

## Assessment of clinical signs of atopic dermatitis: A systematic review and recommendation

Jochen Schmitt, MD, MPH,<sup>a,b,\*</sup> Sinéad Langan, MD, PhD,<sup>c,\*</sup> Stefanie Deckert, MPH,<sup>a,\*</sup> Ake Svensson, MD, PhD,<sup>d</sup> Laura von Kobyletzki, MD, MSc,<sup>e</sup> Kim Thomas, PhD,<sup>f</sup> Phyllis Spuls, MD, PhD,<sup>g</sup> on the behalf of the Harmonising Outcome Measures for Atopic Dermatitis (HOME) Initiative Dresden, Germany, London and Nottingham, United Kingdom, Malmö and Karlstad, Sweden, and Amsterdam, The Netherlands

**Background:** Clinical signs are a core outcome domain for atopic dermatitis (AD) trials. The current lack of standardization of outcome measures in AD trials hampers evidence-based communication.

**Objective:** We sought to provide evidence-based recommendations for the measurement of clinical signs in AD trials and to inform the Harmonising Outcome Measures for Atopic Dermatitis Initiative.

**Methods:** We conducted a systematic review on measurement properties of outcome measurements for clinical signs of AD. We systematically searched MEDLINE and Embase (until October 1, 2012) for validation studies on instruments measuring the clinical signs of AD. Grading of the truth, discrimination, and feasibility of scales; methodological study quality; and recommendations were based on predefined criteria.

**Results:** Sixteen eligible instruments were identified, of which 2 were best validated. The Eczema Area and Severity Index has adequate validity, responsiveness, internal consistency, intraobserver reliability, and intermediate interobserver reliability but unclear interpretability and feasibility. The Severity Scoring of Atopic Dermatitis Index (SCORAD) has

adequate validity, responsiveness, interobserver reliability, and interpretability and unclear intraobserver reliability. Only the objective SCORAD (ie, the clinical signs domain of the SCORAD) is internally consistent. The Six Area, Six Sign Atopic Dermatitis Index severity score and Three Item Severity Score fulfill some quality criteria, but the performance in other required measurement properties is unclear. The Patient-oriented Eczema Measure is reliable and responsive but has inadequate content validity to assess clinical signs of AD. The remaining 11 scales have either (almost) not been validated or performed inadequately.

**Conclusions:** The Eczema Area and Severity Index and SCORAD are the best instruments to assess the clinical signs of AD. The other 14 instruments identified are (currently) not recommended because of unclear or inadequate measurement properties. (J Allergy Clin Immunol 2013;132:1337-47.)

**Key words:** Atopic dermatitis, evidence-based medicine, systematic review, severity of illness index, reliability, validity, responsiveness

From <sup>a</sup>the Centre for Evidence-based Healthcare, University Hospital Dresden; <sup>b</sup>the Institute and Polyclinic of Occupational and Social Medicine, Technical University Dresden; <sup>c</sup>the Faculty of Epidemiology and Population Health, London School of Hygiene and Tropical Medicine; <sup>d</sup>the Institution for Clinical Science, Lund University, Department of Dermatology, Malmö; <sup>e</sup>the Department of Public Health Sciences, Karlstad University; <sup>f</sup>the Centre of Evidence Based Dermatology, University of Nottingham; and <sup>g</sup>the Department of Dermatology, Academic Medical Centre, University of Amsterdam.

\*These authors contributed equally to this work.

This publication describes work conducted on behalf of the Harmonising Outcome Measures for Atopic Dermatitis (HOME) Initiative, which is supported through an independent research grant funded by the National Institute for Health Research (NIHR) under its Programme Grants for Applied Research funding scheme (RP-PG-0407-10177). The views expressed in this publication are those of the author or authors and not necessarily those of the National Health Service, the NIHR, or the Department of Health.

Disclosure of potential conflict of interest: S. Langan has received one or more grants from or has one or more grants pending with the National Institute for Health Research (NIHR; UK Department of Health). L. von Kobyletzki is a Board member for 100000DK, is employed by 156000SEK, and has received one or more grants from or has one or more grants pending with 452500SEK. K. Thomas has been supported by an NIHR Programme Grant for Applied Research. The rest of the authors declare that they have no relevant conflicts of interest.

Received for publication May 3, 2013; revised July 11, 2013; accepted for publication July 11, 2013.

Available online September 12, 2013.

Corresponding author: Jochen Schmitt, MD, MPH, Centre for Evidence-based Healthcare, University Hospital Dresden, Fetscherstraße 74, 01307 Dresden, Germany.

E-mail: jochen.schmitt@uniklinikum-dresden.de.

0091-6749/\$36.00

© 2013 American Academy of Allergy, Asthma & Immunology

http://dx.doi.org/10.1016/j.jaci.2013.07.008

Atopic dermatitis (AD) is a major medical condition that causes substantial burden to patients, their families, and society.<sup>1-4</sup> Various interventions exist, many of which have been assessed in randomized controlled trials.<sup>5</sup> However, clinical trials are only as credible as their end points.<sup>6</sup>

A previous systematic review<sup>7</sup> published in 2007 indicated that most of the published and applied outcome measures for AD had either been inadequately validated or had performed poorly when validated. None of the outcome measures identified in our 2007 article met all the requirements for measurement instruments to be “highly recommended” for use.<sup>7</sup>

Because of substantial variation in AD outcome measure<sup>8</sup> trials, interventions are not comparable. The lack of standardization of AD outcome measures currently makes evidence-based decision making impossible and hampers scientific communication. The international, multiprofessional Harmonising Outcome Measures for Eczema (HOME) Initiative aims to standardize and validate outcome measurements for AD.<sup>9</sup> By using Delphi Study methodology,<sup>10</sup> the following 4 criteria of AD have been defined as core outcome domains to be assessed in all future AD trials<sup>11,12</sup>: (1) clinical signs measured with a physician-assessed instrument; (2) symptoms measured with a patient-assessed instrument; (3) health-related quality of life; and (4) long-term control of flares.

The next crucial step in the process of standardization of AD outcome measurements is to identify appropriate instruments to measure each of the 4 core outcome domains of AD. There was broad international consensus among clinicians, patients, and

*Abbreviations used*

AD:	Atopic dermatitis
BSA:	Body surface area
COSMIN:	Consensus-based Standards for the Selection of Health Measurement Instruments
EASI:	Eczema Area and Severity Index
HOME:	The Harmonising Outcome Measures for Eczema Initiative
oSCORAD:	Objective Scoring Atopic Dermatitis Index
POEM:	Patient-oriented Eczema Measure
PO-SCORAD:	Patient-oriented Scoring Atopic Dermatitis Index
SA-EASI:	Self-administered Eczema Area and Severity Index
SASSAD:	Six Area, Six Sign Atopic Dermatitis Index
SCORAD:	Scoring Atopic Dermatitis Index
TIS:	Three Item Severity Score

methodologists that the quality criteria “truth, discrimination, and feasibility”<sup>13</sup> need to be met for AD outcome measures to be recommended by the HOME initiative.<sup>12</sup>

The objectives of this systematic review were to (1) systematically assess measurement properties of outcome measurements for AD signs, (2) provide evidence-based recommendations for the measurement of clinical signs in future AD trials, and (3) prioritize future validation studies concerning instruments to assess the clinical signs of AD.

## METHODS

We conducted a systematic review on the measurement properties of outcome measures for clinical signs of AD according to an *a priori* study protocol. The protocol of our review was registered in the international prospective register of systematic reviews (PROSPERO; registration no. CRD42013003935, [http://www.crd.york.ac.uk/PROSPERO/display\\_record.asp?ID=CRD42013003935](http://www.crd.york.ac.uk/PROSPERO/display_record.asp?ID=CRD42013003935)).

### Eligibility criteria

On the basis of the research question, inclusion and exclusion criteria were defined for literature search and study selection (Table I). Only articles reporting on AD (population) and validation studies (study design) that investigated clinical signs (outcome) were considered. Articles in which an eligible scale was published first (“inauguration articles”) and articles that subsequently explicitly investigated the measurement properties of eligible scales were eligible for this review. Articles that reported an eligible scale (eg, outcome in a clinical trial without any explicit validation) were not considered eligible.

### Literature search

A comprehensive systematic literature search was carried out in MEDLINE (through the PubMed interface) and Embase (through the Ovid interface) for all articles published until October 1, 2012. A specific search string was developed and included search terms for AD, severity of illness, and measurement properties using a combination of medical subject headings and key words. For validation studies, sensitive search terms recommended by Terwee et al<sup>14</sup> were applied. We did not use language restrictions for selecting studies. Further details on the search string are provided in Table E1 in this article's Online Repository at [www.jacionline.org](http://www.jacionline.org).

### Study selection and data extraction

Initially, titles and abstracts of studies identified by means of electronic searches were screened independently by 3 reviewers (J.S., S.L., and S.D.). Subsequently, full texts of the remaining studies were independently examined by the same reviewers to determine whether the inclusion criteria clearly related to the *a priori* defined research question. Full texts that did not fulfill the

**TABLE I.** Inclusion and exclusion criteria of study selection

	Inclusion	Exclusion
Population	AD (synonyms: atopic eczema, eczema, neurodermatitis)	All others
Outcome	Clinical signs (eg, edema, erythema, redness, lichenification) and severity by means of multiple items	Exclusively symptoms, quality of life, or other domains investigated biomarker or skin physiology measures as comparators
Study design	Validation studies or inauguration articles with $\geq 1$ of the following measurement properties: content validity, construct validity, internal consistency, reliability, responsiveness, floor or ceiling effects, interpretability, acceptability	Articles that report an eligible scale without any explicit validation linguistic validation studies, which assess discriminant validity only

predefined criteria were excluded. All disagreements were discussed in consensus teleconferences. The reasons for exclusion (eg, irrelevant populations, absence of clinical signs, and articles that report an eligible scale without any explicit validation) were documented for each article. In addition to the electronic literature search, all reference lists of studies included and key articles on this topic were screened for potentially relevant studies. Data from eligible studies, including study and scale characteristics, were independently extracted by 4 review teams (J.S./A.S., S.L./L.vK., K.T./S.D., and P.S./J.S.) using standardized and pilot-tested evidence tables. These 4 review teams also extracted data on measurement properties of outcome measures for clinical signs of AD and data on the methodological quality of the studies included. Disagreements among the 4 review teams were resolved by consensus within each team.

### Assessment of measurement properties of outcome measures for clinical signs of AD

Measurement properties assessed in this review were content validity, construct validity, internal consistency, interobserver reliability, intraobserver reliability, responsiveness/sensitivity to change, floor and ceiling effects, interpretability, and acceptability/ease of use.<sup>15</sup> The predefined criteria of rating properties of outcome measures are in accordance with the recommendations of the Consensus-based Standards for the Selection of Health Measurement Instruments (COSMIN) group<sup>16</sup>; the OMERACT filter,<sup>13</sup> which has been adopted by the HOME initiative<sup>12</sup>; and the criteria applied in our previous review on AD outcome measures.<sup>7</sup> Table II provides details concerning definitions of the measurement properties assessed and the criteria for an adequate (+), intermediate (+/-), and inadequate (-) rating. Criterion validity refers to the extent to which a measurement relates to a gold standard. The absence of a gold standard to measure the clinical signs of AD was the main reason to perform this review. Therefore the aspect of criterion validity was not considered in this review.

### Assessment of the methodological quality of included studies

The COSMIN checklist<sup>17,18</sup> was used to evaluate the methodological quality of included studies. On the basis of the extended version of this checklist,<sup>19</sup> an overall methodological quality score was calculated by using a 4-point scale (ie, poor, fair, good, or excellent) for each item of the COSMIN checklist. As recommended, the overall score for each item was obtained by taking the lowest score for any question within the item.<sup>19</sup> The COSMIN checklist was not applicable for the psychometric properties of floor or ceiling effect, interpretability, and ease of use.

### Data synthesis

For each instrument, the evidence concerning each of the abovementioned measurement properties was summarized separately. Considering all evidence

**TABLE II.** Methods and definitions to rate the quality of outcome measures

Measurement property (name)	Measurement property (description)	Inclusion in OMERACT filter	Criteria for adequate rating (+)	Criteria for intermediate rating (+/–)	Criteria for inadequate rating (–)
Content validity	Extent to which the domain “signs” is comprehensively represented by the items in the instrument	Truth	Clinicians OR patients were involved in item selection; clinicians AND patients consider $\geq 90\%$ of items as representative for the domain atopic eczema signs	Clinicians OR patients were involved in item selection; clinicians AND patients consider at 70% to 89% of items as representative for the domain atopic eczema signs	NEITHER clinicians NOR patients were involved in item selection; clinicians OR patients consider $< 70\%$ of items as representative for the domain atopic eczema signs
Construct validity	Extent to which an instrument truly measures the construct “atopic eczema signs”	Truth	Two different instruments that aim to measure signs of atopic eczema show high correlation (ie, factor loading/correlation coefficient $> 0.70$ ); OR specific hypotheses were formulated AND $\geq 75\%$ of the results are in accordance with these hypotheses	Two different instruments that aim to measure signs of atopic eczema show high correlation (ie, factor loading/correlation coefficient 0.60-0.69)	Two different instruments that aim to measure signs of atopic eczema do not show high correlation (ie, factor loading/correlation coefficient $< 0.60$ ); AND/OR specific hypotheses were formulated BUT $< 75\%$ of hypotheses were confirmed, despite adequate design and methods
Internal consistency	Extent to which items in the scale “signs” are interrelated (measure the same construct)	Discrimination	Factor analyses performed on adequate sample size (7 times the number of items and $> 100$ ) AND Cronbach $\alpha$ calculated per dimension AND Cronbach $\alpha$ of 0.70-0.95	No factor analysis OR doubtful design or method	Cronbach $\alpha$ $< 0.70$ or $> 0.95$ , despite adequate design and methods
Test-retest reliability/ intrarater variability	Extent to which repeated measurement by the same investigator provides identical results	Discrimination	Correlation coefficient $> 0.90$ OR percentage variation $< 5\%$ OR coefficient of variation $< 10\%$	Correlation coefficient 0.80-0.90 OR percentage variation 5% to 10% OR coefficient of variation 10% to 20%	Correlation coefficient $< 8.80$ Percentage variation $> 10\%$ OR coefficient of variation $> 20\%$
Interobserver reliability	Extent to which application of the instrument by different investigators provides identical results	Discrimination	Intraclass correlation $> 0.80$ OR weighted $\kappa$ $> 0.60$ OR coefficient of variation $< 20\%$ OR ANOVA (percentage variance explained by observer) $< 10\%$	Intraclass correlation 0.60-0.80 OR weighted $\kappa$ 0.40-0.60 OR coefficient of variation 20% to 30% OR ANOVA (percentage variance explained by observer) 10% to 20%	Intraclass correlation $< 0.60$ OR weighted $\kappa$ $> 0.4$ OR coefficient of variation $> 30\%$ OR ANOVA (percentage variance explained by observer) $> 20\%$
Sensitivity to change/ responsiveness	Ability of questionnaire to detect clinically important changes over time	Discrimination	MIC defined; MIC $>$ SDC	MIC defined, but doubtful design or method; MIC $>$ SDC	MIC not defined; MIC $\leq$ SDC
Interpretability	Degree to which one can assign qualitative meaning to quantitative scores (ie, ranges for clear/almost clear, mild, moderate, and severe eczema)	Feasibility	Ranges of the scale have been defined that represent “clear/almost clear,” “mild,” “moderate,” and “severe” atopic eczema/atopic eczema signs AND MIC defined	Mean and SD scores presented of $\geq 3$ relevant subgroups of patients but no clear ranges in scale for qualitative meaning of severity of signs defined OR no MIC defined	No information found on interpretation
Floor or ceiling effects	Number of respondents who achieved the lowest or highest possible score	Not required by OMERACT filter	$\leq 15\%$ of the respondents (of validation study but not in a given RCT) achieved the highest or lowest possible scores	Doubtful design or method	$> 15\%$ of the respondents (of validation study but not in a given RCT) achieved the highest or lowest possible scores despite adequate design and method
Acceptability/ ease of use	Degree to which the score can be applied easily, given constraints of time and money	Feasibility	Time to administer $< 7$ min AND score can be used without charge AND no specific tools needed	Time to administer 7-10 min AND score can be used without charge AND no specific tools needed	Time to administer $> 10$ min OR score cannot be used without charge OR specific tool needed

MIC, Minimal important change; RCT, randomized controlled trial; SDC, smallest detectable change.

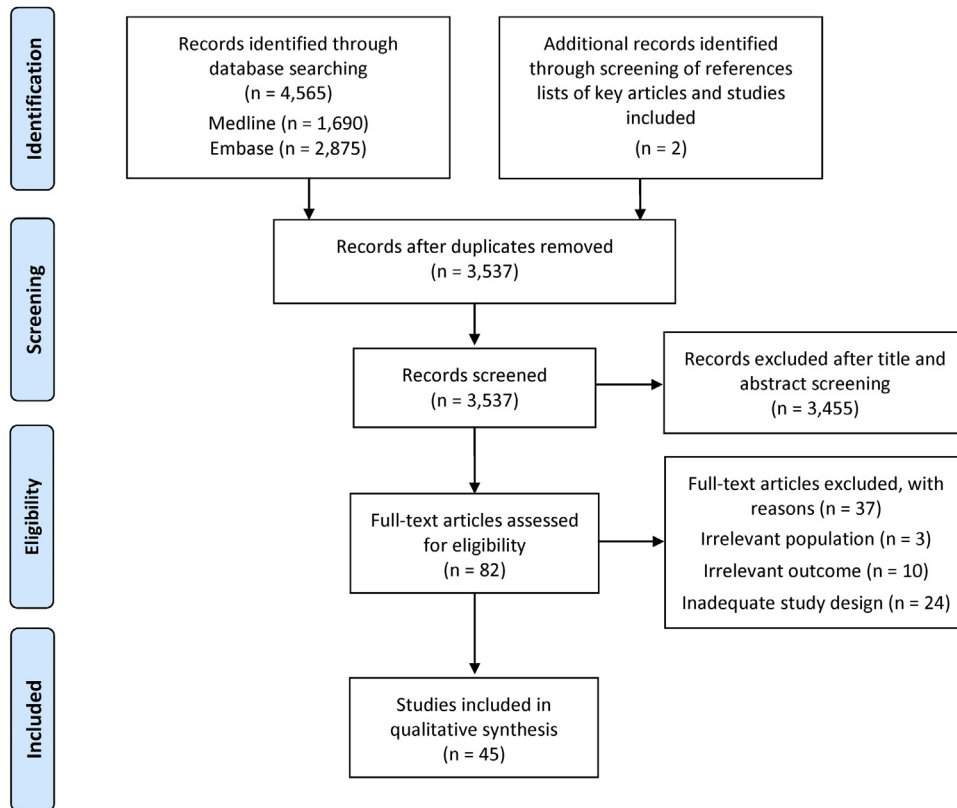


FIG 1. Flow chart of study selection procedure (according to the PRISMA statement).<sup>64</sup>

available, an overall scale quality concerning content validity, construct validity, internal consistency, interobserver reliability, intraobserver reliability, responsiveness/sensitivity to change, floor or ceiling effects, interpretability, and acceptability/ease of use of each outcome measure was rated as adequate (+), intermediate (+/-), or inadequate (-). If 2 studies came to different conclusions concerning a specific measurement property of a specific scale, the study with the higher methodological quality according to the COSMIN checklist<sup>17-19</sup> informed the overall rating.

### Predefined criteria for recommendations

For each instrument identified in the review, a standardized recommendation for use or required future validation work was prepared, depending on the measurement properties.

According to the results of the HOME II meeting,<sup>12</sup> all 3 criteria of the OMERACT filter (ie truth, discrimination, and feasibility)<sup>13</sup> have to be met by an outcome measure to be recommended by the HOME initiative.

Four categories of recommendation were specified as follows:

- The outcome measure meets the requirements of truth (content validity and construct validity), discrimination (reliability, internal consistency, and sensitivity to change), and feasibility (interpretability and ease of use) and is recommended for use.
- The outcome measure meets 2 or more requirements, but performance in other required measurement properties is unclear, so that the outcome measure has the potential to be recommended in the future depending on the results of further validation studies.
- The outcome measure has inadequate quality in at least 1 required measurement property ( $\geq 1$  rating of "minus") and therefore is not recommended to be used.
- The outcome measure has (almost) not been validated. Its performance in all or most relevant quality requirements is unclear, so that it is not recommended to be used until further validation studies clarify its measurement properties.

## RESULTS

We identified 45 eligible articles.<sup>7,20-63</sup> Of these, 43 studies were identified by means of electronic search, and 2 studies<sup>47,57</sup> were identified through hand search. The flow diagram (Fig 1)<sup>64</sup> provides information on study identification and the selection process. Table E2 in this article's Online Repository at [www.jacionline.org](http://www.jacionline.org) presents a list of the articles that were excluded based on full-text screening, including the reason for exclusion. Interrater reliability among the 3 reviewers was substantial<sup>65</sup> for title and abstract screening (Cohen  $\kappa = 0.75$ - $0.79$ ; agreement = 98.9% to 99.3%) and poor to almost perfect<sup>65</sup> for full-text inclusion (Cohen  $\kappa = 0.31$ - $0.84$ ; agreement = 65.5% to 92.3%).

### Study characteristics

The 45 articles reported on 16 different instruments to assess the clinical signs of AD. The characteristics of these 45 articles are summarized by outcome measurement in Table III.<sup>7,20-63</sup> Most validation studies investigated the Scoring Atopic Dermatitis Index (SCORAD)<sup>32</sup> and the Eczema Area and Severity Index (EASI).<sup>54</sup>

For both the SCORAD and EASI, more than 2000 patients with a broad spectrum of age groups and degrees of severity of AD were included in the identified validation studies (Table III).

### Content of the instruments identified

Thirteen different clinical signs are included in the identified instruments, with each instrument including 3 to 8 clinical signs. Erythema is the most frequently included (14/16 instruments). Lichenification is included in 10 of 16 instruments. The items edema/induration/papulation, oozing/crusting/weeping/exudation,

**TABLE III.** Study characteristics of studies included by outcome measurement\*

Outcome measurement	No. of validation studies	Study characteristics				
		Setting: Primary care Secondary/tertiary care	Severity of AD	Study population		
				Total	Per study (range)	Age (y [range])
SCORAD <sup>†32</sup>	26 <sup>22-24,26,28-30,32,34,35,37,45,46,48,49,51-53,55,56,58-63</sup>	Primary care, n = 2; secondary/tertiary care, n = 20; all settings, n = 2; NR, n = 2	Mild to severe, n = 14; moderate to severe, n = 5; severe, n = 1; NR, n = 6	4127	15-823	0-74
PO-SCORAD <sup>58</sup>	2 <sup>52,58</sup>	All secondary/tertiary care	Mild to severe, n = 1; NR, n = 1	504	33-471	2-74
EASI <sup>54</sup>	9 <sup>20,21,29,35,40,47,49,55,61</sup>	Secondary/tertiary care, n = 6; all settings, n = 1; NR, n = 2	Mild to moderate, n = 1; mild to severe, n = 5; severe, n = 1; NR, n = 2	2062	20-1550	0-70
SA-EASI <sup>21</sup>	4 <sup>21,23,24,56</sup>	All secondary/tertiary care	Moderate to severe, n = 3; NR, n = 1	240	35-98	0-17
SASSAD <sup>33</sup>	5 <sup>33,35,42,56,61</sup>	Secondary/tertiary care, n = 3; all settings, n = 1; NR, n = 1	Mild to severe, n = 3; moderate to severe, n = 1; mild to moderate, n = 1	340	6-180	0-67
POEM <sup>36</sup>	2 <sup>36,49</sup>	All settings, n = 1, NR, n = 1	Mild to severe, n = 1; severe, n = 1	120	40-80	1-58
BCSS <sup>57</sup>	1 <sup>51</sup>	Secondary/tertiary care	NR	82	NA	0-67
ADAM <sup>43</sup>	2 <sup>43,44</sup>	All secondary/tertiary care	Mild to severe, n = 1; moderate to severe, n = 1	224	53-171	0-16
ADASI <sup>38,39</sup>	1 <sup>39</sup>	Secondary/tertiary care	NR	100	NA	NR
ADQ <sup>34</sup>	1 <sup>34</sup>	Secondary/tertiary care	Mild to severe	68	NA	0-12
OSAAD <sup>53</sup>	3 <sup>37,45,53</sup>	All secondary/tertiary care	Mild to severe, n = 2; moderate to severe, n = 1	92	22-38	0-57
SSS <sup>31</sup>	2 <sup>31,51</sup>	All secondary/tertiary care	All mild to severe	96	7-34	0-67
TIS <sup>60</sup>	7 <sup>30,32,35,51,59,60,62</sup>	Primary care, n = 2; secondary/tertiary care, n = 4; all settings, n = 1	Mild to severe, n = 6; NR, n = 1	926	66-180	0-67
W-AZS <sup>50</sup>	0	NA	NA	NA	NA	NA
Unnamed scale 1 <sup>25</sup>	1 <sup>25</sup>	Secondary/tertiary care	NR	28	NA	0-8
Unnamed scale 2 <sup>25</sup>	1 <sup>25</sup>	Secondary/tertiary care	NR	28	NA	0-8

ADAM, Atopic Dermatitis Assessment Measure; ADASI, Atopic Dermatitis Area and Severity Index; ADQ, Atopic Dermatitis Quickscore; BCSS, Basic Clinical Scoring System; NA, not applicable; NR, not reported; OSAAD, Objective Severity Assessment of Atopic Dermatitis; SSS, Simple Scoring System; W-AZS, New Scoring System for Assessment of the Extent and Severity of Skin Inflammation Index in Atopic Dermatitis Patients.

\*Studies without reference to a specific outcome measure<sup>7,41</sup> are not considered here.

†SCORAD<sup>32</sup> and oSCORAD<sup>22</sup> were considered as a single instrument.

and excoriation are included in 9 of 16 instruments. Dryness and scaling are assessed in 8 and 6 instruments, respectively. Cracking/fissuring, vesicles, (de)pigmentation, flaking, bleeding, and erosions are included in 3 or less instruments to assess the clinical signs of AD (Table IV).\* All instruments except the Patient-oriented Eczema Measure (POEM)<sup>36</sup> and the Three Item Severity Score (TIS)<sup>60</sup> include an assessment of intensity and extent (affected body surface area [BSA], clinical signs at specific locations, or both) of clinical signs. The EASI<sup>54</sup>; TIS<sup>60</sup>; Six Area, Six Sign Atopic Dermatitis Index (SASSAD)<sup>33</sup>; and Basic Clinical Scoring System<sup>57</sup> are unidimensional instruments (ie, they only assess the domain clinical signs of AD). The other instruments included are multidimensional and also include symptoms of AD (n = 11),† epidermal function (n = 1),<sup>37</sup> and/or complications of AD (n = 2).<sup>25</sup> For the SCORAD,<sup>32</sup> which includes clinical signs and symptoms of

AD, a unidimensional version only including the intensity and extent of clinical signs, the objective Scoring Atopic Dermatitis Index (oSCORAD),<sup>22</sup> is available.

### Studies investigating content validity

Two studies<sup>7,35</sup> on the content validity of measures to assess the intensity and extent of clinical signs were identified. Charman et al<sup>35</sup> assessed clinical signs and BSA involvement in 180 consecutive patients with AD and analyzed their relationship with a patient global bother score by means of bivariate and multivariate regression. Excoriations, erythema, and edema/papulation were independent predictors of patient-rated disease severity. BSA involvement, which might not be reliably assessed,<sup>41</sup> was not linearly related with the global bother score of greater than 30% BSA involvement.<sup>35</sup> Schmitt et al<sup>7</sup> asked 12 consumers and 6 dermatology experts to rate the content validity of all items and domains included in the identified outcome measures for AD on a 5-point Likert scale. Both experts and consumers considered the intensity

\*References 21, 25, 31-34, 36, 38, 39, 43, 50, 53, 54, 57, 58, and 60.

†References 21, 25, 31, 32, 34, 36, 38, 43, 50, and 58.

**TABLE IV.** Description of outcome measurements identified for clinical signs of AD

Name of scale	Items of signs domain								
	Erythema	Edema/induration/ papulation	Oozing/crusting/ weeping/exudation	Excoriation	Lichenification	Dryness	Scaling	Cracking/ fissuring	Vesicles
ADAM, 1999 <sup>43</sup>	●			●	●		●		
ADASI, 1991 <sup>38,39</sup>	●	●	●		●		●		
ADQ, 2008 <sup>34</sup>	●		●			●	●	●	
BCSS, 1995 <sup>57</sup>									
EASI, 1998 <sup>54</sup>	●	●		●	●				
OSAAD, 2003 <sup>53</sup>	●		●	●	●				
POEM, 2004 <sup>36</sup>			●			●		●	
PO-SCORAD <sup>58</sup>	●	●	●	●	●	●	●	●	
SA-EASI, 2002 <sup>21</sup>	●	●		●		●			
SASSAD, 1996 <sup>33</sup>	●		●	●	●	●		●	
SCORAD, 1993 <sup>32</sup>	●	●	●	●	●	●			
SSS, 1989 <sup>31</sup>	●	●	●	●	●		●		●
TIS, 1999 <sup>60</sup>	●	●		●					
W-AZS, 2005 <sup>50</sup>	●	●	●				●		●
Unnamed scale 1, 2004 <sup>25</sup>	●	●			●	●			
Unnamed scale 2, 2004 <sup>25</sup>	●				●	●			

ADAM, Atopic Dermatitis Assessment Measure; ADASI, Atopic Dermatitis Area and Severity Index; ADQ, Atopic Dermatitis Quickscore; BCSS, Basic Clinical Scoring System; OSAAD, Objective Severity Assessment of Atopic Dermatitis; SSS, Simple Scoring System; W-AZS, New Scoring System for Assessment of the Extent and Severity of Skin Inflammation Index in Atopic Dermatitis Patients.

\*The oSCORAD does not contain symptoms.

**TABLE V.** Number of studies assessing the measurement properties of outcome measures for clinical signs of AD

Measurement property (name)	ADAM <sup>43</sup>	ADASI <sup>38,39</sup>	ADQ <sup>34</sup>	BCSS <sup>57</sup>	EASI <sup>54</sup>	OSAAD <sup>53</sup>	POEM <sup>36</sup>
Content validity	2 <sup>7,35</sup>	2 <sup>7,35</sup>	2 <sup>7,35</sup>	2 <sup>7,35</sup>	2 <sup>7,35</sup>	2 <sup>7,35</sup>	3 <sup>7,35,36</sup>
Construct validity	—	—	1 <sup>34</sup>	1 <sup>51</sup>	5*	3 <sup>37,45,53</sup>	—
Internal consistency	1 <sup>44</sup>	—	—	—	2 <sup>29,40</sup>	—	1 <sup>36</sup>
Intraobserver reliability	—	—	—	—	1 <sup>47</sup>	—	1 <sup>36</sup>
Interobserver reliability	1 <sup>43</sup>	—	—	1 <sup>51</sup>	2 <sup>29,47</sup>	1 <sup>37</sup>	—
Responsiveness	—	—	—	—	2 <sup>40,49</sup>	1 <sup>37</sup>	2 <sup>36,49</sup>
Floor or ceiling effects	—	—	—	—	—	2 <sup>37,53</sup>	—
Interpretability	—	—	—	—	—	—	—
Acceptability	—	1 <sup>39</sup>	1 <sup>34</sup>	—	—	1 <sup>37</sup>	1 <sup>36</sup>

ADAM, Atopic Dermatitis Assessment Measure; ADASI, Atopic Dermatitis Area and Severity Index; ADQ, Atopic Dermatitis Quickscore; BCSS, Basic Clinical Scoring System; OSAAD, Objective Severity Assessment of Atopic Dermatitis; SSS, Simple Scoring System; W-AZS, New Scoring System for Assessment of the Extent and Severity of Skin Inflammation Index in Atopic Dermatitis Patients.

\*References 20, 21, 29, 55, and 61.

†References 21, 23, 24, and 56.

‡References 23, 24, 29, 30, 34, 37, 45, 51-53, 55, 56, 58-61, and 63.

§References 28, 29, 32, 45, 46, 48, and 62.

||References 22, 23, 26, 29, 32, 34, 37, 51, 59, 60, and 62.

¶References 32, 51, 59, 60, and 62.

of lesions and the extent of AD as “very important” criteria for the assessment of the severity of AD. Excoriations, erythema, edema/papulation, dryness, lichenification, and oozing/crusting/exudation were considered “important” or “very important” by consumers and experts.<sup>7</sup>

### Measurement properties of outcome measures for clinical signs of AD

Table V<sup>7,20-63</sup> summarizes the number of studies assessing the different measurement properties of each outcome measure identified. From the total of 146 assessed (sub)studies included, the COSMIN checklist was applicable for 123 validation (sub)studies. Of these, 17 (14%) were rated as having excellent, 8 (6%)

as having good, 54 (44%) as having fair, and 44 (36%) as having poor quality. The most frequent reason for fair or poor quality was low sample size. Table E3 in this article’s Online Repository at [www.jacionline.org](http://www.jacionline.org) provides detailed information about the study base, methods, results, quality, and conclusions of each of these validation (sub)studies. The SCORAD is the only instrument for which each measurement property of interest has been investigated according to our predefined criteria (Tables II and V).

Table VI<sup>7,20-63</sup> provides a summary of measurement properties of each outcome measure for clinical signs of AD and recommendations. None of the studies met all of our predefined requirements of truth, discrimination, and feasibility. Currently, the EASI and SCORAD are the most valid and reliable instruments to assess clinical signs of AD.

TABLE IV. (Continued)

(De)pigmentation	Flaking	Bleeding	Erosions	Consideration of intensity and extent of signs		Other domains		
				Only intensity of signs	Intensity AND extent of signs	Symptoms	Epidermal function	Complications
					•	•		
					•	•		
					•	•		
					•			
					•			
					•		•	
	•	•			Frequency of signs	•		
		•				•		
						•		
						•		
						•*		
•						•		
				•				
•			•			•		
			•			•		•
			•			•		•

TABLE V. (Continued)

PO-SCORAD <sup>58</sup>	SA-EASI <sup>21</sup>	SASSAD <sup>33</sup>	SCORAD <sup>32</sup>	SSS <sup>31</sup>	TIS <sup>60</sup>	W-AZS <sup>50</sup>	Unnamed scale 1 <sup>25</sup>	Unnamed scale 2 <sup>25</sup>
2 <sup>7,35</sup>	2 <sup>7,35</sup>	2 <sup>7,35</sup>	3 <sup>7,32,35</sup>	2 <sup>7,35</sup>	2 <sup>7,35</sup>	2 <sup>7,35</sup>	2 <sup>7,35</sup>	2 <sup>7,35</sup>
2 <sup>52,58</sup>	4 <sup>†</sup>	2 <sup>56,61</sup>	17 <sup>‡</sup>	1 <sup>51</sup>	3 <sup>30,59,60</sup>	—	—	—
—	—	—	7 <sup>§</sup>	—	1 <sup>32</sup>	—	—	—
—	—	1 <sup>42</sup>	1	—	—	—	—	—
—	1 <sup>24</sup>	1 <sup>42</sup>	11 <sup>  </sup>	2 <sup>31,51</sup>	5 <sup>¶</sup>	—	1 <sup>25</sup>	1 <sup>25</sup>
—	—	—	2 <sup>37,49</sup>	—	—	—	—	—
—	—	1 <sup>33</sup>	2 <sup>32,62</sup>	1 <sup>31</sup>	2 <sup>59,60</sup>	—	—	—
—	—	—	2 <sup>22,27</sup>	—	1 <sup>59</sup>	—	—	—
1 <sup>58</sup>	1 <sup>24</sup>	1 <sup>33</sup>	2 <sup>32,62</sup>	—	1 <sup>59</sup>	—	1 <sup>25</sup>	1 <sup>25</sup>

The EASI and SCORAD both have adequate content validity.<sup>7,35</sup> High correlation between the EASI and SCORAD ( $r = 0.84\text{--}0.93$ <sup>29,61</sup>) indicates adequate construct validity of both instruments. Correlation between the EASI and the Self-administered Eczema Area and Severity Index (SA-EASI;  $r = 0.17\text{--}0.62$ <sup>21</sup>) and between the SCORAD and the Patient-oriented Scoring Atopic Dermatitis Index (PO-SCORAD;  $r = 0.27\text{--}0.79$ <sup>52,58</sup>) was weaker, suggesting that the patient-assessed modifications of the EASI and SCORAD measure different constructs than the original investigator-assessed instruments.<sup>66</sup> For an overview of the strength of correlation between the different instruments included in this review, please refer to Table E4 in this article's Online Repository at [www.jacionline.org](http://www.jacionline.org). The EASI is internally consistent (Cronbach

$\alpha = 0.94$ <sup>29</sup>). The SCORAD is a composite score, and therefore internal consistency within the constructs assessed (signs and symptoms) is relevant. The objective parts of the SCORAD that assess the intensity and extent of clinical signs are strongly interrelated ( $r = 0.64\text{--}0.86$ ),<sup>28,48</sup> suggesting internal consistency of the objective SCORAD.<sup>22</sup> The EASI had adequate intraobserver reliability and intermediate interobserver reliability in a population of 20 children and adults with different degrees of severity of AD assessed by 15 trained investigators.<sup>47</sup> Interobserver reliability of the SCORAD was adequate in different populations of children and adults with different degrees of disease severity (see Table E3 for details). Intraobserver reliability of the SCORAD was investigated in one study,<sup>32</sup> but grading of intraobserver reliability according to

**TABLE VI.** Summary of measurement properties of outcome measures for clinical signs of AD and recommendations

Measurement property (name)	ADAM <sup>43</sup>	ADASI <sup>38,39</sup>	ADQ <sup>34</sup>	BCSS <sup>57</sup>	EASI <sup>54</sup>	OSAAD <sup>53</sup>	POEM <sup>36</sup>
Content validity	–	+/–	–	–	+	–	–
Construct validity	NR	NR	+/–	–	+	+/–	NR*
Internal consistency	+	NR	NR	NR	+	NR	+
Intraobserver reliability	NR	NR	NR	NR	+	NR	+
Interobserver reliability	+/–	NR	NR	+	+/–	+	NA
Sensitivity to change	NR	NR	NR	NR	+	NR	+
Floor or ceiling effects	NR	NR	NR	NR	NR	+	NR
Interpretability	NR	NR	NR	NR	NR	NR	NR
Acceptability	NR	+	+	NR	NR	–	+
Recommendation	C	D	C	C	B	C	C

Recommendations are defined as follows: *A*, outcome measure meets the requirements of truth (content validity and construct validity), discrimination (reliability, internal consistency, and sensitivity to change), and feasibility (interpretability and ease of use) and is recommended for use; *B*, outcome measure meets 2 or more requirements, but performance in other required measurement properties is unclear, so that the outcome measure has the potential to be recommended in the future depending on the results of further validation studies; *C*, outcome measure has inadequate quality in at least 1 required measurement properties ( $\geq 1$  rating of "minus") and therefore is not recommended to be used; and *D*, outcome measure has (almost) not been validated. Its performance in all or most relevant quality requirements is unclear, so that it is not recommended to be used until further validation studies clarify its measurement properties.

*ADAM*, Atopic Dermatitis Assessment Measure; *ADASI*, Atopic Dermatitis Area and Severity Index; *ADQ*, Atopic Dermatitis Quickscore; *BCSS*, Basic Clinical Scoring System; *NA*, not applicable; *NR*, not reported, indicating measurement property unclear; *OSAAD*, Objective Severity Assessment of Atopic Dermatitis; *W-AZS*, New Scoring System for Assessment of the Extent and Severity of Skin Inflammation Index in Atopic Dermatitis Patients.

+, Positive rating indicating "adequate" measurement property; +/-, intermediate rating indicating "intermediate" measurement property; –, inadequate rating indicating "inadequate" measurement property (please refer to Table II for definitions).

\*Compared with other clinical sign measures.

†Further research is necessary to clarify whether an outcome measure with intermediate content validity is sufficient to capture clinical signs of AD in clinical trials.

our predefined criteria was not possible. In a study that used an investigator global assessment as the anchor, the minimal clinical important difference was 8.7 points for the SCORAD, 8.2 for the objective SCORAD, and 6.6 for the EASI,<sup>49</sup> indicating adequate sensitivity to change. Reports on the feasibility of the EASI (ie, floor or ceiling effects), interpretability, and ease of use were not identified. The SCORAD had no floor or ceiling effects in 2 studies.<sup>32,62</sup> Score ranges to define mild, moderate, and severe AD have been defined by expert opinion for the SCORAD<sup>27</sup> and oSCORAD.<sup>22</sup> The time to perform the SCORAD has been reported to be "about 10 minutes."<sup>32</sup>

The SASSAD and TIS fulfill some quality criteria (Table VI), but the performance in other required measurement properties is unclear. These outcome measures might have the potential to be recommended in the future depending on the results of further validation studies. However, both the SASSAD and TIS only have intermediate content validity, which is a disadvantage compared with the EASI and SCORAD.

Seven outcome measures (the Atopic Dermatitis Assessment Measure, Atopic Dermatitis Quickscore, Basic Clinical Scoring System, Objective Severity Assessment of Atopic Dermatitis, POEM, Simple Scoring System, and New Scoring System for Assessment of the Extent and Severity of Skin Inflammation Index in Atopic Dermatitis Patients [W-AZS]<sup>31,34,36,43,50,53,57</sup>) have inadequate quality in at least 1 of the required measurement properties and are therefore not recommended to be used to assess clinical signs of AD. The POEM is internally consistent, reliable, responsive, and feasible<sup>36,49</sup> but is primarily a symptoms scale that does not capture the content required of a clinical signs instrument.

Five outcome measures (unnamed scale 1, unnamed scale 2, PO-SCORAD, SA-EASI, and the Atopic Dermatitis Area and Severity Index<sup>21,25,39,58</sup>) have almost not been validated, and their measurement properties are unclear. These outcome measures<sup>21,25,39,58</sup> are currently not recommended (Table VI). Table

E3 provides a summary of all validation studies on the 16 instruments to assess clinical signs of AD.

## DISCUSSION

This systematic review identified, summarized, and critically appraised 16 different outcome measures to assess the severity of clinical signs of AD. Although none of these instruments completely met all the predefined criteria for adequate truth, discrimination, and feasibility, substantial validation work has been undertaken since our previous review.<sup>7</sup>

The EASI<sup>54</sup> and SCORAD<sup>32</sup> are the best instruments available to measure clinical signs of AD.

The SCORAD is a valid, internally consistent, responsive, interpretable composite score that includes the intensity and extent of clinical signs of AD and the severity of AD symptoms. The SCORAD has adequate interobserver reliability and no floor or ceiling effects. The oSCORAD is internally consistent. Intraobserver reliability of the SCORAD is unclear and requires further investigation. Disadvantages of the SCORAD include the difficulty of identifying an "average representative area" for the assessment of the intensity of clinical signs and the underlying assumption that the extent of AD is linearly related to severity, which has been shown to be only true in patients with up to 30% BSA affected.<sup>35</sup>

The EASI is valid and internally consistent and has adequate intraobserver reliability, intermediate interobserver reliability, and adequate responsiveness. One advantage of the EASI is its unidimensionality; that is, it only measures clinical signs of AD and does not include symptoms. Another advantage of the EASI is that it measures the intensity of the lesions at multiple body parts (rather than relying on a representative lesion). Ease of use and floor or ceiling effects of the EASI have not yet been reported in a validation study. However, from clinical trials that applied the EASI,<sup>8</sup> we have some evidence concerning the absence of floor or



TABLE VI. (Continued)

PO-SCORAD <sup>58</sup>	SA-EASI <sup>21</sup>	SASSAD <sup>33</sup>	SCORAD <sup>32</sup>	SSS <sup>31</sup>	TIS <sup>60</sup>	W-AZS <sup>50</sup>	Unnamed sale 1 <sup>25</sup>	Unnamed sale 2 <sup>25</sup>
+/-	+	+/-	+	-	+/-	-	+/-	+/-
+/-	+/-	+	+	NR	+/-	NR	NR	NR
NR	NR	NR	+/-	NR	+/-	NR	NR	NR
NR	NR	NR	NR	NR	NR	NR	NR	NR
NA	NA	+/-	+	NR	+/-	NR	+	+
NR	NR	NR	+	NR	NR	NR	NR	NR
NR	NR	+	+	+	+	NR	NR	NR
NR	NR	NR	+	NR	+	NR	NR	NR
+	NR	+	+/-	NR	+	NR	+	+
D	D	B†	B	C	B†	C	D	D

ceiling effects, and personal experience indicates that the time needed to perform the EASI and oSCORAD is similar. The main disadvantage of the EASI is that interpretability data are lacking (ie, definition of ranges of the EASI that represent mild, moderate, and severe clinical signs of AD). Ideally, interpretability of the EASI should be investigated with a global (gestalt) assessment of the severity of AD signs as a reference measure.

Clinical signs are typically assessed by an investigator.<sup>66</sup> However, for the EASI and SCORAD, patient-assessed variants have been introduced. As highlighted in this review, content and measurement properties differ substantially between the EASI and SA-EASI and between the SCORAD and PO-SCORAD, so that these patient-assessed instruments cannot be used interchangeably with the original investigator-assessed outcome measures. On the basis of the existing evidence, the SA-EASI and PO-SCORAD are not recommended to assess clinical signs of AD.

The conceptual framework (ie, the underlying relationship between the items of a measurement instrument and the construct [domain] “clinical signs of AD” to be measured)<sup>66</sup> follows a formative model. This means that the items determine or cause the construct to be assessed.<sup>66</sup> In such a situation, content validity is of utmost importance. Content validity has been defined as “the degree to which the content of a measurement instrument is an adequate reflection of the construct to be measured.”<sup>17</sup> Two studies<sup>7,35</sup> included in the review informed content validity and suggest that the extent of AD and the intensity of excoriations, erythema, edema/papulation, dryness, lichenification, and oozing/crusting/exudation were considered as “important” or “very important” by consumers and experts. The EASI and SCORAD are the only outcome measures that have adequate content validity according to our predefined criteria. Although the SASSAD and TIS fulfill some important quality criteria and therefore received recommendation B (ie, possibly recommendable if validation

gaps are closed), a consensus involving the relevant stakeholders is necessary to clarify whether an outcome measure with intermediate content validity is sufficient to capture the clinical signs of AD in clinical trials.

### Strengths and limitations of this systematic review

We applied predefined criteria to rate the measurement properties of instruments and make recommendations. The protocol was preregistered. We used a validated, highly sensitive search strategy for validation studies,<sup>14</sup> which identified all articles included in our previous review<sup>7</sup> listed in PubMed, Embase, or both. We did not consider unpublished data or studies only published as an abstract because we believed that relevant information would be missing. We applied the COSMIN checklist methodology<sup>15,17-19,66</sup> to rate study quality and define the criteria for adequate measurement properties. Even though this checklist was originally developed for the assessment of the methodological quality of patient-reported outcome measures,<sup>17</sup> we believe that it is also applicable for non-patient-reported outcome measures.

Reliability is substantially influenced by the variability between patients and cannot be considered a fixed characteristic of an instrument, such as content validity.<sup>66</sup> The degree of variability of the severity of clinical signs of AD was heterogeneous across the populations assessed in the reliability studies, so that the reliability results might not be directly comparable between the different instruments.

### Recommendations for researchers, clinicians, and decision makers

Clinical sign measures have been the primary end point in the vast majority of all clinical trials undertaken to investigate the efficacy of interventions for AD.<sup>5,6,8</sup> Clinical signs assessed by a physician have been included in the core set of outcome domains

for AD trials based on a 100% consensus between the 43 panel members participating in our previous multiperspective, international Delphi consensus study on outcome measurement in patients with AD.<sup>11</sup>

This review indicates that either the EASI or SCORAD should be used to assess the clinical signs of AD in future clinical trials. If the SCORAD is used, it should be reported as a profile (ie, the scores for the intensity and extent of AD signs and the severity of AD symptoms should be reported separately).

The ultimate goal from the perspective of evidence-based medicine is to achieve worldwide consensus to consistently apply a single valid, reliable, and feasible instrument to measure the clinical signs of AD in all future clinical trials. The HOME Initiative ([www.homeforeczema.org/](http://www.homeforeczema.org/)) is dedicated to achieving this goal.

### Key messages

- Valid, reliable, and feasible outcome measures are a prerequisite to evidence-based health care.
- This systematic review on outcome measures for clinical signs of AD identified and critically appraised 16 different instruments.
- The EASI and SCORAD have the best measurement properties and are recommended for use in future clinical trials.

### REFERENCES

1. Bieber T. Atopic dermatitis. *N Engl J Med* 2008;358:1483-94.
2. Carroll CL, Balkrishnan R, Feldman SR, Fleischer AB Jr, Manuel JC. The burden of atopic dermatitis: impact on the patient, family, and society. *Pediatr Dermatol* 2005;22:192-9.
3. Suh DC, Sung J, Gause D, Raut M, Huang J, Choi IS. Economic burden of atopic manifestations in patients with atopic dermatitis—analysis of administrative claims. *J Manag Care Pharm* 2007;13:778-89.
4. Williams HC. Clinical practice. Atopic dermatitis. *N Engl J Med* 2005;352:2314-24.
5. Schmitt J, Schakel K, Schmitt N, Meurer M. Systemic treatment of severe atopic eczema: a systematic review. *Acta Derm Venereol* 2007;87:100-11.
6. Schmitt J, Apfelbacher CJ, Flohr C. Eczema. *Clin Evid (Online)* 2011;2011.
7. Schmitt J, Langan SM, Williams HC. What are the best outcome measurements for atopic eczema?—A systematic review. *J Allergy Clin Immunol* 2007;120:1389-98.
8. Rehal B, Armstrong A. Health outcome measures in atopic dermatitis: a systematic review of trends in disease severity and quality-of-life instruments 1985-2010. *PLoS One* 2011;6:e17520.
9. Schmitt J, Williams HC. Harmonising Outcome Measures for Eczema (HOME). Report from the first international consensus meeting (HOME 1), 24 July 2010, Munich, Germany. *Br J Dermatol* 2010;163:1166-8.
10. Pill J. The Delphi method: substance, context, a critique and an annotated bibliography. *Socio-Economic Planning Science* 1971;5:57-71.
11. Schmitt J, Langan S, Stamm T, Williams HC. Core outcome domains for controlled trials and clinical recordkeeping in eczema: international multiperspective Delphi consensus process. *J Invest Dermatol* 2011;131:623-30.
12. Schmitt J, Spuls P, Boers M, Thomas K, Chalmers J, Roekvisch E, et al. Towards global consensus on outcome measures for atopic eczema research: results of the HOME II meeting. *Allergy* 2012;67:1111-7.
13. Boers M, Brooks P, Strand CV, Tugwell P. The OMERACT filter for outcome measures in rheumatology. *J Rheumatol* 1998;25:198-9.
14. Terwee CB, Jansma EP, Riphagen II, de Vet HC. Development of a methodological PubMed search filter for finding studies on measurement properties of measurement instruments. *Qual Life Res* 2009;18:1115-23.
15. Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol 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;63:737-45.
16. Terwee CB, Bot SD, de Boer MR, van der Windt DA, Knol DL, Dekker J, et al. Quality criteria were proposed for measurement properties of health status questionnaires. *J Clin Epidemiol* 2007;60:34-42.
17. Mokkink LB, Terwee CB, Knol DL, Stratford PW, Alonso J, Patrick DL, et al. The COSMIN checklist for evaluating the methodological quality of studies on measurement properties: a clarification of its content. *BMC Med Res Methodol* 2010;10:22.
18. Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, et al. The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study. *Qual Life Res* 2010;19:539-49.
19. Terwee CB, Mokkink LB, Knol DL, Ostelo RW, Bouter LM, de Vet HC. Rating the methodological quality in systematic reviews of studies on measurement properties: a scoring system for the COSMIN checklist. *Qual Life Res* 2012;21:651-7.
20. Shim WH, Park HJ, Kim HS, Kim SH, Ko HC, Kim MB, et al. Does the EASI score reflect itch severity? *Ann Allergy Asthma Immunol* 2011;106:540-1.
21. Housman TS, Patel MJ, Camacho F, Feldman SR, Fleischer AB Jr, Balkrishnan R. Use of the Self-Administered Eczema Area and Severity Index by parent caregivers: results of a validation study. *Br J Dermatol* 2002;147:1192-8.
22. Kunz B, Oranje AP, Labreze L, Stalder JF, Ring J, Taieb A. Clinical validation and guidelines for the SCORAD index: consensus report of the European Task Force on Atopic Dermatitis. *Dermatology* 1997;195:10-9.
23. Mazzotti E, Paradisi M, Provini A, Tabolli S. Reproducibility of severity rating for atopic dermatitis. *Acta Derm Venereol* 2008;88:60-1.
24. Mazzotti E, Di Pietro C, Mastroeni S, Provini A, Paradisi M, Tabolli S. Childhood atopic dermatitis. The relationship between parental and dermatologist assessment of disease severity. *Italian Journal Dermatol Venerol* 2005;140:389-95.
25. Muraio Y, Mae H, Konishi T, Gao X, Tanizawa T. Semiquantitative scoring in children under treatment for atopic dermatitis. *Pediatr Int* 2004;46:330-6.
26. Oranje AP, Stalder JF, Taieb A, Tasset C, de Longueville M. Scoring of atopic dermatitis by SCORAD using a training atlas by investigators from different disciplines. ETAC Study Group. *Early Treatment of the Atopic Child. Pediatr Allergy Immunol* 1997;8:28-34.
27. 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;41:149-55.
28. Pucci N, Novembre E, Cammarata MG, Bernardini R, Monaco MG, Calogero C, et al. Scoring atopic dermatitis in infants and young children: distinctive features of the SCORAD index. *Allergy* 2005;60:113-6.
29. Rullo VE, Segato A, Kirsh A, Sole D. Severity scoring of atopic dermatitis: a comparison of two scoring systems. *Allergol Immunopathol (Madr)* 2008;36:205-11.
30. Cosickic A, Skocic F, Colic-Hadzic B, Jahic M. Clinical characteristics and estimation severity of the atopic dermatitis in children. *Med Arh* 2010;64:178-82.
31. Costa C, Rilliet A, Nicolet M, Saurat JH. Scoring atopic dermatitis: the simpler the better? *Acta Derm Venereol* 1989;69:41-5.
32. Severity scoring of atopic dermatitis: the SCORAD index. Consensus Report of the European Task Force on Atopic Dermatitis. *Dermatology* 1993;186:23-31.
33. Berth-Jones J. Six area, six sign atopic dermatitis (SASSAD) severity score: a simple system for monitoring disease activity in atopic dermatitis. *Br J Dermatol* 1996;135(suppl 48):25-30.
34. Carel K, Bratton DL, Miyazawa N, Gyorkos E, Kelsay K, Bender B, et al. The Atopic Dermatitis Quickscore (ADQ): validation of a new parent-administered atopic dermatitis scoring tool. *Ann Allergy Asthma Immunol* 2008;101:500-7.
35. Charman CR, Venn AJ, Williams H. Measuring atopic eczema severity visually: which variables are most important to patients? *Arch Dermatol* 2005;141:1146-51.
36. 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.
37. Angelova-Fischer I, Bauer A, Hippler UC, Petrov I, Kazandjieva J, Bruckner T, et al. The Objective Severity Assessment of Atopic Dermatitis (OSAAD) score: validity, reliability and sensitivity in adult patients with atopic dermatitis. *Br J Dermatol* 2005;153:767-73.
38. Bahmer FA. ADASI score: atopic dermatitis area and severity index. *Acta Derm Venereol Suppl (Stockh)* 1992;176:32-3.
39. Bahmer FA, Schafer J, Schubert HJ. Quantification of the extent and the severity of atopic dermatitis: the ADASI score. *Arch Dermatol* 1991;127:1239-40.
40. Barbier N, Paul C, Luger T, Allen R, De Prost Y, Papp K, 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.

41. Charman CR, Venn AJ, Williams HC. Measurement of body surface area involvement in atopic eczema: an impossible task? *Br J Dermatol* 1999;140:109-11.
42. Charman CR, Venn AJ, Williams HC. Reliability testing of the Six Area, Six Sign Atopic Dermatitis severity score. *Br J Dermatol* 2002;146:1057-60.
43. Charman D, Varigos G, Horne DJ, Oberklaid F. The development of a practical and reliable assessment measure for atopic dermatitis (ADAM). *J Outcome Meas* 1999;3:21-34.
44. Charman DP, Varigos GA. Grades of severity and the validation of an atopic dermatitis assessment measure (ADAM). *J Outcome Meas* 1999;3:162-75.
45. Hachisuka J, Takeuchi S, Kido M, Fukiwake N, Furue M. Severity of disease, rather than xerosis, correlates with pruritus in patients with atopic dermatitis. *Int J Dermatol* 2009;48:374-8.
46. Haecck IM, ten Berge O, van Velsen SG, de Bruin-Weller MS, Bruijnzeel-Koomen CA, Knol MJ. Moderate correlation between quality of life and disease activity in adult patients with atopic dermatitis. *J Eur Acad Dermatol Venereol* 2012;26:236-41.
47. Hanifin JM, Thurston M, Omoto M, Cherill R, Tofte SJ, Graeber M. The eczema area and severity index (EASI): assessment of reliability in atopic dermatitis. EASI Evaluator Group. *Exp Dermatol* 2001;10:11-8.
48. Hon KL, Leung TF, Wong Y, Fok TF. Lesson from performing SCORADs in children with atopic dermatitis: subjective symptoms do not correlate well with disease extent or intensity. *Int J Dermatol* 2006;45:728-30.
49. Schram ME, Spuls PI, Leeftang MM, Lindeboom R, Bos JD, Schmitt J. EASI, (objective) SCORAD and POEM for atopic eczema: responsiveness and minimal clinically important difference. *Allergy* 2012;67:99-106.
50. Silny W, Czarnicka-Operacz M, Silny P. The new scoring system for evaluation of skin inflammation extent and severity in patients with atopic dermatitis. *Acta Dermatovenerol Croat* 2005;13:219-24.
51. Sprickelman AB, Tupker RA, Burgerhof H, Schouten JP, Brand PL, Heymans HS, et al. Severity scoring of atopic dermatitis: a comparison of three scoring systems. *Allergy* 1997;52:944-9.
52. Stalder JF, Barbarot S, Wollenberg A, Holm EA, De RL, Seidenari S, et al. Patient-oriented SCORAD (PO-SCORAD): a new self-assessment scale in atopic dermatitis validated in Europe. *Allergy* 2011;66:1114-21.
53. Sugarman JL, Fluhr JW, Fowler AJ, Bruckner T, Diepgen TL, Williams ML. The objective severity assessment of atopic dermatitis score: an objective measure using permeability barrier function and stratum corneum hydration with computer-assisted estimates for extent of disease. *Arch Dermatol* 2003;139:1417-22.
54. Tofte SJ, Graeber M, Cherill R, Omoto M, Thurston M, Hanifin JM. Eczema area and severity index (EASI): A new tool to evaluate atopic dermatitis. *J Eur Acad Dermatol Venereol* 1998;11(suppl 2):S197.
55. Tremp M, Knafla I, Burg G, Wuthrich B, Schmid-Grendelmeier P. 'EASIdig'—a digital tool to document disease activity in atopic dermatitis. *Dermatology* 2011;223:68-73.
56. van Velsen SG, Knol MJ, Haecck IM, Bruijnzeel-Koomen CA, Pasmans SG. The Self-administered Eczema Area and Severity Index in children with moderate to severe atopic dermatitis: better estimation of AD body surface area than severity. *Pediatr Dermatol* 2010;27:470-5.
57. Verwimp JJM, Bindels JG, Barents M, Heymans HSA. 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):S39.
58. Vourc'h-Jourdain M, Barbarot S, Taieb A, Diepgen T, Ambonati M, Durosier V, et al. Patient-oriented SCORAD: a self-assessment score in atopic dermatitis. A preliminary feasibility study. *Dermatology* 2009;218:246-51.
59. Willemssen MG, van Valburg RW, Dirven-Meijer PC, Oranje AP, van der Wouden JC, Moed H. Determining the severity of atopic dermatitis in children presenting in general practice: an easy and fast method. *Dermatol Res Pract* 2009;2009:357046.
60. Wolkerstorfer A, de Waard van der Spek FB, Glazenburg EJ, Mulder PG, Oranje AP. 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.
61. Yang HJ, Jeon YH, Pyun BY. Evaluation of patient's subjective severity using various scoring system in Korean children with atopic dermatitis. *Asian Pac J Allergy Immunol* 2010;28:130-5.
62. Schäfer T, Dockery D, Kramer U, Behrendt H, Ring J. Experiences with the severity scoring of atopic dermatitis in a population of German pre-school children. *Br J Dermatol* 1997;137:558-62.
63. Weisshaar E, Diepgen TL, Bruckner T, Fartasch M, Kupfer J, Lob-Corzilius T, et al. Itch intensity evaluated in the German Atopic Dermatitis Intervention Study (GADIS): correlations with quality of life, coping behaviour and SCORAD severity in 823 children. *Acta Derm Venereol* 2008;88:234-9.
64. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *Int J Surg* 2010;8:336-41.
65. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977;3:174.
66. de Vet HC, Terwee CB, Mokkink LB, Knol DL. Measurement in medicine—a practical guide. New York: Cambridge University Press; 2011.