and analysis. *Acta Derm Venereol Suppl (Stockh)* 1989;144:15-19.
19. Leshem YA, Hajar T, Hanifin JM, et al. What the Eczema Area and Severity Index score tells us about the severity of atopic dermatitis: An interpretability study. *Br J Dermatol* 2015;172:1353-1357.
20. Reitamo S, Wollenberg A, Schopf E, et al. Safety and efficacy of 1 year of tacrolimus ointment monotherapy in adults with atopic dermatitis. The European Tacrolimus Ointment Study Group. *Arch Dermatol* 2000;136: 999-1006.
21. Feldman SR, Krueger GG. Psoriasis assessment tools in clinical trials. *Ann Rheum Dis* 2005;64(suppl 2):ii65-II68. discussion ii69-73.
22. Futamura M, Leshem YA, Thomas KS, et al. A systematic review of Investigator Global Assessment (IGA) in atopic dermatitis (AD) trials: Many options, no standards. *J Am Acad Dermatol* 2016;74: 288-294.
23. Charman C, Williams H. Outcome measures of disease severity in atopic eczema. *Arch Dermatol* 2000;136:763-769.
24. Barbier N, Paul C, Luger T, et al. Validation of the Eczema Area and Severity Index for atopic dermatitis in a cohort of 1550 patients from the pimecrolimus cream 1% randomized controlled clinical trials programme. *Br J Dermatol* 2004;150:96-102.
25. Sowden JM, Berth-Jones J, Ross JS, et al. Double-blind, controlled, crossover study of cyclosporin in adults with severe refractory atopic dermatitis. *Lancet* 1991;338:137-140.
26. Berth-Jones J, Graham-Brown RA, Marks R, et al. Long-term efficacy and safety of cyclosporin in severe adult atopic dermatitis. *Br J Dermatol* 1997;136:76-81.
27. Granlund H, Erkkö P, Sinisalo M, et al. Cyclosporin in atopic dermatitis: Time to relapse and effect of intermittent therapy. *Br J Dermatol* 1995;132:106-112.
28. Berth-Jones J, Finlay AY, Zaki I, et al. Cyclosporine in severe childhood atopic dermatitis: A multicenter study. *J Am Acad Dermatol* 1996;34: 1016-1021.
29. Van Leent EJ, Graber M, Thurston M, et al. Effectiveness of the ascomycin macrolactam SDZ ASM 981 in the topical treatment of atopic dermatitis. *Arch Dermatol* 1998;134:805-809.
30. Rajka G, Langeland T. Grading of the severity of atopic dermatitis. *Acta Derm Venereol Suppl (Stockh)* 1989;144:13-14.
31. Carroll CL, Balkrishnan R, Feldman SR, et al. The burden of atopic dermatitis: Impact on the patient, family, and society. *Pediatr Dermatol* 2005;22:192-199.
32. Schmitt J, Schmitt NM, Kirch W, et al. Outpatient care and medical treatment of children and adults with atopic eczema. *J Dtsch Dermatol Ges* 2009;7:345-351.
33. 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;72:146-163.
34. Bahmer FA, Schafer J, Schubert HJ. Quantification of the extent and the severity of atopic dermatitis: The ADASI score. *Arch Dermatol* 1991;127: 1239-1240.
35. 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;79:356-359.
36. Verwimp JJ, Bindels JG, Barents M, et al. Symptomatology and growth in infants with cow's milk protein intolerance using two different whey-protein hydrolysate based formulas in a Primary Health Care setting. *Eur J Clin Nutr* 1995;49(suppl 1):S39-S48.
37. Holm EA, Wulf HC, Thomassen L, et al. Assessment of atopic eczema: Clinical scoring and noninvasive measurements. *Br J Dermatol* 2007;157:674-680.
38. Chopra R, Vakharia P, Sacotte R, et al. Relationship between EASI and SCORAD severity assessments for atopic dermatitis. *J Allergy Clin Immunol* 2017;140:1708-1710.e1.
39. Schmitt J, Williams H, Home Development Group. Harmonising Outcome Measures for Eczema (HOME). Report from the First International Consensus Meeting (HOME 1), 24 July 2010, Munich, Germany. *Br J Dermatol* 2010;163:1166-1168.
40. Kim DH, Li K, Seo SJ, et al. Quality of life and disease severity are correlated in patients with atopic dermatitis. *J Korean Med Sci* 2012;27: 1327-1332.
41. Haeck IM, ten Berge O, van Velsen SG, et al. Moderate correlation between quality of life and disease activity in adult patients with atopic dermatitis. *J Eur Acad Dermatol Venereol* 2012;26:236-241.
42. Zhao CY, Tran AQ, Lazo-Dizon JP, et al. A pilot comparison study of four clinician-rated atopic dermatitis severity scales. *Br J Dermatol* 2015;173: 488-497.
43. Chopra R, Vakharia P, Silverberg JI, et al. Lack of uniformity in the methodology of atopic dermatitis clinical trials: A systematic review. *J Am Acad Dermatol* 2017;76:AB422.
44. Charman C, Chambers C, Williams H. Measuring atopic dermatitis severity in randomized controlled clinical trials: What exactly are we measuring? *J Invest Dermatol* 2003;120:932-941.
45. Zhao CY, Hao EY, Oh DD, et al. A comparison study of clinician-rated atopic dermatitis outcome measures for intermediate- to dark-skinned patients. *Br J Dermatol* 2017;176:985-992.
46. Heinl D, Prinsen CA, Deckert S, et al. Measurement properties of adult quality-of-life measurement instruments for eczema: A systematic review. *Allergy* 2016;71:358-370.

47. Chalmers JR, Simpson E, Apfelbacher CJ, et al. Report from the fourth international consensus meeting to harmonize core outcome measures for atopic eczema/dermatitis clinical trials (HOME initiative). *Br J Dermatol* 2016;175:69-79.
48. Ganemo A, Svensson A, Svedman C, et al. Usefulness of Rajka & Langeland eczema severity score in clinical practice. *Acta Derm Venereol* 2016;96:521-524.
49. Hon KL, Ma KC, Wong E, et al. Validation of a self-administered questionnaire in Chinese in the assessment of eczema severity. *Pediatr Dermatol* 2003;20:465-469.
50. 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;140:1513-1519.