

Scoring of atopic dermatitis by SCORAD using a training atlas by investigators from different disciplines

Oranje AP, Stalder JF, Taïeb A, Tasset C, de Longueville M. Scoring of atopic dermatitis by SCORAD using a training atlas by investigators from different disciplines.
Pediatr Allergy Immunol 1997; 8: 28–34. © Munksgaard 1997.

The ETAC (Early Treatment of the Atopic Child) study, a multi-national double-blind placebo-controlled randomized trial, has been in progress since 1994. Fifty-six centers in Europe and Canada participate in this study. A total of 817 children with atopic dermatitis [AD] were recruited. The severity of AD was scored using the SCORAD (objective criteria). Ninety-eight investigators (mostly pediatricians) were trained by three members of the European Task Force on Atopic Dermatitis [ETFAD] to standardize their objective SCORAD scoring (system developed by the ETFAD). The experts selected photographs and prepared a training atlas. The percentages of photographs assessed by the 98 non-expert investigators below, within and above the range of evaluations by the three experts were calculated. Taking over- and underscoring together, edema/papulation was the easiest intensity item to score (82% within the range by the experts). The global symptom score, as well as lichenification, edema/papulation, oozing and excoriation registered by physicians with dermatological experience were not statistically significantly different from those by others. Erythema was statistically significantly better scored by those with dermatological experience. The results of the Euclidean Distance method showed that the item excoriations gave the largest distance. Erythema and excoriations were scored better by dermatologically experienced physicians (*t*-test, $p=0.042$ and $p=0.063$ respectively), but lichenification was better scored by non-dermatologically experienced physicians ($p=0.013$). The extent of surface area involved in the disease was calculated on 3 sets of photographs. Most evaluations by the 98 nonexpert investigators were within the range of the experts. Dermatologically experienced physicians scored significantly better than the others (*t*-test, $p=0.006$). This training program is useful for standardizing the scoring in AD and indicates that SCORAD can be used by investigators from different disciplines.

Arnold P. Oranje¹, Jean-François Stalder², Alain Taïeb³, Chantal Tasset⁴ and Marc de Longueville⁴ on behalf of the ETAC Study Group*

¹Department of Dermato-Venereology (Pediatric Dermatology Unit), Rotterdam, The Netherlands,

²Department of Dermatology, CHU Nantes, France,

³Hôpital Pellegrin-Enfants, Bordeaux, France, and

⁴UCB, Belgium

Arnold P. Oranje, Dr, MD PhD, Department of Dermato-venereology, Pediatric Dermatology Unit, University Hospital Rotterdam, dr. Molewaterplein 60, 3015 GJ Rotterdam, The Netherlands

Key words: atopic dermatitis; SCORAD; ETAC

Received: 1 November 1996

Accepted: 30 April 1997

The European Task Force on Atopic Dermatitis obtained consensus on assessment methods in atopic dermatitis [AD] to be used in clinical therapeutic and other trials (1). The final choice for the evaluation system was based on simplicity and easy routine use at outpatient clinics. The system was called SCORAD-index. Extent and subjective symptoms account for around 20% each in the total score and

intensity items account for 60% (1). The test sheet is shown in Fig. 1. The time necessary for this scoring system is about 5–10 minutes depending on the experience of the evaluating physician.

The ETAC (Early Treatment of the Atopic Child) study, started in 1994, is a multi-national double-blind placebo-controlled randomized trial. Fifty-six centers in Europe and Canada are participating. A total of 817 children with atopic dermatitis (AD) were recruited. The main hypothesis under test is

*ETAC - Study Group: see List of authors

whether the onset of asthma can be prevented by early and long-term treatment with cetirizine. Cetirizine is a nonsedating antihistamine. The investigators (mostly pediatricians) were trained by three members of the European Task Force on Atopic Dermatitis (ETFAD) to standardize their objective SCORAD scoring (system developed by the ETFAD). The experts' selected slides, that were copied as photographs, and prepared a training atlas. The goals were

1. to train the investigators and
2. to minimize the difference in scoring the severity of atopic dermatitis.

We report the results of a statistical analysis of SCORAD assessments of photographs in a training manual/atlas by 101 investigators.

Materials and methods

Investigators and centers

The investigators involved in the training were from different disciplines. Most were pediatricians (n=69), whereas 22 were (pediatric) dermatologists or physicians with dermatological experience. The three experts were dermatologists with specialization in pediatric dermatology (Alain Taïeb=at, Jean-François Stalder=jfs, Arnold Oranje=apo). The three experts also participated in the consensus group that developed the testing system: SCORAD-index.

Illustrations

The first report prepared in 1993 by the Task Force on Atopic Dermatitis showing photographs of elementary AD lesions was used for the initial training of the investigators (1). A training manual/atlas for the investigators was prepared by the three expert-investigators (at, jfs and apo) in co-operation with the scientific group of the ETAC sponsor. The experts selected 27 photographs and prepared a training atlas for use in the training program (1).

Data analysis

Total surface-extent of the lesions was evaluated in photographs 1 + 2, 3 + 4, 5 + 6 (front & back sites of three patients) and 7. The extent was rounded to a multiple of 5 (2.5 and 7.5 are rounded upwards). Photograph 7 was not used in the analysis because of poor quality.

The intensity of 5 symptoms for photographs 8 to 27: erythema, edema/papulation, oozing/crusts, excoriation and lichenification were assessed on a four-point ordinal scale (0, 1, 2, and 3). Dryness (an

additional symptom to be scored in the SCORAD) was not used. For each photograph, a global symptom score was calculated as the average intensity of the first five symptoms. For the statistical analysis the results of photograph 8, 11, 13, 16, and 17 were not used. The reasons being poor quality or pigmented skin.

Statistical analysis

All statistical tests were performed two-tailed at a significance level of 5%. Calculation of the percentage of photographs below, within and above the expert range, for each symptom and for the global symptom score was performed. A descriptive analysis of these variables is presented. The evaluations by the physicians with dermatological experience were compared with those without dermatological experience using the *t*-test.

The Euclidean Distance is the distance between the evaluations of each investigator and the average expert score (2). The Euclidean Distance was calculated for each symptom, the global symptom score and the surface of the lesions. The distance expresses a value that counts negative and positive distance data together (how far are the evaluations given by an investigator from the average expert-data, whatever the direction of this distance is [under- or over-evaluation]). The following formula is used to compute this distance:

$$D(i, h) = \sqrt{(I[i, 1] - I[h, 1])^2 + (I[i, 2] - I[h, 2])^2 + \dots + (I[i, 15] - I[h, 15])^2}$$

- I (i,l) = score by investigator i and photograph l.
- I (h,l) = score by average expert h and photograph l.
- D (i,h) = Euclidean Distance between investigator and average expert.

In order to illustrate the meaning of the Euclidean distance for the "global symptom score" consider some examples of evaluation by an investigator. If this investigator gives:

- for all 15 photos the same score as the average expert the Euclidean distance equals 0.
- for six of the photos the same score as the average expert and for nine photos a difference of one unit from the average expert the Euclidean distance equals 3.
- all 15 photos a difference of one unit from the average expert the Euclidean distance equals $3.9 (= \sqrt{15})$.

Average distances to the average expert scores were compared between the investigators with and without dermatological experience, using the *t*-test.

SCORAD
EUROPEAN TASK FORCE
ON ATOPIC DERMATITIS

Last Name First Name

Date of Birth DD/MM/YY

Date of Visit

INSTITUTION

PHYSICIAN

Topical Steroid used:
Potency (brand name)

Amount / Month (6)

Number of flares / Month

Figures in parenthesis for children under two years

A: EXTENT Please indicate the area involved

B: INTENSITY

CRITERIA	INTENSITY
Erythema	
Edema/Papulation	
Oozing/Crusts	
Excoriation	
Lichenification	
Dryness II	

MEANS OF CALCULATION
INTENSITY ITEMS (average representative area)
0= absence
1= mild
2= moderate
3= severe
• Dryness is evaluated on uninvolved areas

C: SUBJECTIVE SYMPTOMS
PRURITUS+SLEEP LOSS

SCORAD $A/5+7B/2+C$

Visual analog scale (average for the last 3 days or nights)
PRURITUS (0 to 10) 0 10
SLEEP LOSS (0 to 10) 0 10

TREATMENT:

REMARKS:

Fig. 1. SCORAD-index scoring system

Results

The 5 intensity items of atopic dermatitis [AD]: erythema, edema/papulation, oozing/crusts, excoriation and lichenification were scored by the 98 investigators and the three experts on the photographs 8 to 27 (excluding photographs 8, 11, 13, 16, and 17) using the objective SCORAD criteria (Fig. 1) (1).

Fig. 2 shows the percentages of photos scored by the investigators below, within and above the range of the three experts. Erythema shows the most under-scoring by 20%, whereas lichenification and oozing show most over-scoring (20 and 19% respectively). Taking over- and under-scoring together, edema/papulation is the easiest intensity item to score (82% within the range by the experts). The scorings by the investigators with dermatological experience for all items except erythema are not statistically significant different from those by the non experienced ones (Table 1). Erythema is better

scored by dermatologically experienced investigators (*t*-test, $p < 0.001$, $p = 0.024$ and $p = 0.057$ for the comparisons of percentages of photographs below, within and above the expert range, respectively).

Fig. 3 shows the range of global symptom scores evaluated by the three experts for each of the photos. No more or slightly more than 1-point difference between the expert-scores is observed. Because of the limited number of experts (= 3), statistical analysis of these data was not possible.

The surface extent of the lesions of atopic dermatitis calculated according to the rules of nine. The extent is calculated from photographs 1 + 2, 3 + 4, 5 + 6 and 7. Photograph 7 was excluded because of quality-variations in different versions of the atlas. Most evaluations by the 98 non-expert investigators were within the range of the experts. The mean Euclidean Distances between the investigators and the average expert were compared between dermatologically experienced physicians and the others: derma-

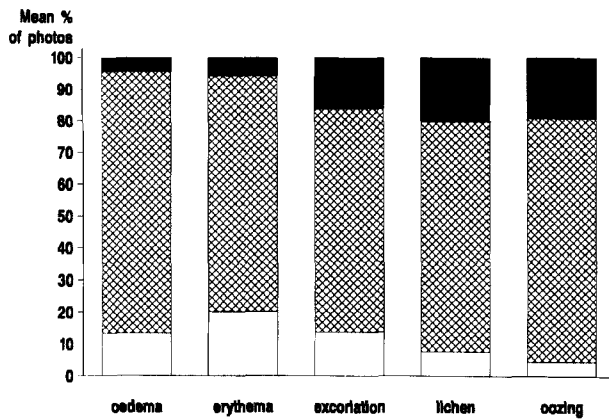


Fig. 2. Mean percentage of photos scored by the non-expert investigators (n=98) below (□), within (X) and above [N] the expert range for the 5 intensity items.

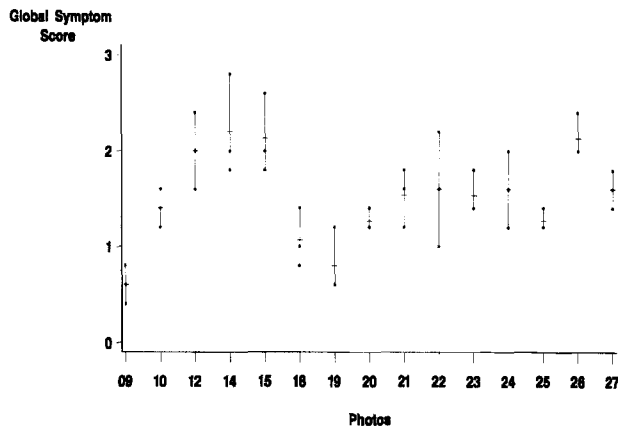


Fig. 3. Global symptom score evaluated by the three experts showing the individual and mean score for each photograph (numbering photographs is according to included photographs).

tologically experienced physicians scored significantly better (*t*-test, $p=0.006$, Table 2).

Figs 4A and 4B show the back of the patient in the original photographs 5 and 6. The shading (Fig. 4B) expresses the extent. The results of the non-experts were compared with those of the experts. Fig. 4C shows the scores by non-experts and experts on photographs 5 + 6. The 25 to 75 percentile borders ranged from 45% to 69% by the non-experts and the minimum and maximum values are respectively 45% and 60% for the experts.

The Euclidean Distance data are given in Fig. 5 and Table 2. The items excoriations, oozing and lichenification yielded the largest distance to the average expert value (more than 3), but for erythema six non-expert investigators scored too high (Fig. 5). Descriptive statistics for the Euclidean Distances between the investigators and the average expert value showed 90th percentile distances of 3.63 for erythema, 3.58 for edema/papulation, 4.64 for ooz-

Table 1. Percentage of photos scored below, within and above the range of the three experts: comparison between the investigators with and without dermatological experience

Intensity Item	Investigators with dermatological experience Mean % of photo $\pm$ SD	Investigators without dermatological experience Mean % of photos $\pm$ SD	<i>p</i> +
Erythema			
% below	10.61 $\pm$ 10.82	23.04 $\pm$ 20.85	< 0.001*
% within	79.39 $\pm$ 9.63	72.23 $\pm$ 20.13	0.024*
% above	10.00 $\pm$ 11.77	4.72 $\pm$ 7.27	0.057
Edema/Papulation			
% below	13.03 $\pm$ 10.98	13.55 $\pm$ 14.61	0.878
% within	81.82 $\pm$ 10.53	82.15 $\pm$ 15.05	0.924
% above	5.15 $\pm$ 9.41	4.31 $\pm$ 6.64	0.697
Oozing/crusts			
% below	4.24 $\pm$ 7.57	4.48 $\pm$ 5.88	0.878
% within	72.12 $\pm$ 15.07	77.33 $\pm$ 13.77	0.130
% above	23.64 $\pm$ 16.42	18.19 $\pm$ 14.29	0.132
Excoriation			
% below	9.72 $\pm$ 11.21	14.65 $\pm$ 12.29	0.095
% within	72.34 $\pm$ 11.69	69.71 $\pm$ 11.90	0.363
% above	17.94 $\pm$ 14.76	15.64 $\pm$ 9.18	0.494
Lichenification			
% below	7.88 $\pm$ 10.82	7.39 $\pm$ 9.84	0.842
% within	67.45 $\pm$ 13.78	73.44 $\pm$ 13.11	0.065
% above	24.68 $\pm$ 14.62	19.17 $\pm$ 12.41	0.082
Global symptom score			
% below	10.91 $\pm$ 13.88	18.76 $\pm$ 19.68	0.084
% within	58.18 $\pm$ 24.70	59.34 $\pm$ 17.51	0.839
% above	30.91 $\pm$ 28.25	21.90 $\pm$ 17.30	0.168

**p* < 0.05: statistically significant difference.
+ *t*-test.

ing/crusts, 4.29 for excoriations, 4.25 for lichenification and 2.42 for global symptom scores.

Table 2 shows the difference in scoring by dermatologically experienced physicians and the others. Statistically significant difference was found for erythema (*t*-test, $p=0.042$). However, for lichenification a difference in favor of the non-experienced ones (*t*-test, $p=0.013$) was noted.

The distances of the global symptom scores by the non-expert investigators to the average score by the experts were also compared. Calculation shows, that seven (7%) of the investigators scored within a distance of 1 and 83 (85%) within a distance of 2. Fifteen (15%) investigators scored the global symptoms above a distance of 2 and only two (2%) above the distance of 3.

Discussion

Assessing the severity of atopic dermatitis [AD] is difficult. The tool used should be reproducible and

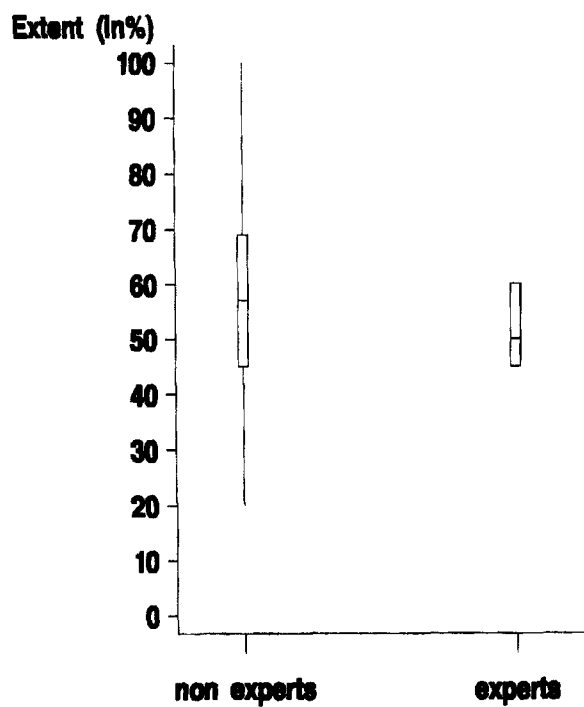
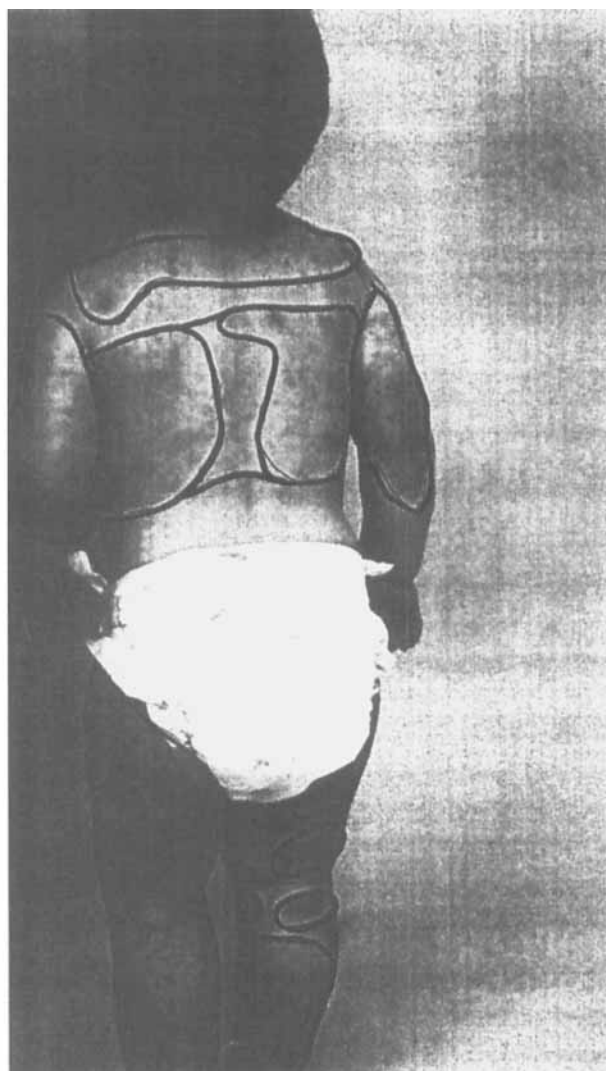


Fig. 4. The surface extent of the lesions of atopic dermatitis calculated according to the rules of nine on photograph 5 and 6 of the atlas. The back of patient in the original photographs 5 and 6 is presented (Fig. 4A). The shading (Fig. 4B) express the extent. The results of the non-experts are compared with those of the experts. The mean surface by the non-experts (Fig. 4C) was 57 (range 20–100) and that by the experts 52 (range 45–60).

Table 2. Euclidean Distance between investigators and average expert; comparison between the investigators with and without dermatological experience

Item	Investigators with dermatological experience n=22	Investigators without dermatological experience n=75	p+
Surface of the lesions (%)			
mean ± SD	19.78 ± 9.87	28.44 ± 18.84	0.006*
Erythema			
mean ± SD	2.38 ± 0.50	2.70 ± 0.99	0.042*
Edema: papulation			
mean ± SD	2.65 ± 0.53	2.61 ± 0.74	0.775
Oozing/crusts			
mean ± SD	3.31 ± 0.89	3.12 ± 0.86	0.373
Excoriation			
mean ± SD	3.15 ± 0.73	3.53 ± 0.85	0.063
Lichenification			
mean ± SD	3.48 ± 0.74	3.04 ± 0.71	0.013*
Global symptom score			
mean ± SD	1.69 ± 0.70	1.58 ± 0.44	0.496

*p < 0.05; statistically significant difference.

+ t-test.

accurate. The European Task Force Group on AD [ETFAD] developed a new assessment tool called SCORAD-index (1). Its guidelines for evaluating AD are defined and calculated. A weighted formula $A/5 + 7B/2 + C$ is used with a maximum score of 103. "A" is defined as the extent of AD using the rule of nine. "B" is defined as the objective intensity items. These items range in severity from 0, 1, 2 and 3. "C" includes subjective items (pruritus and sleep loss) measured by Visual Analogue Scale (1). In routine practice, the index is very accurate and reproducible. An instructive CD-ROM was developed by Stalder in 1995 (3).

European Task force on Atopic Dermatitis [ETFAD] training-programs were undertaken in Bordeaux, Hamburg and Rotterdam with good results obtained by trained (pediatric) dermatologists on "live" patients (4). Inter-observer variability in scoring for objective SCORAD criteria were evaluated. The extent of lesions showed inter-observer variability mostly for patients with lesions of moderate intensity involving 20–60% of the body surface. Intensity items showed less variability, but most difficult were oozing/crusts and lichenifications. Low and high scoring profiles were noted. The beneficial effect of training was evident. Subsequently, the scoring guidelines were optimized (4). The results of trials would be more comparable by training and correcting individual misinterpretations. Therefore, a training program was organized for the investigators participating in the ETAC trial.

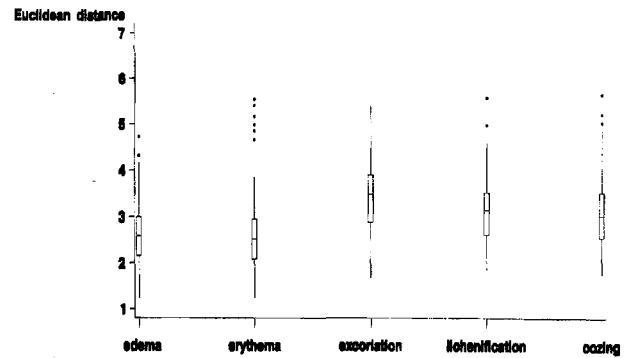


Fig. 5. Euclidean Distances between investigators and average expert for the 5 intensity items (erythema, edema/papulation, oozing/crusts, excoriation and lichenification).

In the present study on evaluation of the training atlas, the most difficult item to score was excoriation. Edema/papulation was the easiest item and that is surprising because in general terms this is a difficult item to evaluate from photographs. Dryness (an additional symptom to be scored in the SCORAD) could not be considered in photographs. Often, one has to palpate the patient to score the dryness. Excoriations showed more than 30% misjudging by the non-expert investigators.

The Euclidean Distance method combines positive and negative differences in one value. Using this method, the results showed that the item excoriations produces the largest distance. Erythema and excoriations were scored better by dermatologically experienced physicians (*t*-test, $p=0.042$ and $p=0.063$ respectively), but lichenification was scored better by non-dermatologically experienced physicians ($p=0.013$).

Measuring of the extent is not a problem when the bordering limits are clear, otherwise it is difficult. This is illustrated in Figs 4A and 4B. In other studies, it was also shown that evaluation of the surface extent is especially difficult when the lesions are not clearly circumscribed (4).

For all the intensity-items a good agreement was achieved between the three experts, seldom more than a range of 1 point-score (based on the limited number of individuals, a statistical evaluation is not possible). The extent evaluations were also within a small range.

In all studies performed to date, the most difficult items to score were lichenification and evaluation of the surface extent (1, 4). However, the use of photographs and the variation in the color qualities impose major restrictions in the current study. The quality of photographs was not ideal and influenced the results negatively. Therefore, several photographs were excluded from evaluation. Pigmented

skin was also difficult to evaluate and these photographs were also excluded from evaluation. However when experienced in evaluating pigmented skin, darkening interpreted as erythema is not problematic to evaluate [4, own experience]. Statistical calculation, provided it is accepted that a score of the Euclidean Distance within a distance of 3 is correct, allow the conclusion that the majority of the investigators scored correctly. For the global symptom scores 85% of them scored within a distance of 2.

In conclusion, this training program illustrates a unique study in which an objective and a subjective instrument such as SCORAD, has been evaluated by 101 investigators. The results indicate that after training the SCORAD can also be used as an instrument by non-dermatologically trained physicians to score AD. Difficult items were lichenification, erythema and excoriations (in this cue). The study did not involve subjective items and dryness (this item cannot be evaluated in photographs). The training and the study will have positive influences on the results of the evaluation of AD in the trial patients and probably will improve the quality of the ETAC (early treatment atopic child) study. However, a quality control study of re-evaluation after training would have answered this question. Such a study has, to date, not yet been undertaken.

Acknowledgements

This study is sponsored by UCB Pharma. The authors are indebted to dr. Bhupendra Tank Department of Dermato-Venereology, Rotterdam for correcting English.

References

1. European Task Force on Atopic dermatitis. Severity scoring of atopic dermatitis: the SCORAD index (consensus report of the European Task Force on atopic dermatitis). *Dermatology* 1993; 186: 23-31.
2. KAUFMANN L, ROUSSEUW PJ. Finding group in data. An introduction to cluster analysis (Euclidean Distance). J Wiley & Sons, New York. 1990.
3. STALDER JF. Training of using SCORAD-index (CD-ROM). 1995.
4. TAIEB A, LABRÈZE L, STALDER J-F, ORANJE AP, KUNZ B. Clinical validation of the SCORAD-index. *Eur J Dermatol* 1995; 5: 69.

List of authors

Scientific advisory board

L. Businco (Roma)
G. Casimir (Brussels)
M. Kjellman (Linköping)
K. Knol (Groningen)
J.L. Menardo (Montpellier)
C. Naspitz (São Paulo)
U. Wahn (Berlin)
J.O. Warner (Southampton)

Investigators and co-investigators

J. Abou-Mrad (Tours)
G. Bellon (Lyon)
E. Bodart (Yvoir)
M. Borres- I.L. Strannegård (Gothenburg)
J. Botey Sala (Barcelona)
T. Bourrier (Nice)
G. Cavagni - M Gardenghi (Brescia)
R.D. Clifford - P. Roy (Dorchester)
J.P. Darras (Béziers)
F.M. de Benedictis - P. Pazzelli (Perugia)
L.E. De Raeve - A. Kempinaire (Brussels)
G. Dutau - F. Rancé (Toulouse)
I. Eichler (Vienna)
J. L. Fauquert - A. Labbé (Clermont-Ferrand)
A. Fiocchi (Milano)
R. Fölster-Holst (Kiel)
K. D. Foote (Winchester)
P. Fortunati (Verona)
R.W. Griffioen - J.H. Sillevius Smitt - J.C. van Nierop (Amsterdam)
A. Grimfeld - F. Sahraoui (Paris)
M. H. Guillet (Brest)
D. Gustafsson - L. Ekholm (Örebro)
I. Huttegger (Salzburg)
J. Leclercq-Foucart (Liege)
R. Lever - G. Ward (Glasgow)
W. Lipschutz (Antwerpen)
C. Marguet (Rouen)
M. Martinez-San Ireneo (Madrid)
M. Masi - A. Miniaci (Bologna)
P. Meglio - P. Lucenti - T. Moretti (Roma)
C. Möller -- G. Forsberg (Umeå)
F. Muñoz-Lopez - M.T. Giner (Barcelona)
K. Negele (Münich)
A. P. Oranje - A. W. Wolkerstorfer - H. J. Neijens (Rotterdam)
A. M. Oudesluys-Murphy - R.N. Sukhai (Rotterdam)
K. Paul - U. Klettke (Berlin)
M. Petrus - M. Rhabbour (Tarbes)
I. Pollock (Enfield)
A. Prehn (Zürich)
S. Ridout - S. M. Tariq (Newport-Isle of Wight)
J. Robert (Lyon)
U. Schauer - C. Rieger (Bochum)
R. Seligmann (Luxemburg)
F. E. R. Simons (Winnipeg)
V. Špicák - M. Fuchs (Prague)
J. F. Stalder - F. Phéline (Nantes)
A. Taieb - L. Labbe (Bordeaux)
D. Teillac-Hamel (Paris)
J. K. van der Woude (Enschede)
J. P. Van Biervliet (Brugge)
D. Vieluf - J. Ring (Hamburg)
G. von Pilgrim (Mainz)
S. Wille - A. Warner (Helsingborg)
E. Young (Amersham)

and the UCB ETAC® TEAM