JAPANESE DERMATOLOGICAL ASSOCIATION
THE JOURNAL OF DERMATOLOGY
doi: 10.1111/j.1346-8138.2011.01285.x
Journal of Dermatology 2011; **38**: 625–631

SPECIAL REPORT

English version of the interim report published in 1998 by the members of the Advisory Committee on Atopic Dermatitis Severity Classification Criteria of the Japanese Dermatological Association

Hikotaro YOSHIDA,[†] Toshiyuki AOKI,[‡] Masutaka FURUE, Hachiro TAGAMI, Fumio KANEKO, Fujio OHTSUKA, Kiyoshi NISHIOKA, Kiyoshi TODA, Masako MIZOGUCHI, Hideo NAKAYAMA, Zenro IKEZAWA, Masahiro TAKIGAWA, Jiro ARATA, Shoso YAMAMOTO, Yoichi TANAKA, Masahide ISHIGAKI,[§] Tadashi KUSUNOKI,[§] Kunihiko YOSHIKAWA

Advisory Committee on Atopic Dermatitis Severity Classification Criteria of Japanese Dermatological Association

ABSTRACT

The Japanese Dermatological Association established an advisory committee in 1995 to set up severity scoring systems for atopic dermatitis (AD). Its interim report was published in Japanese in the *Japanese Journal of Dermatology* (**108**: 1491–1496, 1998) by Chairman Hikotaro Yoshida. Because of the strong demand for an English version, we have decided to publish the report in English. This prospective study was designed to evaluate the status of 259 AD patients using Method 1, which involves a simple global evaluation of disease severity; Method 2, which involves global evaluation by summing severity scores obtained from five body regions (i.e. the head and neck, anterior and posterior trunks, and upper and lower limbs); Method 3, which consists of both assessment of the extent of involved areas at each of the five body regions and that of the severity scores of each eruption component observed in the most severely affected body region; and Method 4, which consists of the evaluation of only subjective components (daytime pruritus and sleep disturbance). Employing the results obtained with Method 1 as a tentative benchmark, we analyzed its correlation with those of Methods 2, 3 and 4 to statistically assess the validity and reliability of these methods. Method 2, Method 3 and the portion of Method 4 involving evaluation of only the subjective symptom of daytime pruritus but not the sleep disturbance were considered useful in evaluating AD severity.

Key words: atopic dermatitis, pruritus, severity classification.

INTRODUCTION

In 1995, the Japanese Dermatological Association created an advisory committee to establish severity classification criteria for atopic dermatitis (AD). When this committee was founded, there were two classification methods used across the world, namely, a classification method prepared by Costa *et al.*,[1] and a Severity Scoring of Atopic Dermatitis (SCORAD) index that was prepared by AD specialists from nine European countries.[2] However, some researchers identified weaknesses in these methods. The weakness of the former was that the process of developing evaluation criteria had not been revealed

Correspondence: Masutaka Furue, M.D., Ph.D., Department of Dermatology, Kyushu University, 3-1-1 Maidashi, Higashiku, Fukuoka 812-8582, Japan. Email: furue@dermatol.med.kyushu-u.ac.jp
[†]Chair, [‡]Vice Chair and [§]adviser for statistical analysis.
Received 5 February 2011; accepted 6 April 2011.

and that of the latter was that the process of defining a representative site for each eruption component was too complicated to conduct at busy outpatient clinics. Accordingly, it was our intention to establish an original Japanese severity classification for AD that would cover these weak points as much as possible.

METHODS

First, the committee laid out the following principles: (i) skin eruption is the most important factor in establishing severity criteria; and (ii) such severity classification criteria should be established on a statistical basis.

An evaluation sheet based on these principles was prepared for each patient (Figs 1,2). Each patient was evaluated by using the following four different methods.

Method 1: Global evaluation

After examining the patient's entire body, scoring was conducted using the following system: 1, very mild; 2, mild; 3, moderate; 4, severe; and 5, very severe. Severity of the lesions as well as the extent of involved body surface was considered (Fig. 1).

Method 2

The entire body surface was divided into five body regions as shown in Figure 1, and evaluation of disease severity was performed for each region in a manner similar to global evaluation from 0 (none) to 4 (very severe) to calculate the total score (Fig. 1).

Method 3

Comprehensive evaluation was based on the extent of AD lesions and each eruption component; this method is a modified version of the SCORAD index. The extent of involved area was assessed by determining the extent of AD lesions in each of the five body regions in the following fashion: 0, none; 1, 0–1/3; 2, 1/3–2/3; 3, 2/3 to less than the entire surface; and 4, entire surface. By selecting the most severely affected body regions, severity scoring was conducted for each eruption component of AD lesions, namely, erythema, papule, erosion, crust, excoriation, lichenification, prurigo, depilation, scale, pigmentation and dry skin, on a 5-grade system (0, none; 1, mild; 2, moderate; 3, severe; and 4, very severe). Committee members agreed that the eruption components scale, dry skin and pigmentation should be excluded from evaluation of the extent of

Method 1 Global evaluation
1, very mild; 2, mild; 3, moderate; 4, severe; 5, very severe.
Very severe: As active lesions are found nearly all over the body, this condition is considered particularly serious among severe cases.
Very mild: AD condition is generally very mild, or mild changes are only partially observed. This condition rarely requires therapy, or can be easily treated with minimum treatment. Thus, it cannot be diagnosed as AD without additional data or information.

Method 2 Severity scoring based on body locations (combined evaluation of severity and extent of involved area)
(Severity and extent of involved area should be evaluated. However, evaluation of the area should include only the eight lesion components mentioned in Method 3 (Fig. 2), excluding scale, dry skin, and pigmentation.)
0, none; 1, mild; 2, moderate; 3, severe; 4, very severe

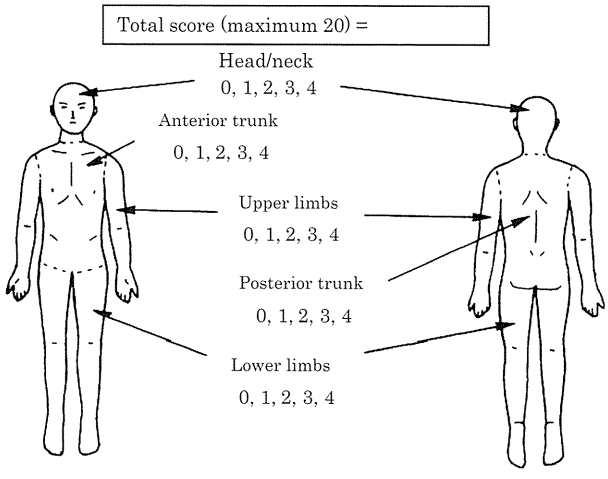


Figure 1. Evaluation Methods 1 and 2. AD, atopic dermatitis.

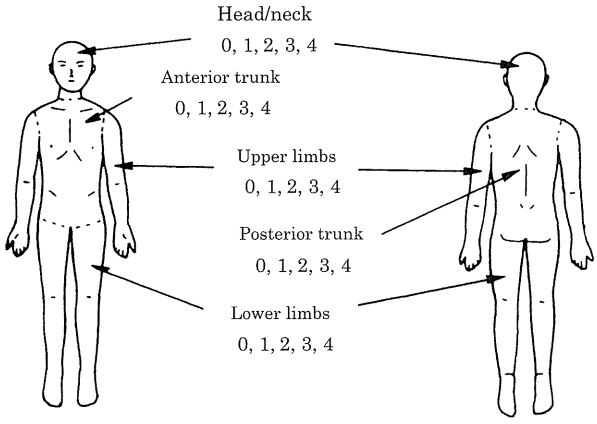
Method 3 Method for comprehensively evaluating the eruption components and areas of involvement
Please assess the (1) eruption component (in the most severely affected body part) and (2) areas of involvement separately.

(1) Eruption component

Erythema	0,	1,	2,	3,	4
Papule	0,	1,	2,	3,	4
Erosion	0,	1,	2,	3,	4
Crust	0,	1,	2,	3,	4
Excoriation	0,	1,	2,	3,	4
Lichenification	0,	1,	2,	3,	4
Prurigo	0,	1,	2,	3,	4
Depilation	0,	1,	2,	3,	4
Scale	0,	1,	2,	3,	4
Pigmentation	0,	1,	2,	3,	4
Dry skin	0,	1,	2,	3,	4

Evaluation criteria
0, none; 1, mild; 2, moderate; 3, severe; 4, very severe.
Depilation, caused by scratching (excluding alopecia).

(2) Area
The areas with eight eruption components as discussed in Method 3 (1) should be determined. Please exclude scale, pigmentation and dry skin.



Evaluation criteria
0, none; 1, less than 1/3; 2, 1/3–2/3; 3, 2/3 to less than the entire surface; 4, entire surface.

Method 4 Evaluation based on subjective symptoms

Daytime pruritus	0,	1,	2,	3,	4
Sleep disturbance	0,	1,	2,	3,	4

Evaluation criteria
0, none; 1, weak; 2, slightly strong; 3, strong; 4, very strong.

Figure 2. Evaluation Methods 3 and 4.

AD lesions because they are considered secondary changes produced in the final stage of AD lesions (Fig. 2).

Method 4: Evaluation based on subjective symptoms

Questions were separately asked concerning daytime pruritus and sleep disturbances at night according to the 5-step evaluation system. However, this evaluation was used only for reference to emphasize the objective evaluation of observable skin changes (Fig. 2).

To ensure the reliability of these AD severity evaluation methods focusing on skin lesions (Methods 1, 2 and 3), interexaminer agreement (including consistency) was also required with regard to each evaluated eruption component. In this study, each patient was evaluated by five observers from the same institution. Interexaminer agreement was also determined by calculating the κ coefficient as described by Davis and Fleiss[3] because its average can be used as an index of reliability of several observers. In cases of complete agreement, the κ coefficient is 1 (maximum), whereas it is 0 (minimum) in cases of non-agreement. When the level of agreement is lower than the expected non-agreement level, a negative value is obtained. Although no reliable reports have been published regarding the score for significant agreement, one report defined a score of 0.4 or higher as "fair agreement" and that of 0.75 or higher as "excellent agreement".[4]

We were also required to evaluate the total scores in order to assess reliability. For this purpose, the intraclass correlation coefficient (ICC)[5] was used.

ICC, an index for the reliability of continuous data evaluation, is defined by the ratio of the measured value dispersion to true value dispersion. If there is no error, the ICC is set at the maximum value of 1, whereas ICC approaches the minimal value of 0 when the error increases.

To prove the acceptability of the evaluation methods for AD severity classification criteria, data validity should also be included for scientific justification. However, there is no established standard for AD severity classification. Thus, the committee decided to prove the validity of these methods by conducting an exploratory analysis of the extent of correlation between Method 1 and Method 2, or that between Method 1 and Method 3. Components showing high correlation were considered important for severity classification.

Because ease of use is important in decreasing the burden on physicians and patients, feasibility of the evaluation method was assessed using the rate of missing data, mistaken entries and impressions stated by the examiners.

Patients that participated in the present study were those treated in the dermatology clinics of 10 institutions to which each committee member belonged. The method validity was assessed for 259 patients examined at these 10 institutions, whereas the reliability was evaluated for 95 patients from five institutions. Thus, some patients were included in both reliability and validity analyses.

RESULTS

Background data of the 259 subjects who participated in the validity study and those of the 95 subjects evaluated in the reliability study are shown in Table 1. No particular bias was observed between these patient groups with regard to their sex, age and AD duration. They were classified according to severity scores obtained from global evaluation (Method 1) (Table 2).

To determine the correlation between Method 1 and Method 2, the sum of the scores of the five body regions in Method 2 was calculated, and a scatter diagram was prepared to compare the results obtained from Method 1 (Fig. 3). The obtained correlation coefficient between these methods (γ) was 0.88, indicating that the results obtained using Method 2 showed a strong correlation with those of Method 1.

The correlation coefficient (γ) was 0.76 between the results of global evaluation of Method 1 and the sum of eruption component scores assessed using Method 3 (Fig. 4), whereas it was 0.81 between those of global evaluation in Method 1 and total involvement area evaluated using Method 3 (Fig. 5).

Table 1. Patient backgrounds

Background factor	Status	No. of subjects examined for validity (%)	No. of subjects examined for reliability (%)
Sex	Male	134 (51.7)	54 (56.8)
	Female	125 (48.3)	41 (43.2)
Patient classification	Outpatient	230 (88.8)	56 (58.9)
	Inpatient	29 (11.2)	39 (41.1)
Age (years)	Unknown	1 (0.4)	0 (0.0)
	<10	30 (11.6)	20 (21.1)
	10–19	72 (27.8)	28 (29.5)
	20–29	104 (40.2)	37 (38.9)
	30–39	49 (18.9)	8 (8.4)
	40–49	2 (0.8)	2 (2.1)
	50–59	1 (0.4)	0 (0.0)
	Mean ± SD	21.8 ± 9.3	18.9 ± 9.9
	Range	0.5–58.2	0.1–45.0
	Median	21.7	19.7
Duration of disease	Unknown	5 (1.9)	3 (3.2)
	<3 months	2 (0.8)	0 (0.0)
	<6 months	5 (1.9)	3 (3.2)
	<1 year	5 (1.9)	4 (4.2)
	<5 years	41 (15.8)	13 (13.7)
	<10 years	33 (12.7)	18 (18.9)
	<20 years	92 (35.5)	32 (33.7)
	<30 years	62 (23.9)	19 (20.0)
	<40 years	14 (5.4)	1 (1.1)
	<50 years	0 (0)	2 (2.1)
	Mean ± SD	13.8 ± 8.9	12.9 ± 9.2
	Range	0.1–36.0	0.3–44.0
	Median	14	12.5

SD, standard deviation.

Table 2. Severity determined using global evaluation (Method 1)

Severity	No. of subjects (%)
Very mild	12 (4.6)
Mild	55 (21.2)
Moderate	123 (47.5)
Severe	53 (20.5)
Very severe	16 (6.2)
Total	259 (100.0)

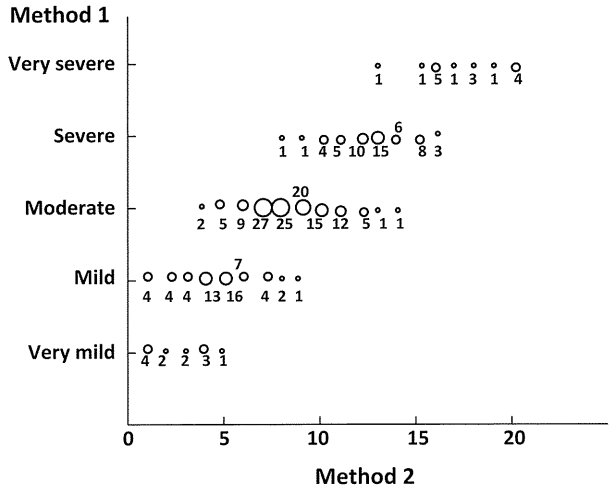


Figure 3. Analysis of correlation between Method 1 and Method 2 (γ = 0.88).

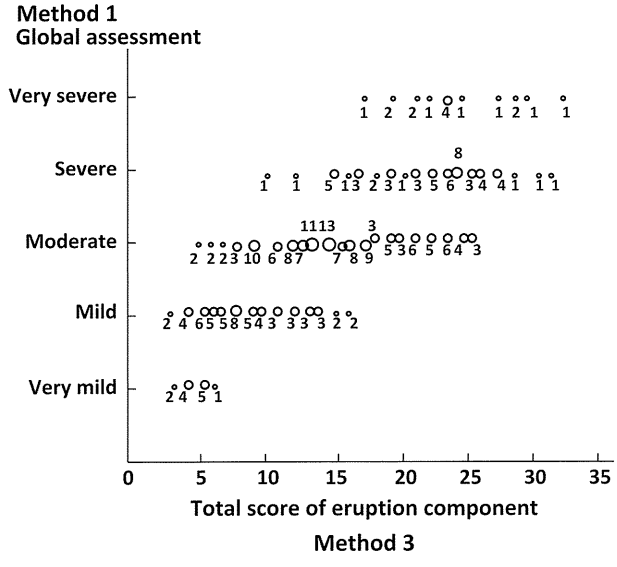


Figure 4. Analysis of correlation between Method 1 (global assessment) and Method 3 (eruption component) (γ = 0.76).

We further determined the κ coefficient, the index of reliability, for each eruption component. During this process, the following two comments were made by examiners: "It is difficult to differentiate between 'severe' and 'very severe'", and "A potential bias in the statistical procedure will occur because the total number of very severe cases is so small". Thus, we included the very severe cases in the severe cases group. The κ coefficient for interexaminer reliability

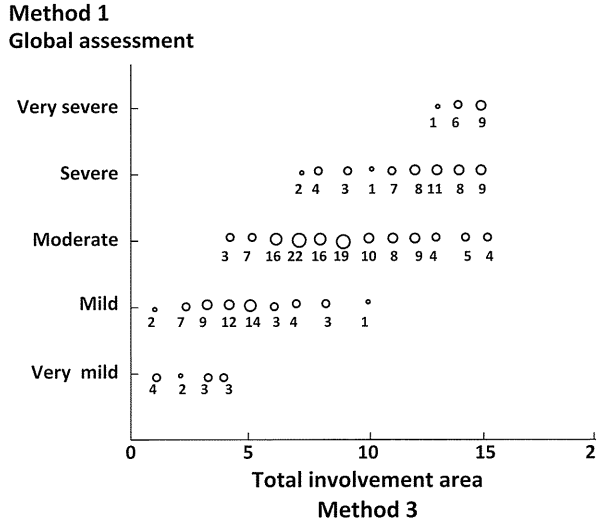


Figure 5. Analysis of correlation between Method 1 (global assessment) and Method 3 (involvement area) (γ = 0.81).

for each eruption component observed at different institutions is shown in Table 3. The mean of the obtained κ coefficients and their respective standard deviations are also shown in Table 3. These results indicate that none of the mean values for eruption components or those for the extent of involvement exceeded 0.3. Among the 11 eruption components, the mean values in only five components (i.e. excoriation, crust, erosion, papule and erythema) were 0.2 or higher. Those giving mean values below 0.15 were prurigo, scale and depilation. However, at some institutions, higher values were observed for these factors, including 0.72 for crust, 0.68 for dry skin and 0.61 for lichenification. In contrast, negative values were observed for some components at a few institutions.

The reliability for utilizing the sum of scores for eruption components and that for the involved area was examined by the ICC. Table 4 shows the results, indicating a value of 0.79 for eruption components and 0.88 for involved area.

The validity of these 11 eruption components was assessed by examining the correlation coefficient (γ) of each eruption component. The correlation coefficients (γ) for global evaluation based on Method 1 were 0.62 for erythema, 0.60 for papule, 0.66 for erosion, 0.57 for crust, 0.64 for excoriation, 0.65 for lichenification, 0.44 for prurigo, 0.17 for depilation, 0.47 for scale, 0.3 for pigmentation and 0.42 for dry

Table 3. κ Coefficient for each component (determined by the same examiners at five institutions, A–E)

Eruption component and area	Institution						
	A	B	C	D	E	Mean	SD
Erythema	0.1959	0.2014	0.1825	−0.0887	0.5522	0.2087	0.1018
Papule	0.2494	0.2364	0.3499	0.0341	0.3333	0.2406	0.0563
Erosion	0.2141	0.2565	0.0965	0.1930	0.4958	0.2512	0.0665
Crust	0.2605	0.1059	−0.0634	0.2308	0.7244	0.2516	0.1313
Excoriation	0.2480	0.3860	0.1656	0.0278	0.4615	0.2578	0.0773
Lichenification	0.1577	0.0908	0.1874	−0.1161	0.6063	0.1852	0.1178
Prurigo	0.0943	0.1386	−0.0516	0.0964	0.0854	0.0726	0.0324
Depilation	0.2899	0.2757	−0.0435	−0.0937	0.2453	0.1347	0.0837
Scale	0.0477	0.0593	0.1331	0.0878	0.3182	0.1292	0.0495
Pigmentation	0.2755	0.2823	0.0493	−0.0714	0.3985	0.1868	0.0858
Dry skin	0.1873	0.0665	0.0613	−0.1250	0.6774	0.1735	0.1355
Head and neck	0.3448	0.1686	0.1820	−0.2076	0.2857	0.1547	0.0963
Anterior trunk	0.3287	0.2576	0.3887	0.0156	0.2908	0.2563	0.0640
Upper limbs	0.3308	0.2591	0.2623	0.0156	0.4014	0.2538	0.0650
Posterior trunk	0.2468	0.3491	0.3274	0.0156	0.3310	0.2540	0.0621
Lower limbs	0.4650	0.2607	0.0536	0.0278	0.2553	0.2125	0.0798
No. of cases	24	20	12	3	4		
No. of examiners	5	4	4	4	4		

Table 4. Intraclass correlation coefficient (ICC)

	Estimate	SE	ICC
Sum of the scores for each eruption component			
Mean of true values	16.86	0.45	0.79
Dispersion in true values	44.95	8.63	
Dispersion in examiners	0.90	0.68	
Dispersion in measurement errors	11.20	1.11	
Sum of the scores for involved area			
Mean of true values	9.30	0.27	0.88
Dispersion in true values	35.98	3.05	
Dispersion in examiners	2.32	0.20	
Dispersion in measurement errors	2.70	0.23	

SE, standard error.

Table 5. Evaluation based on subjective symptoms

	Degree	No. of subjects (%)
Daytime pruritus	None	7 (2.7)
	Weak	93 (35.7)
	Slightly intense	97 (37.5)
	Intense	40 (15.4)
	Very intense	22 (8.5)
Sleep disturbance	None	71 (27.5)
	Weak	60 (23.2)
	Slightly intense	62 (23.9)
	Intense	44 (17.0)
	Very intense	22 (8.5)

skin. Among these, components exhibiting low γ values, such as prurigo, depilation, scale, pigmentation and dry skin, also revealed low multiple correlation coefficients, indicating that their impact on the severity evaluation was low.

The study feasibility was examined with regard to missing entries, mistaken entries in the evaluation sheet and evaluator impression, but no difficulties were observed.

Evaluation results of the subjective symptom of pruritus (Method 4) are shown in Table 5. The γ correlation with global evaluation obtained using Method 1 was 0.698 for daytime pruritus, in contrast to 0.4 or below for sleep disturbance (Table 6). Thus, the γ correlation was lower than that noted between Method 1 and Method 3 (Figs 4,5), but there was some correlation between daytime pruritus and Method 1.

Table 6. Correlation coefficient for subjective symptoms and other evaluation elements

	Method 1	Method 2	Sum of the scores for eruptions evaluated using Method 3	Sum of the areas of eruptions evaluated using Method 3
Daytime pruritus	0.698	0.62	0.58	0.62
Sleep disturbance	<0.4	<0.4	<0.4	<0.4

DISCUSSION

We assembled a committee consisting of individuals from 10 institutions across Japan to establish scientific AD severity classification criteria and carried out a joint study to assess the reliability and validity of these criteria.

In total, 259 patients were examined using Method 1 (global evaluation), Method 2 (severity evaluation in five body regions), Method 3 (evaluation of eruption component in the most severely affected body part and extent of involved areas in five body regions) and Method 4 (evaluation based on subjective symptoms). Using the results of Method 1 as a tentative benchmark, we analyzed their correlation with results obtained using Methods 2, 3 and 4. We found that Methods 1 and 2 showed the strongest correlation, followed by Methods 1 and 3.

Evaluation of the reliability and validity based on the observation of eruption components in Method 3 indicates that it is reasonable to exclude prurigo, depilation and scale from the assessment. The obtained results also indicated that pigmentation and dry skin are less important for severity evaluation.

Furthermore, a 4-step evaluation of eruption components, which included placing the "very severe" group into the "severe" group, was thought to be more practical than the 5-step evaluation consisting of "very severe", "severe", "moderate", "mild" and "none".

In Method 4, which is used to evaluate severity based on subjective symptoms, reliability was found to be accurate only for daytime pruritus but not for sleep disturbance. Therefore, the latter was not considered useful in evaluating severity.

In this study, we were not able to identify an available method of verifying the global evaluation itself, because this evaluation depends on the general physician judgment of AD patient clinical states. Therefore, we employed global evaluation as the basis for assessing other evaluation methods. From these comparisons, Method 2 (global evaluation by summing severity in five body regions), Method 3 (evaluation of each eruption component in the most severely affected body part and extent of involved areas in five body regions) and daytime pruritus of Method 4 were all considered useful in evaluating AD severity. In contrast, the subjective symptom of sleep disturbance was not observed to be useful in evaluation.

PERSPECTIVES AND ISSUES FOR THE FUTURE

On the basis of the results of the present study, addressing the following points is considered necessary to improve the AD severity evaluation method:

1 Severity evaluation based on each body region is simple and practical. If theoretically justified, it will be a desirable method.

2 Further analyses, as indicated below, are necessary for the evaluation of the skin lesions and their involved areas.
   (i) Interexaminer reliability should be improved by streamlining the evaluated components of target skin lesions. Accomplishing this will simplify the evaluation method.
   (ii) Is it necessary to weight the eruption component? If so, how should this be accomplished?
   (iii) Is it necessary to weight the location of the lesion? If so, how should this be accomplished?

3 Is it possible to combine the evaluation of daytime pruritus with other items in order to develop a more useful method?

4 Should evaluation of quality of life be introduced?

REFERENCES

1 Costa C, Rilliet A, Nicolet M, Saurat JH. Scoring atopic dermatitis: the simpler the better? *Acta Derm Venereol* 1989; **69**: 41–45.
2 Consensus report of the European Task Force on Atopic Dermatitis. Severity scoring of atopic dermatitis: the SCORAD index. *Dermatology* 1993; **186**: 23–31.
3 Davis M, Fleiss JL. Measuring agreement for multinominal data. *Biometrics* 1982; **38**: 1047–1051.
4 Sprikkelman AB, Tupker RA, Burgerhof H *et al.* Severity scoring of atopic dermatitis: a comparison of three scoring systems. *Allergy* 1997; **52**: 944–949.
5 Streiner DL, Norman GR. *Health Measurement Scales, A Practical Guide to their Development and Use*, 2nd edn. Oxford Medical Publications, Oxford, 1996; 104–127.

JAPANESE DERMATOLOGICAL ASSOCIATION
THE JOURNAL OF DERMATOLOGY
doi: 10.1111/j.1346-8138.2011.01286.x
Journal of Dermatology 2011; **38**: 632–644

SPECIAL REPORT

English version of the concluding report published in 2001 by the Advisory Committee on Atopic Dermatitis Severity Classification Criteria of the Japanese Dermatological Association

Toshiyuki AOKI,* Hikotaro YOSHIDA,† Masutaka FURUE, Hachiro TAGAMI, Fumio KANEKO, Fujio OHTSUKA, Kiyoshi NISHIOKA, Kiyoshi TODA, Masako MIZOGUCHI, Masamitsu ICHIHASHI, Hiroaki UEKI, Hideo NAKAYAMA, Zenro IKEZAWA, Masahiro TAKIGAWA, Jiro ARATA, Osamu KORO, Shoso YAMAMOTO, Yoichi TANAKA, Masahide ISHIGAKI,‡ Tadashi KUSUNOKI,‡ Kunihiko YOSHIKAWA

Advisory Committee on Atopic Dermatitis Severity Classification Criteria of Japanese Dermatological Association

ABSTRACT

The Japanese Dermatological Association established an advisory committee in 1995 to develop a severity scoring system for atopic dermatitis (AD). Its interim and concluding reports were published in Japanese in the *Japanese Journal of Dermatology* (**108**: 1491–1496, 1998 and **111**: 2023–2033, 2001). Because of the strong demand for an English version, we have decided to publish the reports in English. This manuscript is the English version of the concluding report. The interim report suggested that eruption components such as erythema, papule, erosion, crust, excoriation and lichenification with extent of involved areas in five body regions, including the head and neck, anterior and posterior trunks, and upper and lower limbs, were important items for assessing AD severity. Additionally, it was recommended that streamlining of eruption components was mandatory for improving the statistical validity and reliability. The committee members subsequently concentrated their efforts on this task, and finally proposed an Atopic Dermatitis Severity Classification Criteria of the Japanese Dermatological Association.

Key words: atopic dermatitis, severity classification, pruritus.

INTRODUCTION

A severity classification system for atopic dermatitis (AD) has strongly been desired by clinicians because it is essential for clinical researchers to have a scoring method to serve as an indicator of exacerbation/improvement and a method of communicating treatment outcome. Such attempts have been made in the past, and several reports have become available. However, it is difficult to assess which method is the most effective, and no reports have assessed the efficacy of individual methods.

Members of the Advisory Committee on Atopic Dermatitis Severity Classification Criteria of the Japanese Dermatological Association, formed in August 1995, have repeatedly met under the chairmanship of Dr Hikotaro Yoshida to discuss how to scientifically address this issue. In September 1998, the interim

Correspondence: Masutaka Furue, M.D., Ph.D., Department of Dermatology, Kyushu University, 3-1-1 Maidashi, Higashiku, Fukuoka 812-8582, Japan. Email: furue@dermatol.med.kyushu-u.ac.jp
*Chair, †Former Chair and ‡adviser for statistical analysis.
Received 5 February 2011; accepted 6 April 2011.

report was published in the *Japanese Journal of Dermatology*.[1] Subsequent committee members aimed for the completion of AD severity classification criteria. Visually classifying eruptions that could not be easily assessed quantitatively was difficult. Applying a statistical technique, which has been adopted for this purpose, the committee solved various problems and developed criteria that would be applicable in practical use. Severity classification for eruption was a difficult challenge, and some problems remain unsolved. Here, we report the progress and results of the committee discussions.

INTERIM REPORT AND THE REMAINING ISSUES

In the interim report,[1] the global evaluation (Method 1), which was used to determine severity after examining the patient's entire body (''very severe'', ''severe'', ''moderate'', ''mild'' and ''very mild''), was considered a tentative benchmark. This method was compared with three other evaluation methods, including the sum of global severity scores from five body regions (Method 2), combined severity of the components and extent of eruption (Method 3), and subjective symptom-based evaluation (Method 4).

In Method 2, the skin was divided into five regions (head/neck, anterior trunk, posterior trunk, upper limbs and lower limbs), and the degree and extent of eruption were combined in a 4-step evaluation (''severe'', ''moderate'', ''mild'' and ''none'') in a fashion similar to that in global evaluation. Method 2 exhibited a high correlation with global evaluation (Method 1) (correlation coefficient $\gamma = 0.88$).

In Method 3, the sum of the scores of 11 eruption components (e.g. erythema, papule) and the sum of the scores for the extent of involved area determined in each of the five body regions were both highly correlated to global evaluation (Method 1) ($\gamma = 0.76$ and 0.81, respectively). To prove the validity of the eruption component evaluation, the correlation coefficient between eruption components and global evaluation (Method 1) was calculated. The correlation coefficient for erosion, excoriation, lichenification, erythema, papule and crust was relatively high, whereas that for scale, prurigo and dry skin was slightly lower, and pigmentation and depilation were very low. Interexaminer consistency was calculated using Davis and Fleiss's κ coefficient to prove the reliability of the evaluated items. The consistency of both the area extent and components of eruption were low and unsatisfactory. However, the κ coefficient was substantially different from institution to institution, indicating that reliability can be improved through training.

Daytime pruritus and sleep disturbance were examined in Method 4. Daytime pruritus was highly correlated with global evaluation (Method 1) ($\gamma = 0.70$), whereas the correlation was very low between sleep disturbance and global evaluation (Method 1) ($\gamma < 0.4$).

On the basis of the above results, the four issues below were discussed in ''Perspectives and issues for the future'' in the interim report:

1 Severity evaluation based on each body region is simple and practical. If theoretically justified, it will be a desirable method.

2 Further analyses, as indicated below, are necessary for the evaluation of the skin lesions and their involved areas.
- (i) Interexaminer reliability should be improved by streamlining the evaluated components of target skin lesions. Accomplishing this will simplify the evaluation method.
- (ii) Is it necessary to weight the eruption component? If so, how should this be accomplished?
- (iii) Is it necessary to weight the location of the lesion? If so, how should this be accomplished?

3 Is it possible to combine the evaluation of daytime pruritus with other items in order to develop a more useful method?

4 Should evaluation of quality of life (QOL) be introduced?

The issues discussed in this report are as follows.

Issue 1: Severity by body region

To assess severity by body region (Method 2), global evaluation was repeated in each of the five body regions. Method 2 showed potential as an evaluation method, but a majority of the committee members determined that objectivity and a scientific basis were lacking because all eruption components were comprehensively evaluated at once. Therefore, the committee did not further discuss this method.

However, during the course of discussion, an alternative method was proposed; each of the five body

regions were to be evaluated by considering the intensity and area extent of three eruption components (erythema/acute papule, exudation/crust and chronic papule/nodule/lichenification as discussed below), and the total score was to be calculated. This method was adopted by the Second Committee as Method 3.

Issue 2: Eruption components and area extent

a Eruption components should be streamlined or integrated to improve evaluation validity and reliability. Because this was considered an extremely important and difficult process, the members concentrated their efforts on this task.

b Weighting of eruption components included designating one of the eruption components (e.g. erosion/exudation) as more important than others (e.g. papule), and this component would therefore carry greater weight.

c Weighting of area extent was used to place more emphasis on a specific body region (e.g. face) relative to other regions.

It was difficult to scientifically discuss (b) and (c), and a majority of the members considered that these items were not as urgent as (a), and therefore no further discussion was conducted.

Issue 3: Evaluation of pruritus

Although the usefulness of daytime pruritus for severity classification was demonstrated, this issue was not discussed because objective eruption classification was the committee's main focus.

Issue 4: QOL evaluation

Because the QOL associated with AD was not well understood, this issue was not discussed.

Issue 2 (a) was the only agenda item discussed and resolved by the Second Committee.

COMPOSITION OF THE SECOND COMMITTEE AND FREQUENCY OF MEETINGS

Because of the size of the committee, working groups were formed to ensure efficiency, cut expenditure and prepare proposals for the Second Committee. When necessary, a core member meeting was held.

The Second Committee met four times (the last meeting involved only document circulation), the working group met four times and the core members met twice (Table 1).

METHODOLOGY

The severity of AD eruption was visually and subjectively evaluated. This subjective method had to be standardized to introduce objectivity. To accomplish this, a statistical technique for calculating validity and reliability was adopted as had been reported in the interim report.[1,2]

Validity was assessed depending on whether the targets were adequately included in the evaluation items. Because there was no absolute standard for severity classification, we used global evaluation (Method 1) as a tentative benchmark. Spearman's rank correlation coefficient was used to calculate the correlation between each observation item and the global evaluation (Method 1), and validity was assessed.

Reliability indicated reproducibility. Although there was intra- and interexaminer reliability, the latter was examined to assess the consistency among a number of examiners with respect to their evaluation of the same items. At one institution, evaluation was conducted by three to five dermatologists appointed as examiners. Evaluation outcomes for a substantial number of patients from multiple institutions were collected, and Davis and Fleiss's κ coefficient was calculated.[3]

PROCESS AND RESULTS

How to divide body regions?

The proposition in the interim report was adopted as it was, and it was not discussed further by the Second Committee.

Scoring method

In the interim report, five steps (''very severe'', ''severe'', ''moderate'', ''mild'' and ''very mild'') were used in the global evaluation (Method 1), but four steps (''severe'', ''moderate'', ''mild'' and ''none'') were adopted for the evaluation of eruption components, because very severe and very mild cases were rare. Accordingly, the method of the interim report was adopted without changes.

Table 1. Activities of the Advisory Committee on Atopic Dermatitis Severity Classification Criteria

	Activity	Statistical analysis
August 1995	Launch of the Advisory Committee on Atopic Dermatitis Severity Classification Criteria, Japanese Dermatological Association (Chairman: Hikotaro Yoshida)	
October 1995	Revision of the existing draft by committee members Request to Sandoz (now Novartis Pharma) to provide statistical support	
December 1995	Request to Dr Tadashi Kusunoki (then the Japanese Society for Pharmaceutical Epidemiology, now the Japanese Society for Pharmacoepidemiology) to act as statistical advisor	
January 1996	The survey sheet was prepared based on the draft from the committee The survey was started with 16 participating institutions	
April 1996	The preliminary study was conducted in 94 new students with atopic dermatitis as part of the health examination at Nagasaki University	Frequency distribution: dispersion was confirmed in the collected data Correlation coefficient: examination of the validity of evaluation items and comparison with the Costa method Analysis of main elements: investigation on the weighting of evaluation items Multiple regression analysis: investigation on the weighting of evaluation items Investigation of the feasibility of the comparative study with the global evaluation
January 1997 (Tokyo)	The interim data collected from the survey was reported to the Advisory Committee on Atopic Dermatitis Severity Classification Criteria Validity was examined in 226 cases from 10 institutions Reliability was examined in 79 cases from five institutions Missing entries, etc. in the survey sheet were reexamined	Frequency distribution: dispersion was confirmed in the collected data Regression analysis: the correlation coefficient was calculated for the global evaluation and the total evaluation items Validity of evaluation items: the correlation coefficient was calculated for each item Reliability: κ coefficient was calculated for interexaminer consistency
July 1997 (Tokyo)	Based on the survey results, the committee confirmed the policy for reexamining the evaluation items and evaluation steps	
September 1998	Chairman Hikotaro Yoshida published the interim report of the Advisory Committee on Atopic Dermatitis Severity Classification Criteria in the Journal of the Japanese Dermatological Association	
September 1998 (Hiroshima) Working Group Meeting	New chairman: Dr Toshiyuki Aoki The Working Group was launched Opinions of committee members were used as a basis of discussion The objectives of the severity classification were reconfirmed (restricted to severity of eruptions) The number of eruption components was reduced from 11 to four Evaluation guidelines were prepared A pilot study using the slides was conducted	Eruption components were streamlined on the basis of previous data, and their validity was examined Validity of items: the correlation coefficient was calculated for each item
November 1998 (Kobe) Working Group Meeting	The clinical slides for the evaluation were selected	

Table 1. (*Continued*)

	Activity	Statistical analysis
January 1999 (Tokyo) Committee Meeting	The evaluation of eruption elements was presented on slides Then, different slides were used to test the classification twice (the order of slides was changed for the second test)	Eleven committee members participated in the test using 18 slides Inter- and intra-examiner reliability was analyzed (calculation of κ coefficient)
April 1999 (Tokyo) Committee Meeting	The test results from the previous meeting were reported (reliability was good for erythema, exudation/crust, and nodule/lichenification, but bad for papule) Discussion on how to handle papules The evaluation method for eruption areas was not finalized Selected slides were presented	
September 1999 (Hiroshima) Working Group Meeting	Discussion on how to handle papules The number of eruption components was reduced to three The existing method of handling eruption areas was retained Three drafts of the severity classification were compiled	
October 1999 (Tokyo) Committee Meeting	A decision was made that the number of eruption elements should be reduced from four to three A decision was made on the nationwide survey The three classification drafts were reduced to two Committee members participated in testing the three eruption components on slides	Six committee members participated in the test using 18 slides (calculation of κ coefficient)
January 2000 (Hiroshima) Working Group Meeting	Discussion of the results of the tests by the committee members Elevation of κ coefficient to ~0.3 Selection of clinical slides for preparing a photo-guideline Discussion on adequacy and colors of the printed photo-guideline	
April 2000 (Fukuoka) Core Member Meeting	Confirmation of the presentations at the 99th Japanese Dermatological Association Workshop 2 and the survey procedures	
May 2000 (Sendai) Survey of General Dermatologists	Committee activities and severity classifications were presented at the 99th Japanese Dermatological Association Meeting Workshop 2 General dermatologists participated in testing the severity classification by using the photo-guideline	67 dermatologists participated Correlation coefficient and κ coefficient was calculated
2000 Nationwide Survey by Committee Members	Severity classification was tested across Japan in the institutions of the committee members by using the photo-guideline	Data on 85 cases in nine institutions Correlation coefficient and κ coefficient were calculated for each participating institution and for the total
March 2001 (Osaka) Core Member Meeting	Discussion on the Sendai test and nationwide test results Discussion on the statistical analysis of nationwide results Confirmation of the usefulness of the three streamlined eruption components	
March 2001 Committee Meeting through Documents	A decision was made that the severity classification will be based on a 4-step evaluation of three eruption components in five body regions	

Table 2. Eruption components in the interim report and other severity classification criteria

	Test item	Reliability (κ coefficient)	Correlation coefficient (γ)	Costa (8 items)	SCORAD (6 items)	Interim report
Eruption elements (11)	Erythema	0.21	0.62	○	○	○
	Papule	0.24	0.60	○	○	○
	Erosion (oozing)	0.25	0.66	○*	○	○
	Crust	0.25	0.57	○		○
	Excoriation	0.26	0.64	○	○	○
	Lichenification	0.19	0.65	○	○	○
	Prurigo	0.07	0.44			
	Depilation	0.13	0.17			
	Scale	0.13	0.47	○		
	Pigmentation	0.19	0.30	○		
	Dry skin	0.17	0.42		○	
Area of eruption	Head/neck	0.15	0.58			
	Anterior trunk	0.26	0.71			
	Posterior trunk	0.25	0.65			
	Upper limbs	0.25	0.69			
	Lower limbs	0.21	0.62			

*Vesicles.

Streamlining of eruption components

Table 2 shows the reliability of 11 eruption components (κ coefficient) and the correlation coefficient γ in the interim report. Three eruption components (prurigo, depilation and scale) had a low κ coefficient (≤0.13), and two components (pigmentation and dry skin) had a higher κ coefficient than the above three components (0.17–0.19), but with a low correlation coefficient (<0.45). These five eruption components were thought to be insignificant for severity classification. This finding may appear strange because prurigo and depilation (caused by scratching) are often observed in severe cases. We explained that the statistical insignificance of these eruption components was due to the low incidence of prurigo and depilation. Finally, the remaining six eruption components (erythema, papule, erosion, crust, excoriation and lichenification), which had a correlation coefficient of 0.55 or more and a κ coefficient ranging 0.19–0.26, were concluded to be important for severity evaluation.

The next task of the committee was to determine how to streamline these six eruption components. Table 2 also depicts eight eruption components of the Costa evaluation method[4] and six eruption components of the Severity Scoring of Atopic Dermatitis (SCORAD) method,[5] indicating that most eruption components were included in the interim report. In the streamlining process, care should be taken not to double-score similar eruption components, because repeated scoring would increase the total score. On the basis of extensive discussion, we made the following decisions: (i) erythema and papule would be independent components; (ii) erosion would be replaced by "exudation" and integrated with crust; (iii) nodule, a pre-lichenification condition, would be integrated with lichenification; and (iv) excoriation would be deleted.

To further assess the reliability of these four eruption components (erythema, papule, exudation/crust and nodule/lichenification), representative clinical slides were evaluated by the committee members. As shown in Table 3, only the κ coefficient for papule was unexpectedly and extremely low, indicating that interindividual differences were substantial in diagnosing papules.

This result was disappointing to all members. Although some members vaguely speculated that

Table 3. Reliability of the evaluation of four eruption components (evaluation of 18 slides by 11 dermatologists); κ coefficient

	First evaluation	Second evaluation
Erythema	0.39	0.34
Papule	0.11	0.10
Exudation/crust	0.33	0.40
Nodule/lichenification	0.35	0.39

Table 4. Three streamlined eruption components

Erythema/acute papule (including all redness/flushing, acute papule and edema)
Exudation/crust (including erosion caused by scratching)
Chronic papule/nodule/lichenification (chronic papule is the stage prior to nodule/lichenification)

Table 5. Reliability of three streamlined eruption components (evaluation of 18 slides by six dermatologists); coefficient κ

Erythema/acute papule: 0.33
Exudation/crust: 0.30
Chronic papule/nodule/lichenification: 0.30

papule evaluation would be difficult, they did not expect this result. Despite an obvious symptom, the evaluation of "papule" varied among evaluators. The members engaged in a heated debate regarding this issue. Some members expressed radical opinions such as deleting papule evaluation from the list. After a long discussion, a member proposed that papules might include acute and chronic stages. All committee members agreed that acute papules should be included in erythema, whereas chronic papules should be included in the nodule/lichenification group.

Three streamlined components were selected as candidates: (i) erythema/acute papule, including redness/flushing, acute papule and edema; (ii) exudation/crusts, including exudation, crust and scratched erosion; and (iii) chronic papule/nodule/lichenification, including chronic papule, nodule and lichenification. Chronic papule is the stage prior to nodule/lichenification. Table 4 is a summary of these integrated eruption components.

We then re-evaluated the reliability of the three streamlined components using 18 clinical slides (Table 5). It was revealed that the reliability was not adversely affected by separating papules into acute and chronic phases or by combining acute papule with erythema and combining chronic papule with nodule/lichenification.

Method of classifying the extent of involved area

Nearly every member expressed a varying opinion on this issue, ranging from equal splits to narrower range and weighting. No agreement was reached, and therefore the method of the interim report (classification of the area extent at 1/3 intervals in each of the five body regions) was adopted.

Draft for severity classification

Two drafts were compiled regarding severity classification by combining three eruption components and area extent (Tables 6 and 7). In draft 1, each of the three eruption components was to be evaluated in the most seriously affected site of each body region, the area extents were to be evaluated in each of the five body regions and the total was to be calculated. Evaluation would have to be performed 20 times in total. In draft 2, each of the three eruption components was to be evaluated only in the most serious site of the entire body, the area extent was to be evaluated in each of the five body regions and the total was to be calculated. This method is simpler than that used in draft 1 because evaluation was to be performed only eight times, but the outcomes may be inaccurate. If the two drafts produced similar results, draft 2 could be adopted.

Issue of the evaluation site

How do physicians recognize a patient's severity? Some physicians may identify a case as severe because the eruption components are severe, even though the area extent may be small. Some may consider that the severity is dependent on area extent, even if eruption components are not serious. This

Table 6. Draft 1

	Head/neck	Anterior trunk	Posterior trunk	Upper limbs	Lower limbs	Total
Erythema/acute papule						
Exudation/crust						
Chronic papule/nodule/lichenification						
Area of eruption						
Total score						

Each of the three eruption components is evaluated at the most serious site in each of the five body regions (15 evaluations in total).
The area extent is separately evaluated for each of the five body regions (five evaluations in total), and the two evaluations are combined (20 evaluations in total).

Table 7. Draft 2

	Erythema/acute papule		Exudation/crust	Chronic papule/nodule/ lichenification		Total
Eruption component						
	Head/neck	Anterior trunk	Posterior trunk	Upper limbs	Lower limbs	Total
Eruption area Total score						

Each of the three eruption components is evaluated at the most serious site on the entire body (three evaluations in total).
Separately, the area extent is evaluated at each of the five body regions (five evaluations in total), and the two evaluations are combined (eight evaluations in total).

issue remains unresolved, even after the body is divided into five regions or after eruption components are decided. For example, there are two methods of evaluating erythema of the head/neck region. One method is to calculate the average severity of erythema in the entire head/neck region. However, the human brain cannot quickly and accurately determine this average level. The second method is to evaluate the erythema at the most severe site of the head/neck region. After discussing the advantages and disadvantages of the two methods, the committee chose to score the highest point of each eruption component in each body region (the former method). Nevertheless, according to this method, the score would be high, even if the severe area was small. This weakness was to be corrected by separately evaluating the area extent in five body regions and summing the scores. The committee thus decided that draft 1 is more practical and suitable than draft 2.

Preparation of the photo-guideline

The interim report showed that reliability (reproducibility among examiners) of the classification method varied substantially from institution to institution and suggested that evaluation skills could be improved through training.

Accordingly, we created a photo-guideline for training, which consisted of clinical photos, including various levels of severity of each eruption component.

Survey of general dermatologists

At the 99th Annual Meeting of the Japanese Dermatological Association (Sendai, Japan), a workshop entitled "Aiming for establishment of Atopic Dermatitis Severity Classification Criteria" was held with participation of general dermatologists. The educational photo-guidelines regarding the three eruption components were first presented, and participants were asked to evaluate the clinical photo slides. Evaluation was conducted twice, and the slide order was changed for the second slide presentation.

Table 8 shows the κ coefficient for 67 dermatologists in the Japanese Dermatological Association who attended the workshop. The coefficient was not high for any of the three eruption components. While the coefficient for exudation/crust exceeded 0.3, that for erythema/acute papule was substantially below 0.3. The coefficient for chronic papule/nodule/lichenification was close to 0.2. With respect to evaluator age, however, evaluation by the dermatologists in their 20s generally produced high coefficient levels (0.4–0.5), and the results of the second evaluation improved from the first evaluation, indicating the positive effects of training.

Survey of committee members' institutions

The survey sheets, containing drafts 1 and 2 with the educational photo-guidelines, were sent to committee members for final assessment. Nine participating institutions (Fukushima Medical University, Tokyo Medical and Dental University, St. Marianna University School of Medicine, Yokohama City University, Habikino Hospital, Kobe University, Hiroshima University, Kyushu University, and Nagasaki University) reported 85 cases (7–10 cases per institution).

Correlations between global evaluation (Method 1) and evaluation of each item (validity: correlation coefficient γ) (Table 9) and interexaminer reproducibility of evaluation for each item (reliability: κ coefficient) (Table 10) are shown according to institution. Except for a few items evaluated by some institutions, there

Table 8. Reliability of the evaluation by dermatologists in general practice (κ coefficient)

		First test				Second test							
	No. of dermatologists	Erythema/ acute papule		Exudation/ crust		Chronic papule/ nodule/ lichenification		Mean	Erythema/ acute papule		Exudation/ crust		Chronic papule/ nodule/ lichenification		Mean
Total	67	0.24		0.33		0.21		0.26	0.28	↑	0.32		0.21		0.27
Sex															
Male	40	0.23		0.33		0.20		0.25	0.27	↑	0.32		0.18		0.26
Female	27	0.27		0.33		0.23		0.28	0.29	↑	0.32		0.28	↑	0.29
Age (years)															
20–29	7	0.33		0.49		0.21		0.34	0.45	↑	0.48		0.40	↑	0.44
30–39	13	0.23		0.31		0.24		0.26	0.36	↑	0.38	↑	0.22	↑	0.32
40–49	28	0.28		0.27		0.22		0.26	0.27		0.30	↑	0.22	↑	0.26
50–59	10	0.14		0.39		0.19		0.24	0.08		0.19		0.11		0.13
≥60	9	0.22		0.33		0.15		0.23	0.31	↑	0.28		0.16	↑	0.25

Table 9. Validity of drafts 1 and 2 (correlation coefficient)

Institution	A	B	C	D	E	F	G	H	I	Total
No. of cases	9	7	10	10	9	10	10	10	10	85
Draft 1										
Erythema/acute papule	0.72	0.41	0.65	0.62	−0.29	0.67	0.43	0.70	0.69	0.62
Exudation/crust	0.47	0.60	0.38	0.27	0.86	0.71	0.88	0.45	0.56	0.54
Chronic papule/nodule/lichenification	0.90	0.70	0.52	0.90	0.68	0.54	0.40	0.87	0.84	0.74
Total for eruption elements	0.81	0.60	0.62	0.80	0.82	0.82	0.82	0.89	0.77	0.79
Eruption area	0.91	0.60	0.84	0.84	0.72	0.77	0.83	0.88	0.88	0.84
Total score	0.94	0.50	0.78	0.83	0.82	0.87	0.88	0.92	0.87	0.85
Draft 2										
Eruption elements	0.64	0.77	0.45	0.84	0.73	0.69	0.73	0.35	0.86	0.68
Eruption area	0.91	0.60	0.84	0.84	0.72	0.77	0.83	0.88	0.88	0.84
Total score	0.94	0.60	0.90	0.90	0.90	0.90	0.89	0.85	0.90	0.88

Table 10. Reliability of drafts 1 and 2 (κ coefficient)

Institution	A	B	C	D	E	F	G	H	I	Total
No. of cases	9	7	10	10	9	10	10	10	10	85
Draft 1										
Erythema/acute papule	0.88	0.49	0.20	0.44	−0.02	0.19	0.18	0.12	0.17	0.33
Exudation/crust	0.88	0.53	0.26	0.26	0.54	0.48	0.34	0.06	0.25	0.40
Chronic papule/nodule/lichenification	0.82	0.45	0.20	0.38	0.41	0.40	0.16	0.11	0.16	0.37
Mean of eruption elements	0.86	0.49	0.22	0.36	0.31	0.36	0.23	0.10	0.19	0.36
Eruption area	0.87	0.52	0.32	0.36	0.33	0.36	0.25	0.14	0.39	0.42
Mean for draft 1	0.86	0.50	0.24	0.36	0.32	0.36.	0.23	0.11	0.24	0.38
Draft 2										
Eruption elements	0.76	0.22	0.23	0.31	0.09	0.32	0.02	0.00	0.33	0.27
Eruption area	0.87	0.52	0.29	0.38	0.31	0.36	0.25	0.14	0.47	0.42
Mean for draft 2	0.82	0.37	0.26	0.34	0.20	0.34	0.13	0.07	0.40	0.34

were few differences in the correlation coefficient between institutions. However, the κ coefficient was substantially different among institutions. This finding indicates that global evaluation (Method 1) and evaluation of each component does not require significant levels of training to obtain good correlations, whereas reproducibility among examiners does require training.

Table 11. Validity of draft 1 (nine institutions, 85 cases) (correlation coefficient)

	Head/ neck	Anterior trunk	Posterior trunk	Upper limbs	Lower limbs	Total
Erythema/acute papule	0.52	0.64	0.64	0.30	0.50	0.62
Exudation/crust	0.45	0.46	0.52	0.32	0.48	0.54
Chronic papule/nodule/lichenification	0.52	0.62	0.59	0.63	0.63	0.74
Total for eruption components	0.66	0.71	0.74	0.55	0.63	0.79
Area extent	0.74	0.77	0.73	0.62	0.67	0.84
Total score for eruption components and area extent	0.73	0.78	0.80	0.64	0.69	0.85

Validity

The validity data of drafts 1 and 2 are summarized in Tables 11 and 12, respectively, and are compared with the results in the interim report (Table 13).

The correlation coefficient of draft 1 was 0.62 for erythema and acute papule, 0.54 for exudation and crust, and 0.74 for chronic papule, nodule and lichenification. These results are similar to those in the interim report regarding erythema and acute papule, whereas the coefficient for exudation and crust was lower than that for erosion and crust in the interim report. However, the coefficient for chronic papule/nodule/lichenification was higher than that of lichenification in the interim report. The total correlation coefficient of the three eruption components was 0.79, which was comparable to that reported in the interim report. The correlation coefficient for the area extent was 0.84, which was higher than the 0.81 reported in the interim report. The correlation coefficient of the total score for all items was 0.85 (Tables 11 and 13).

In draft 2, the correlation coefficient for erythema and acute papule, exudation and crust, and chronic papule, nodule and lichenification was 0.45, 0.47 and 0.72, respectively; the coefficient for area extent was 0.84 and the coefficient for total items was 0.88 (Table 12). The correlation between the eruption component evaluation and global evaluation was substantially lower than the 0.79 reported in the interim report. However, the correlation between the evaluation of area extent and global evaluation was higher than that of the interim report and comparable to the level reported in draft 1. The correlation coefficient of the total score for all items was 0.88, which was higher than that of draft 1 (Tables 12 and 13).

Table 12. Validity of draft 2 (nine institutions, 85 cases)

	Correlation coefficient
Eruption component	
Erythema/acute papule	0.45
Exudation/crust	0.47
Chronic papule/nodule/lichenification	0.72
Total	0.68
Area extent	
Head/neck	0.74
Anterior trunk	0.77
Posterior trunk	0.72
Upper limbs	0.61
Lower limbs	0.68
Total	0.84
Total score for eruption components and area extent	0.88

Table 13. Validity of drafts 1 and 2 in comparison with the interim report (correlation coefficient)

Eruption element	Interim report	Draft 1	Draft 2
Erythema	0.62	0.62	0.45
Papule	0.60		
Lichenification (nodule)	0.65	0.74	0.72
Erosion (exudation)	0.66	0.54	0.47
Crust	0.57		
Total for eruption components	0.79	0.79	0.68
Area extent	0.81	0.84	0.84
Total score for eruption components and area extent		0.85	0.88

Reliability

Reliability (reproducibility) among examiners was analyzed. Tables 14 and 15 show summarized data

Table 14. Reliability of draft 1 (nine institutions, 85 cases) (κ coefficient)

	Head/neck	Anterior trunk	Posterior trunk	Upper limbs	Lower limbs	Mean
Erythema/acute papule	0.27	0.38	0.36	0.27	0.35	0.33
Exudation/crust	0.36	0.41	0.42	0.36	0.46	0.40
Chronic papule/nodule/lichenification	0.36	0.43	0.34	0.36	0.35	0.37
Mean of eruption components	0.33	0.41	0.37	0.33	0.39	0.36
Area extent	0.42	0.48	0.40	0.35	0.45	0.42
Mean of eruption components and area extent	0.35	0.43	0.38	0.33	0.40	0.38

Table 15. Reliability of draft 2 (nine institutions, 85 cases)

Evaluation item	κ coefficient
Eruption component	
Erythema/acute papule	0.22
Exudation/crust	0.36
Chronic papule/nodule/lichenification	0.24
Mean of eruption components	0.27
Area extent	
Head/neck	0.41
Anterior trunk	0.48
Posterior trunk	0.41
Upper limbs	0.35
Lower limbs	0.44
Mean of area extent	0.42
Mean of eruption components and area extent	0.34

Table 16. Reliability of drafts 1 and 2 in comparison with the interim report (κ coefficient)

Eruption element	Interim report	Draft 1	Draft 2
Erythema	0.21	0.33	0.22
Papule	0.24		
Lichenification (nodule)	0.19	0.37	0.24
Erosion (exudation)	0.25	0.40	0.36
Crust	0.25		
Mean of eruption components	0.23	0.36	0.27
Area extent	0.23	0.42	0.42
Mean of eruption components and area extent		0.38	0.34

from nine institutions. Table 16 shows the comparison among the interim report, draft 1 and draft 2.

The κ coefficient of draft 1 was 0.33 for erythema and acute papule, 0.40 for exudation and crust, and 0.37 for chronic papule, nodule and lichenification. The κ coefficient was 0.36 for the mean of eruption components, 0.42 for area extent and 0.38 for the mean of the total score (Table 14).

In contrast, the κ coefficients of draft 2 for the 3 eruption components were 0.22, 0.36 and 0.24, respectively; the κ coefficient for the mean of eruption components was 0.27 (Table 15). The κ coefficient for the mean of eruption components and area extent was 0.34, which were substantially lower than the levels reported in draft 1 (Table 16).

On the basis of the above results, the conclusion was reached that draft 1 was generally superior to the simplified draft 2.

CONCLUSION

We would like to propose the severity classification criteria of AD shown in Table 17. The entire body is divided into five regions: head/neck, anterior trunk, posterior trunk, upper limbs and lower limbs. Each of the three eruption components is evaluated (15 times in total) at the most seriously affected site of each body region, in accordance with the 4-step (0, 1, 2 and 3) scoring method (highest score: 45). The area extent occupied by the three eruption components in each body region is separately evaluated (five times) in accordance with the 4-step (0, 1, 2 and 3) scoring method (highest score: 15). The above scores (20 evaluation results) are then totaled (60 points is the full score).

The area extent is evaluated in only four steps (none, <1/3, 1/3–2/3 and >2/3) because evaluation will be difficult if there are more size intervals.

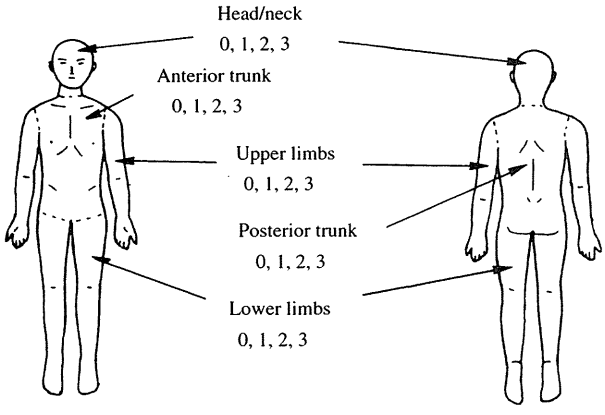
Of the total score of 60 points, 50% is allocated to acute-phase eruption (erythema/acute papule and exudation/crust), 25% to chronic-phase eruption (chronic papule/nodule/lichenification) and 25% to the area extent. Therefore, the result is greatly impacted by acute-phase eruption. Next, as each

Table 17. Severity classification of atopic dermatitis by the Japanese Dermatological Association

This severity classification can be adopted only for cases that are definitely diagnosed as atopic dermatitis
Three eruption components are evaluated in the most severely affected site of each of the five body regions (15 evaluations in total)
The area extent of the five body regions are also evaluated (five evaluations in total)
Both evaluations are combined (20 evaluations in total)
For evaluation of the severity of eruption components in each region, the most severe site is selected for each component
Evaluation of the area extent should be done considering all three components for all five body regions
The highest possible score is 60 points

	Head/neck	Anterior trunk	Posterior trunk	Upper limbs	Lower limbs	Total
Erythema/acute papule						
Exudation/crust						
Chronic papule/nodule/lichenification						
Area extent						

Total Score



Evaluation method
I. Evaluation criteria for three eruption components
0, absent; 1, mild; 2, moderate; 3, severe.
*Explanation of three eruption components
Erythema: All redness, flushing, and edema are included; acute papules: papules not affected by scratching.
Exudation/crusts: Erosion by scratching is included.
Chronic papules: Papules affected by scratching.
Nodules/lichenification: Further progression of chronic papule.
II. Evaluation criteria for area extent
0, no eruption; 1, <1/3; 2, 1/3–2/3; 3, >2/3.

eruption component is evaluated at the most seriously affected site in each of the five body regions, the severity of a limited region has a greater impact than the area extent. As the area extent also receives a certain allocation level, this method can offer well-balanced coverage for AD cases presenting with a variety of features in varying distributions.

After 7 years of activity, our committee proposed AD severity classification criteria that will address the practical needs in a statistically justified manner.

Further studies are necessary to evaluate the clinical value of this method.

ISSUES FOR THE FUTURE

1 The proposed severity classification method cannot be used to evaluate conditions such as scratch-induced depilation and pigmentation. Because the incidence of these symptoms was low, resulting in their low correlation with global evaluation, they were excluded from the evaluated eruption components. Additionally, this method could not be used to evaluate AD characterized mainly by dry skin. Scales and dry skin exhibited low interexaminer reproducibility and were

therefore excluded. When these excluded eruption components have to be monitored by clinicians, another evaluation method should be used. Although interexaminer reproducibility was low for prurigo, this component was retained as a nodule on the list.

2 While the validity of the proposed classification method was good (γ = 0.85), reliability was not as satisfactory (κ coefficient = 0.38). The κ coefficient of 0.8 or more is defined as ''nearly perfect'', 0.6–0.8 as ''substantial'' and 0.4–0.5 as ''moderate''. The κ coefficient obtained using this classification method reached only 0.4 for some components. These results emphasize a necessity for educational training regarding eruption components in acquiring proper evaluation skills.

3 In this committee, we focused on ''severity of eruption in AD'' but not on ''severity of AD''. Although the latter issue cannot be addressed without the former, these two issues are clearly different, and the proposed severity classification is pertinent to the former.

There are many candidate factors other than eruption component and area extent implicated in ''AD severity''.

Pruritus was evaluated in the interim report. Daytime pruritus was found to be correlated with global evaluation (Method 1). Serum lactate dehydrogenase, eosinophil count of peripheral blood, and other laboratory data are known to reflect AD severity. These factors are thought to have a parallel relationship with eruption severity. Meanwhile, the potency of administered anti-inflammatory agents and severity of AD may be inversely related.

The impact of ''pruritus, laboratory data, and anti-inflammatory agents used for treatment'' on the ''severity of eruption in AD'' are not well understood. Further, other indicators for the ''severity of AD'' may incorporate the patient's medical history. These issues must be addressed in the future.

REFERENCES

1 Yoshida H. Interim report from the Advisory Committee on Atopic Dermatitis Severity Classification. *Jpn J Dermatol* 1998; **108**: 1491–1496. (In Japanese)
2 Kusunoki T. Classification for evaluation. *Pharmacia Review* 1991; **28**: 71–86. (In Japanese)
3 Davis M, Fleiss JL. Measuring agreement for multinominal data. *Biometrics* 1982; **38**: 1047–1051.
4 Costa C, Rilliet A, Nicolet M, Saurat JH. Scoring atopic dermatitis: the simpler the better? *Acta Derm Venereol* 1989; **69**: 41–45.
5 Consensus report of European Task Force on Atopic Dermatitis. Severity scoring of atopic dermatitis: the SCORAD index. *Dermatology* 1993; **186**: 23–31.