

The eczema area and severity index (EASI): assessment of reliability in atopic dermatitis

Hanifin JM, Thurston M, Omoto M, Cherill R, Tofte SJ, Graeber M, the EASI Evaluator Group. The eczema area and severity index (EASI): assessment of reliability in atopic dermatitis

Exp Dermatol 2001; 10: 11–18. © Munksgaard, 2001

Abstract: *Objective* – To test the reliability of the eczema area and severity index (EASI) scoring system by assessing inter- and intra-observer consistency. *Design:* Training of evaluators, application, and assessment over 2 consecutive days. *Setting* – An academic center. *Patients* – Twenty adults and children with atopic dermatitis (AD); cohort 1 (10 patients ≥ 8 years) and cohort 2 (10 patients < 8 years). *Interventions* – None. *Main outcome measure* – The EASI was used by 15 dermatologist evaluators to assess atopic dermatitis in cohort 1 and cohort 2 on 2 consecutive days. Inter- and intraobserver reliability were analyzed. *Results* – Overall intra-evaluator reliability of the EASI was in the fair-to-good range. Inter-evaluator reliability analyses indicated that the evaluators assessed the patients consistently across both study days. *Conclusions* – This study demonstrated that the EASI can be learned quickly and utilized reliably in the assessment of severity and extent of AD. There was consistency among the evaluators between consecutive days of evaluation. These results support the use of the EASI in clinical trials of therapeutic agents for AD.

J. M. Hanifin¹, M. Thurston², M. Omoto^{1,3}, R. Cherill², S. J. Tofte¹, M. Graeber² and the EASI Evaluator Group⁴

¹Department of Dermatology, Oregon Health Sciences University, Portland, Oregon; ²Novartis Pharmaceuticals Corporation; ³Shiga University of Medical Science, Shiga, Japan; ⁴The EASI Evaluator Group is alphabetically listed in Acknowledgments

Key words: atopic dermatitis – eczema – psoriasis

Jon M. Hanifin, Professor of Dermatology, Oregon Health Sciences University, 3181 S.W. Sam Jackson Park Road, Portland, OR 97201–3098, USA
Tel.: 503 494 5603. Fax: 503 494 6907
e-mail: hanifinj@ohsu.edu

Accepted for publication 20 November 2000

Accurate assessment of the extent and severity of atopic dermatitis (AD) is essential for quantitating 1) the baseline clinical disease burden and 2) the effectiveness of a treatment regimen being tested (1, 2). The extent of AD is usually determined by examining the patient's skin and estimating the percentage involvement of affected areas, while severity is determined by grading specific signs of eczema (e.g. erythema, induration/edema/papulation, excoriation, lichenification, scaling, and oozing/weeping/crusting), and by eliciting the symptomatic intensity of pruritus. Estimation of changes in these parameters during successive examinations can be subjective, inaccurate, and difficult to standardize.

Over the past 2 decades, there have been considerable efforts to standardize AD assessment and alleviate problems associated with semi-quantitative evaluations. During this time, several tools have been designed for the evaluation of AD, ranging from simple scoring of localized lesional areas (2) to severity and extent measures (3–5) to more complex and comprehensive systems that have not been widely embraced (6, 7). The scoring of atopic der-

matitis (SCORAD) method has been validated, but it mixes observations of signs and the assessment of symptoms, and is weighted toward pediatric populations (7). To date, none of these methods has been uniformly accepted for the reasons listed above and because measures and recordings are not optimally efficient for quick, yet reliable assessment of patients in busy clinic settings.

Investigators would welcome development of a validated, reliable system that is both easy to use and reflective of key morphologic features of AD. The ideal tool for assessing the extent and severity of AD should be efficient and simple, yet comprehensive, consistent, and flexible for a range of patients (i.e., those with mild to severe disease, both pediatric and adult). In addition, it must be effective in assessing therapeutic response to both new and existing therapies.

With these attributes in mind, the eczema area and severity index (EASI) was developed. This instrument was designed by modifying the general scheme used in the psoriasis area and severity index (PASI) (8). The PASI has been widely used and established as a well-accepted and standardized in-

strument for assessing therapeutic response in patients with psoriasis. The PASI system incorporates the principle that disease intensity and the total affected body area are crucial components of assessment for chronic, widespread inflammatory skin disease (8). While the PASI cannot be used for AD because of distinctly different sign severity parameters, the principle of integrating disease extent and sign severity to describe disease has led to the definition of the eczema area and severity index (EASI).

Data from this study support EASI as an effective, easily understood, and simple tool for the evaluation of AD. The individual components of EASI (i.e. body region involvement, severity) can be separated and evaluated independently or in combination to provide a more complete assessment of the patient. In addition, the EASI is comprehensive, evaluating four natural anatomical regions for both severity and extent of key disease signs. The EASI is adaptable to children, the group most commonly affected and whose body proportions have not yet reached those of adults. It focuses on the key acute and chronic signs of inflammation (i.e. erythema, induration/papulation, excoriation, and lichenification). EASI excludes non-key signs such as xerosis and scaling, oozing and crusting, and subjective parameters such as pruritus (7) and sleep loss in order to focus the index on key disease signs and to avoid mixing objective parameters with subjective symptoms.

The goal of the present study was to assess the usefulness and reliability of the EASI in a group of variably affected subjects with AD on 2 successive days.

Materials and methods

Evaluators

A total of 15 evaluators were selected as a representative sample of the population of clinicians who may ultimately use the EASI in clinical practice. Evaluators were dermatologists from the United States with experience in assessing the signs and symptoms of AD in both adults and children. The evaluators came from diverse practice backgrounds, including university clinics, private practice, pediatric dermatology clinics, and health maintenance organizations. The principal investigator did not participate in the EASI evaluations on days 1 and 2.

The same panel of evaluators assessed both the adult and pediatric subjects in the study. All evaluators received the study protocol and EASI evaluation forms 1 week prior to the validation study. On day 1 of the study, the evaluators received 30 min of training with the opportunity to discuss the

evaluation method prior to assessing the patients. The training included a 10-min practice evaluation conducted on 2 patients (1 pediatric and 1 adult) who were not included in the data analysis. The evaluators were instructed to be consistent in their evaluations and to not discuss the evaluations with other participants.

Patients

Patients were recruited from the Oregon Health Sciences University (OHSU) Atopic Dermatitis Clinic population by the principal investigator. Patients enrolled in the study met the following inclusion criteria: 1) Diagnosis of AD according to the criteria of Hanifin & Rajka (9); 2) a range of disease activity from mild to severe AD according to the criteria of Rajka & Langeland (10); and 3) all signed an informed consent approved by OHSU's institutional review board (IRB). The exclusion criteria were: 1) Patients who recently used tar or dyes for treating AD, because this might interfere with the evaluation; 2) patients who were likely to use ultrapotent topical steroids in between day 1 and day 2 of the study, because this could interfere with the consistency of scoring on these 2 days; and 3) any other condition (e.g. immunodeficiency) which, in the opinion of the principal investigator, would render the patient ineligible for the study.

Two cohorts of patients (10 patients per cohort) were evaluated in this study: Cohort 1 consisted of patients aged 8 years and above (mean 43 years), while cohort 2 consisted of patients aged 0 through 7 years (mean 4.3 years). Each cohort of patients was selected by the principal investigator to achieve a broad range of disease severity and body region involvement.

The patients were instructed not to apply any creams or ointments to their dermatitis within 2 h of coming to the clinic on days 1 and 2 of the study. Patients were not permitted to begin any new therapy for their dermatitis during participation in this 2-day study.

On day 0 (screening), all patients were evaluated to assess extent and severity of their dermatitis. On day 1 of the study, patients in cohort 1 were randomly (i.e. in terms of disease severity) assembled into 10 different examination rooms. A clinical assistant was stationed in each room to record the evaluator's observations on to the case report form. The end of each evaluation session was signaled, at which time the evaluators advanced to their next assigned examination room. Evaluators were asked to check their evaluations recorded on the case report form for any errors before continuing on to the next patient. Each evaluator assessed each patient only once on a single day. Evaluators

did not discuss their assessments with the patients or with one another. The patients in cohort 2 were assessed immediately after cohort 1. Patients in cohort 2 were evaluated in the same manner as those in cohort 1. On day 2, both cohorts 1 and 2 were randomly assigned to exam rooms and then assessed using a re-randomized evaluation schedule to determine the day-to-day consistency of scoring.

Eczema area and severity index (EASI)

Regional body surface area tabulation was used to assess the severity of dermatitis over four body areas. In cohort 1 (older patient group), the head and neck (H), upper extremities (U), trunk (T), and lower extremities (L) were assigned proportionate body surface areas of 10% (H), 20% (U), 30% (T), and 40% (L), roughly consistent with the “rule of nines” (Table 1). For young children in cohort 2, a modification was used such that the proportionate body surface areas were 20% (H), 20% (U), 30% (T), and 30% (L) (8). The area within each body region with the key signs of inflammation (as defined below) was estimated as a percentage of the total area of that particular body region and recorded on the case report form. Symptoms (e.g. pruritus), along with secondary signs (e.g. xerosis, scaling) were excluded from the area assessments. This percentage of area involved for each of the four body regions was assigned a proportional score from 0 to 6 during the analysis: 0=no eruption; 1=<10%; 2=<10%–29%; 3=<30%–49%; 4=<50%–69%; 5=<70%–89%; and 6=>90%–100%.

Each of the four body regions was assessed separately for the key signs of erythema, induration/papulation/edema, excoriations, and lichenification. The average degree of severity of each sign in each of the four body regions was assigned a score of 0 to 3 (none, mild, moderate, and severe, respectively) with half-steps allowed. It should be

noted that 1) the buttocks and feet were counted as part of the lower extremities; 2) the internal axillae and groin were counted as part of the trunk; and 3) the external axillae and hands were counted as part of the upper extremities. The total body score for each body region was obtained by multiplying the sum of the severity scores of the four key signs by the area score, then multiplying the result by the constant weighted value assigned to that body region (Table 1). The sum of these scores gives the EASI total, which could range from 0 to a maximum 72.

Statistical analysis

The objective of this study was to determine whether a group of evaluators would be consistent in measurement on a given day (inter-evaluator reproducibility) and whether evaluators would be consistent over more than one occasion (intra-evaluator reproducibility) for assessing extent and severity parameters in AD patients using the EASI. Inferential analyses focused on the primary goal of the study, inter-evaluator reproducibility, and were conducted in two parts. First, a primary analysis was conducted using a random effects model to investigate variability in the overall EASI parameter (11). The correlation coefficient of reliability, also referred to as simply, reliability, and denoted as r -hat, measures the proportion of overall variability explained by subject-to-subject variability. Values of r -hat between 0.4 and 0.75 are interpreted to constitute fair-to-good reliability; values greater than 0.75 indicate excellent reliability (11). In addition, a secondary analysis was conducted to investigate the reliability of assessment in each of the key signs (erythema, infiltration/papulation, excoriations, lichenification) of the EASI using extensions of the kappa statistic (11, 12). These were interpreted in a similar fashion to the r -hat statistic. The kappa coefficient equals unity when there is complete agreement of the raters. When the observed agreement exceeds chance agreement, kappa is positive, with its magnitude reflecting the strength of agreement.

Intra-evaluator reliability was to be addressed descriptively. Simple linear regression analysis of Day 1 score against Day 2 score provided a descriptive overview. Additional mixed effects models, with day as a fixed effect, and random effects for patient, evaluator and interactions were performed on the EASI response.

Results

Evaluation of the data indicated good overall evidence of reliability in rating AD subjects with

Table 1. Eczema area and severity index: calculation for patients 8 years of age and older¹

Body region	EASI Score ^{2,3}
Head/Neck (H)	$(E+I+Ex+L) \times \text{Area} \times 0.1$
Upper limbs (UL)	$(E+I+Ex+L) \times \text{Area} \times 0.2$
Trunk (T)	$(E+I+Ex+L) \times \text{Area} \times 0.3$
Lower limbs (LL)	$(E+I+Ex+L) \times \text{Area} \times 0.4$
EASI =	<i>Sum of the above 4 body region scores</i>

¹For children aged 0–7 years, proportionate areas were head/neck, 20%; upper limbs, 20%; trunk, 30%; and lower limbs, 30%.

²E=Erythema, I=induration/papulation, Ex=excoriation, L=lichenification.

³Where area is defined on a 7-point ordinal scale: 0=no eruption; 1=<10%; 2=<10%–29%; 3=<30%–49%; 4=<50%–69%; 5=<70%–89%; and 6=>90%–100%.

EASI. Scoring was consistent on each day of the study as shown by the trends in patient scores (Fig. 1).

Inter-evaluator reliability

EASI evaluation demonstrated reasonably good overall inter-evaluator reliability (Fig. 2). The r -hat across the two cohorts on day 1 of 0.71, increasing to 0.76 on day 2, supports the overall conclusion that EASI provides a reliable tool in assessing atopic dermatitis. Of the 15 evaluators in the

study, one evaluator was an outlier with a median score of 28.3, which was approximately 10 units higher than any other evaluator. With exclusion of this evaluator from the r -hat analysis, all reliabilities were in the good-to-excellent range (Fig. 3).

When analyzed according to body region, all severity and extent inter-evaluator reliabilities were in either the fair-to-good or the excellent ranges (Fig. 4), with the one exception of the extent assessment on the trunk of patients in cohort 2.

The kappa statistic (intra-class correlation coef-

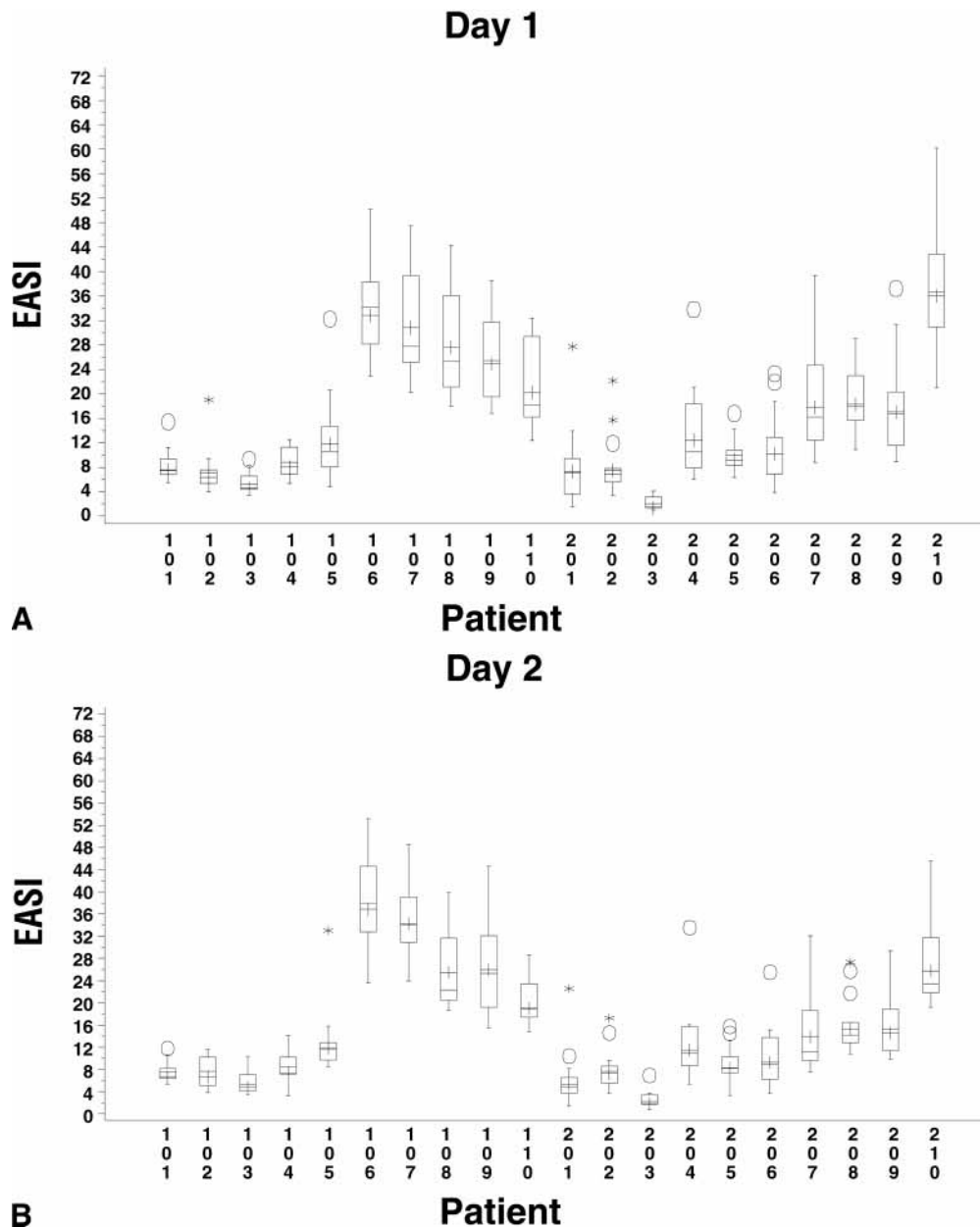


Figure 1. Summary of trends in patient scoring following EASI evaluation demonstrated consistency among the evaluators on both day 1 (A) and day 2 (B). The box plot at each day represents the median (horizontal line through box, mean (+), interquartile range, maximum and minimum values within 1.5 interquartile ranges, outliers (O) and extreme outliers (*)

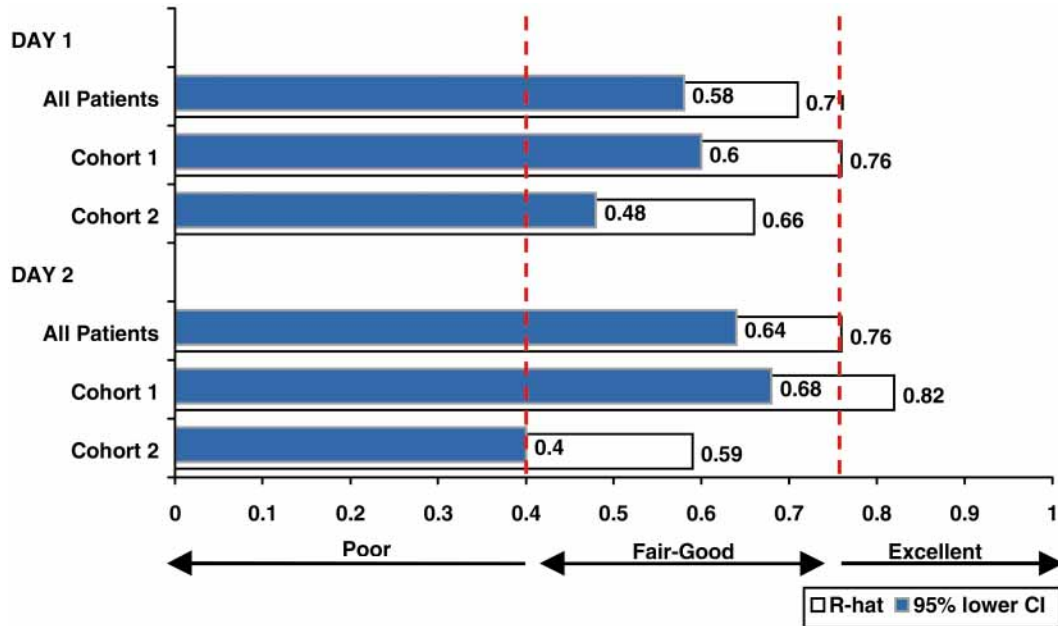


Figure 2. EASI evaluation demonstrated fair-to-good overall inter-evaluator reliability

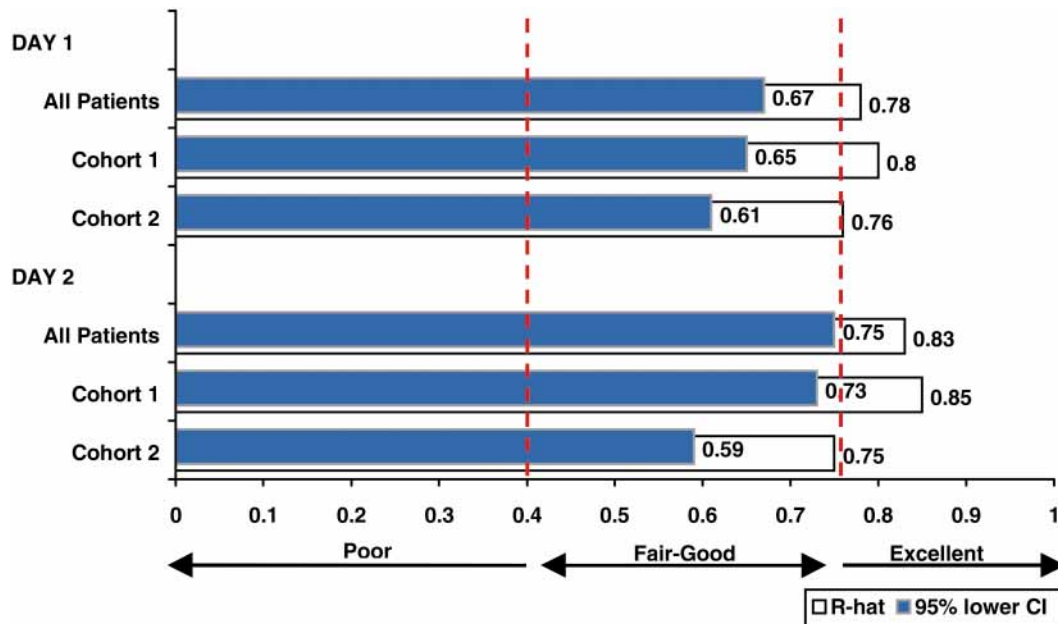


Figure 3. The overall reliability of the EASI increased with the omission of evaluator number 10

ficient) was employed to determine which of the individual key signs of AD most detracted from the reliability of the severity grade (Fig. 5). An overall weighted kappa statistic across all body regions showed fair-to-good intra-rater agreement. The data suggest that induration/papulation was the most difficult sign to assess, since it ranked lowest of the kappa statistics. In contrast, erythema and excoriations were assessed with good

reliability and lichenification was assessed with reasonable reliability. Generally, any changes in individual subject scores between day 1 and day 2 were within a score of 0.5 for severity.

Intra-evaluator reliability

Using regression analyses, complete agreement from day 1 to day 2 for all evaluators would give a

regression coefficient of 1.0. The regression coefficient for cohort 1 was 0.996. For cohort 2, the regression coefficient was lower at 0.657. This was primarily due to patient 210, who was scored lower by

13 of 15 evaluators. This gave a mean score reduction from 36 on day 1 to 28 on day 2, but there was a general slight trend to lower scores on day 2, possibly reflecting a change in subjects' disease.

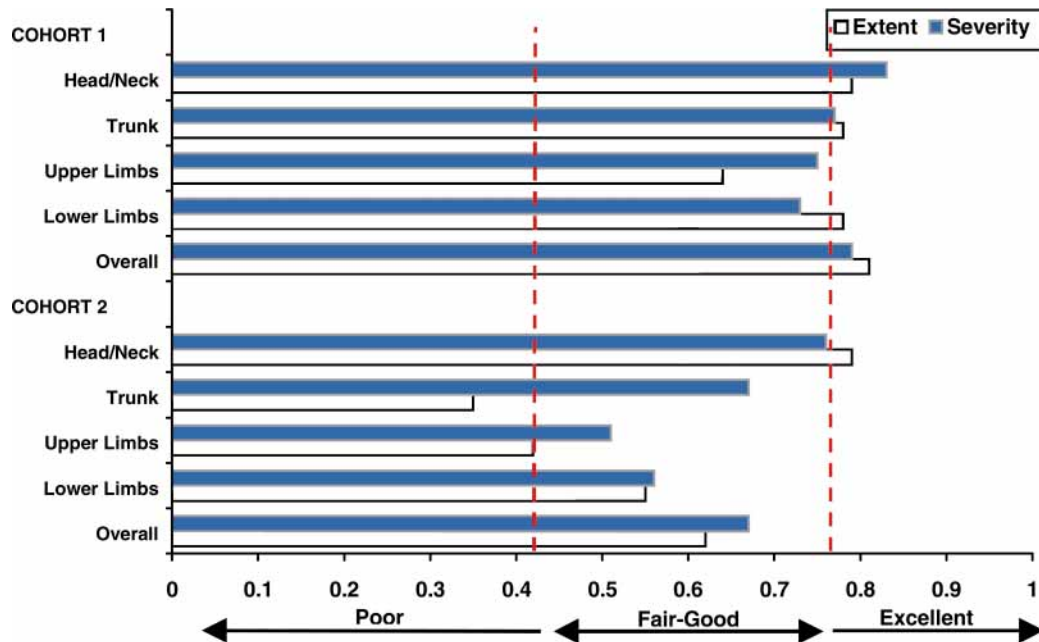


Figure 4. EASI evaluation demonstrated fair-to-good or excellent overall inter-evaluator reliability when analyzed by body region. All severity and extent reliabilities were in either the fair-to-good or the excellent range with the exception of the assessment of the extent on the trunk of patients in cohort 2, which was poor

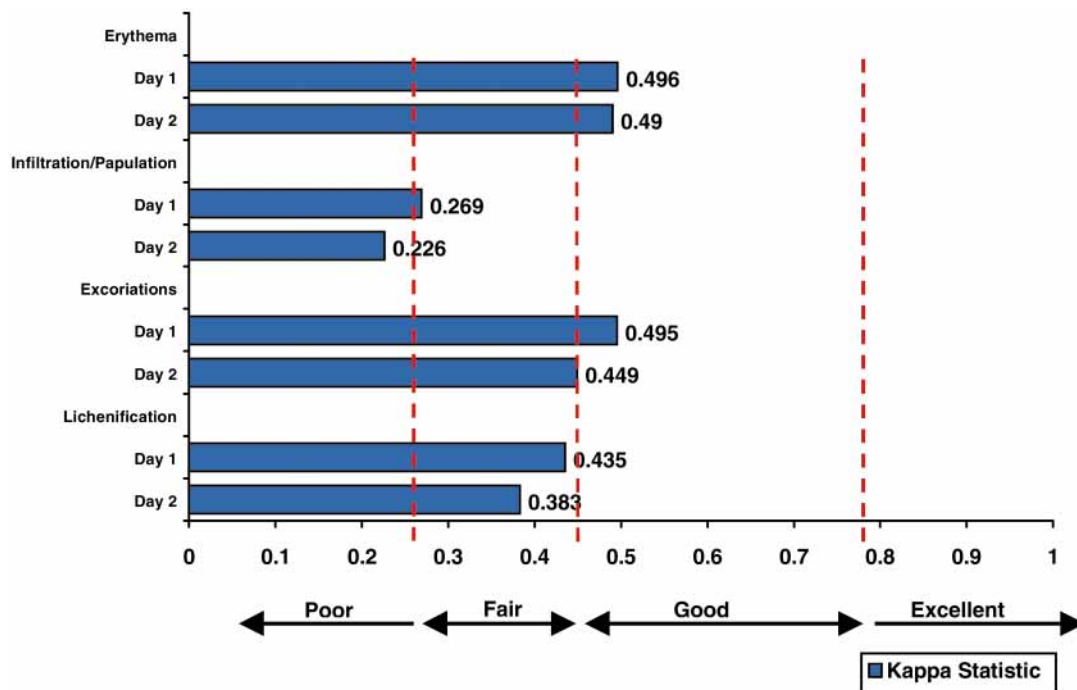


Figure 5. EASI evaluation of individual signs of atopic dermatitis (AD): infiltration/papulation showed the least agreement and erythema and excoriations showed the most agreement among the evaluators

The primary analysis concentrated on the individual study days, noting the consistent findings in both of these sessions. Investigating the effect of the study day further, mixed effects models, with day as a fixed effect, showed no overall effect ($P=0.22$). There was no evidence of a significant effect of day for cohort 1, although there was some evidence for cohort 2 ($P=0.042$). This supports the summary finding made above.

It should be noted that evaluator #10 scored consistently between day 1 and day 2 of the study; while not scoring patients on the EASI as other evaluators did, this evaluator was able to consistently use the tool.

Discussion

Physicians who treat patients diagnosed as having AD recognize that there is a strong need for a standardized grading system that could be used during patient treatment or clinical investigations. Several tools have been developed and evaluated, but none has been consistently accepted and globally utilized to date (2–7). The appropriate tool for assessing AD would be comprehensive yet easy to implement in a clinical setting.

The EASI is an easily understood and simple system that can be used by practitioners and investigators to standardize the baseline evaluation of AD and track changes in AD over time. It was designed by modifying the general scheme used in the PASI scoring system, which has been utilized effectively in the assessment of psoriasis (8, 13). EASI assesses the key signs of eczema over the four natural anatomic divisions of the body – the head and neck, the trunk and the upper and lower extremities. Estimation and recording of only four areas greatly simplifies assessment compared to systems which require “rule of nines”, or scoring of six areas (4, 5, 7).

The present EASI study demonstrated that this instrument can be used reliably in the assessment of severity and extent of AD. Importantly, individual evaluators were consistent in their assessments between day 1 and day 2. Changes in individual subject scores between day 1 and day 2 were generally within a range of one-half of a point. This was a key finding, given that consistency over time is critical for accurate assessment of therapeutic interventions. Our study also highlights some problems that can be encountered in the assessment of younger patients. Scores were generally lower for cohort 2 on the second day. This may reflect the children’s adaptation and reduced anxiety in the study environment. Assessments of children demand an efficient instrument that can be used rapidly and reliably.

Variations in the evaluators’ assessments of the components of EASI suggest greater accuracy might be obtained if the same evaluator assesses a given patient throughout the course of treatment. It is likely that the overall reliability and utility of the EASI can be improved with better training in the assessment of each of the components of EASI. Use of this system over time should also increase reliability.

The fact that induration/papulation was the most difficult sign to assess was not surprising, given that these are the clinical expressions of cellular infiltration and edema and require both visualization and palpation. While papulation and edema can be easily discerned by side lighting, induration requires palpation and can be especially difficult to distinguish in areas of lichenification (7). These parameters are included because they are key signs of AD and of considerable interest when assessing the completeness of a therapeutic response. Recent studies have validated the efficacy of evaluating all the key signs that parallel each other in a statistically significant manner during response to therapy (14).

The results of this study indicate that the EASI provides a comprehensive yet efficient index for assessing AD and for sequentially measuring responses to treatments utilized for AD. Since the EASI is a composite score, it allows for individual components to be assessed both separately or in combination, which provides a more complete evaluation of the patient’s condition. The utility of EASI in detecting changes in disease severity over treatment course, and its potential effectiveness in differentiating between treatment potencies has now been assessed in several clinical trial settings (14–17). These reports have shown statistical agreement between the EASI and various other grading systems. The EASI can be used with confidence in future clinical trials of agents for AD treatment.

Acknowledgements

The authors gratefully acknowledge the substantial contributions made to this work by the Eczema Area and Severity Index (EASI) Evaluator Group:

Michael Adler, MD, Portland, Oregon; Carolyn Hale, MD, Beaverton, Oregon; Adelaide Hebert, MD, Houston, Texas; Robin Hornung, MD, Chicago, Illinois; Michael Jarratt, MD, Austin, Texas; Sewon Kang, MD, Ann Arbor, Michigan; H. Irving Katz, MD, Fridley, Minnesota; Mary Lyons, MD, Beaverton, Oregon; Linda Rabinowitz, MD, Milwaukee, Wisconsin; Farah Shah, MD, Lubbock, Texas; Robert Sidbury, MD, Portland, Oregon; Seth Stevens, MD, Cleveland, Ohio; Michael Tharp, MD, Chicago, Illinois; Annette Wagner, MD, Chicago, Illinois; Kenneth Washenik, MD, PhD; New York, New York.

References

1. Abrams B B. Atopic dermatitis: elements in clinical study design and analysis. *Acta Derm Venereol* (Stockh) 1989; 144 (suppl): 15–19.

2. Hanifin J M. Standardized grading of subjects for clinical research studies in atopic dermatitis: workshop report. *Acta Derm Venereol* (Stockh) 1989; 144 (suppl): 28–30.
3. Seymour J L, Keswick B H, Hanifin J M, Jordan W P, Milligan M C. Clinical effects of diaper types on the skin of normal infants and infants with atopic dermatitis. *J Am Acad Dermatol* 1987; 17: 988–997.
4. Costa C, Rilliet A, Nicolet M, Saurat J H. Scoring atopic dermatitis: the simpler the better? *Acta Derm Venereol* (Stockh) 1989; 69: 41–45.
5. Sowden J M, Berth-Jones J, Ross J S et al. Double-blind, controlled, crossover study of cyclosporin in adults with severe refractory atopic dermatitis. *Lancet* 1991; 338: 137–140.
6. Bahmer F A. Quantification of the extent and the severity of atopic dermatitis: the ADASI score. *Arch Dermatol* 1991; 127: 1239–1240.
7. Stalder J F, Atherton D J, Bieber T et al. (Consensus Report of the European Task Force on Atopic Dermatitis). Severity scoring of atopic dermatitis: the SCORAD index. *Dermatology* 1993; 186: 23–31.
8. Fredriksson T, Pettersson U. Severe psoriasis – oral therapy with a new retinoid. *Dermatologica* 1978; 157: 238–244.
9. Hanifin J M, Rajka G. Diagnostic features of atopic dermatitis. *Acta Derm Venereol* (Stockh) 1980; 92 (suppl): 44–47.
10. Rajka G, Langeland T. Grading of the severity of atopic dermatitis. *Acta Derm Venereol* (Stockh) 1989; 144 (suppl): 13–14.
11. Fleiss J L. Reliability of measurement. *The Design and Analysis of Clinical Experiments*. In: New York, NY: John Wiley & Sons, 1986: 1–32.
12. Fleiss J L, Cohen J, Everitt B S. Large sample standard errors of kappa and weighted kappa. *Psychol Bull* 69: 72: 323–327.
13. Fleischer A B Jr., Feldman S R, Rapp S R et al. Disease severity measures in a population of psoriasis patients: the symptoms of psoriasis correlate with self-administered psoriasis area severity index scores. *J Invest Dermatol* 1996; 107: 26–29.
14. Hanifin J M, Ling M R, Langley R et al. Tacrolimus ointment for the treatment of atopic dermatitis in adult patients: Part I, Efficacy. *J Am Acad Dermatol* 2001; 44: (in press).
15. Luger T, Van Leent E J M, Graeber M et al. SDZ ASM 981: An emerging safe and effective treatment for atopic dermatitis. (in press).
16. Kang S, Lucky A, Pariser D, Lawrence I, Hanifin J, and the Tacrolimus Ointment Study Group. A long-term, non-comparative trial to evaluate the safety of topically applied tacrolimus as treatment of atopic dermatitis in pediatric patients. *J Am Acad Dermatol* 2001; 44: (in press).
17. Paller A, Eichenfield L F, Leung D et al. Topically applied tacrolimus ointment vs. vehicle ointment in pediatric patients with atopic dermatitis. *J Am Acad Dermatol* 2001; 44: (in press).