

CLINICAL REPORTS

Scoring the Severity of Atopic Dermatitis: Three Item Severity Score as a Rough System for Daily Practice and as a Pre-screening Tool for Studies

A. WOLKERSTORFER¹, F. B. DE WAARD VAN DER SPEK¹, E. J. GLAZENBURG², P. G. H. MULDER³ and A. P. ORANJE¹

¹Department of Dermato-Venereology, University Hospital, Rotterdam, ²Medical Department, Glaxo Wellcome BV, Zeist and ³Department of Epidemiology and Biostatistics, Erasmus University, Rotterdam, The Netherlands

Different scoring systems have been developed to determine the severity of atopic dermatitis. The SCORAD (SCORing Atopic Dermatitis), one of the best validated systems, is suited for clinical trials, but is too complicated and time consuming for routine clinical use. The TIS score (Three Item Severity score), a simplified system, is based on the evaluation of erythema, oedema/papulation and excoriation on a scale from 0 to 3. In order to determine the value of the TIS score we conducted a prospective study in 126 children with mild to severe atopic dermatitis. Both the TIS score and the SCORAD were assessed by trained investigators. Interobserver agreement was investigated in 20 children by comparing the independently performed scores of three investigators. A positive correlation was found between the TIS score and the SCORAD (Rank Spearman $r_s=0.86$; $p<0.0005$). The item which correlated best with the SCORAD was excoriation ($r_s=0.72$; $p<0.0005$) followed by oedema/papulations ($r_s=0.66$; $p<0.0005$). Interobserver agreement which was calculated by Cohen's kappa (κ) was "excellent" for SCORAD ($\kappa=0.82$; $p<0.001$) and "fair" for TIS score ($\kappa=0.58$; $p<0.01$). We conclude that the TIS score is a rough, though reliable and simple system for scoring atopic dermatitis. It is particularly suitable in general practice, for routine clinical use and for screening purposes in clinical trials. For research purposes, the objective SCORAD offers a more detailed and comprehensive assessment. **Key words:** SCORAD; TIS; atopic dermatitis; scoring system.

(Accepted March 4, 1999.)

Acta Derm Venereol 1999; 79: 356–359.

A. P. Oranje, MD, PhD, Department of Dermato-Venereology, Dr Molewaterplein 60, NL-3050 GJ Rotterdam, The Netherlands. E-mail: oranje@alkg.azr.nl.

Atopic dermatitis is a common chronically relapsing skin disease affecting 8.9–20.4% of those born after 1970 (1). Patients with atopic dermatitis (AD) account for about 30% of dermatological consultations in general practice (2). Although guidelines give a good framework in managing AD (3), the chronically relapsing course may disappoint not only the patient, but also the physician (4). In the follow-up of such a fluctuating disease a more informative way of recording than "the patient seems better or worse" is necessary. Therefore assessing the severity of AD as objectively and reproducible as possible is extremely important, not only for research purposes, but also in clinical

practice. However, a best scoring system for all purposes is not available (5). The choice of the system will depend on whether it is used for clinical trials, clinical routine assessment or at busy general practice. A system that is intended for research should be sufficiently discriminating and comprehensive to cover the various clinical manifestations of AD. As a result, such a scoring system will be complicated and rather time-consuming. By way of contrast, simplicity and time spent for scoring are crucial issues in a scoring system for daily routine use.

Many different scoring systems have been proposed for assessing the severity of AD (1). These systems are based on the evaluation of 1 or more of the following items: 1) extent; 2) a selection of intensity items; 3) subjective signs (pruritus, sleep loss); and 4) history of eczema.

For research purposes in AD, systems have been described by Hanifin (6), Bahmer et al. (ADASI) (7), Sowden et al. (Leicester score) (8) and Harper et al. (9). The lack of standardization prompted work groups to achieve a consensus on how to score AD. In the UK, the Joint Workshop on Management of Atopic Eczema has recommended the Leicester system (8) for research purposes while in the USA the "Eczema area and severity index" (EASI) has been proposed (Clinical Dermatology 2000, Singapore). Both systems rely on the recording of different signs at different defined body sites.

In Europe, the European Task Force on Atopic Dermatitis has developed and evaluated a composite severity index based on a broad consensus by dermatologists. The resulting SCORAD index (10) consisted of information on the extent, the intensity and subjective symptoms. The SCORAD index has been used in several immunological and clinical trials since then. The objective part of the SCORAD (extent, intensity) has been further validated with regards to the inter-observer variability in 2 studies with patients (11) and with the aid of a pictorial atlas (12). The subjective part (pruritus, sleep loss), which appeared to be a cause of large variations, has been scrapped by the work group except for the follow-up of individual patients. The objective SCORAD is an excellent system for trials, but is too complicated and time-consuming for a routine clinical setting. Therefore, we have developed a simple scoring system called "Three Item Severity (TIS) score", which is based on only 3 intensity items (erythema, oedema/papulation and excoriations). The aim of the study was to evaluate the TIS score in routine clinical practice and to investigate the correlation with the SCORAD.

MATERIAL AND METHODS

Patients

A total of 126 children (mean age 3.9 years; range 4 months to 16 years) with atopic dermatitis according to Sampson (13) (younger than 2 years) and Williams et al. (14) (older than 2 years) was recruited from the outpatient unit of the Paediatric Dermatology Department at the University Hospital Rotterdam. All children with AD visiting our outpatient unit between July 1997 and March 1998 entered the study. The severity of the disease was evaluated using the SCORAD and the TIS score in all children. To assess inter-observer variation, 20 of these children were scored simultaneously by 3 different physicians: an expert dermatologist involved in the development of the SCORAD, another trained dermatologist and a trained non-dermatologist.

SCORAD

The objective SCORAD is a scoring system based on the assessment of extent and intensity in a standardized manner. The complete system is called SCORAD index (10) and also includes the assessment of subjective symptoms (pruritus, sleep loss) on a visual analogue scale. The extent of lesions is scored by applying the rule of nine after drawing the lesions on an evaluation form. The intensity is determined by grading each of the 6 items on a scale from 0 to 3 (erythema, oedema/papulation, oozing/crusts, excoriation, lichenification and dryness). Each item should be scored on the most representative area for a given intensity item. Finally the total score is the sum of extent/5 + 7 × intensity/2. Owing to this formula extent accounts for about 25% and intensity for about 75% of the total score. The range of the objective SCORAD lies between 0 and 83. Based on the objective SCORAD, the severity of AD can be classified into mild (<15), moderate (between 15 and 40) and severe (≥40) AD. The objective SCORAD reflects the modified consensus of the European Task Force on Atopic Dermatitis. Unlike the initial version of the SCORAD index (10), subjective symptoms are not included in the scoring of AD (11).

TIS score

The TIS score is the sum of 3 intensity items scored on a scale from 0 to 3 (erythema, oedema/papulation, excoriations). Similar to the objective SCORAD, each item should be scored on the most representative lesion. This means that different items may be scored on different sites. The range of the TIS score lies between 0 and 9.

Statistical analysis

Correlation between objective SCORAD and TIS score and between objective SCORAD and the different intensity items was calculated

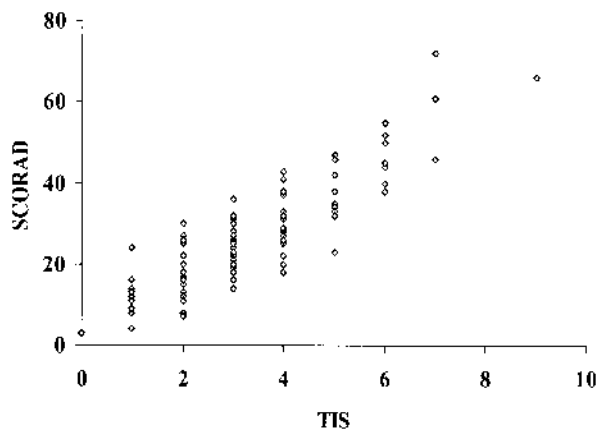


Fig. 1. Correlation between objective SCORAD and TIS score in 126 children with AD. Rank Spearman correlation $r_s=0.86$, $p<0.0005$.

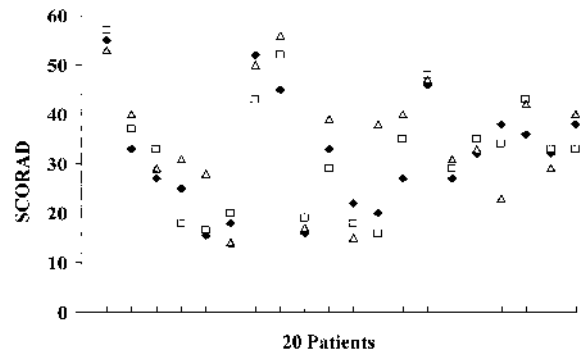


Fig. 2. Objective SCORAD in 20 children by 3 different physicians (◆ expert dermatologist, ▲ dermatologist, □ non-dermatologist). Inter-observer agreement was excellent: $\kappa=0.82$.

using the Rank Spearman's correlation. Inter-observer agreement for both objective SCORAD and TIS score was calculated as the intraclass correlation coefficient using kappa (κ). Agreement between observers was also calculated for each scoring item separately. A $\kappa \leq 0.4$ represents poor agreement; $0.75 > \kappa > 0.4$ represents fair agreement and $\kappa \geq 0.75$ represents excellent agreement (15).

RESULTS

The investigated study population ($n=126$) consisted of children with mild ($n=34$), moderate ($n=78$) and severe ($n=14$) AD according to the objective SCORAD. A positive correlation was observed between TIS score and objective SCORAD (Rank Spearman's $r_s=0.86$ $p<0.0005$) (Fig. 1). The intensity item which correlated best with the objective SCORAD was excoriation ($r_s=0.72$ $p<0.0005$) followed by oedema/papulation ($r_s=0.66$ $p<0.0005$), oozing/crusts ($r_s=0.6$ $p<0.0005$), erythema ($r_s=0.56$ $p<0.0005$), lichenification ($r_s=0.56$ $p<0.0005$) and dryness ($r_s=0.32$ $p<0.0005$). Extent, according to the rule of nine, also correlated well with the objective SCORAD ($r_s=0.82$ $p<0.0005$). Inter-observer agreement was assessed in 20 patients and was "excellent" for objective SCORAD ($\kappa=0.82$; $p<0.001$) and "fair" for TIS score ($\kappa=0.58$; $p<0.01$) (Figs 2 and 3). Furthermore, Figs. 2 and 3 show that a trained physician (non-dermatologist) scores AD as good as a trained dermatologist.

When each scored item was calculated separately (Table I), we observed that the inter-observer agreement was the highest

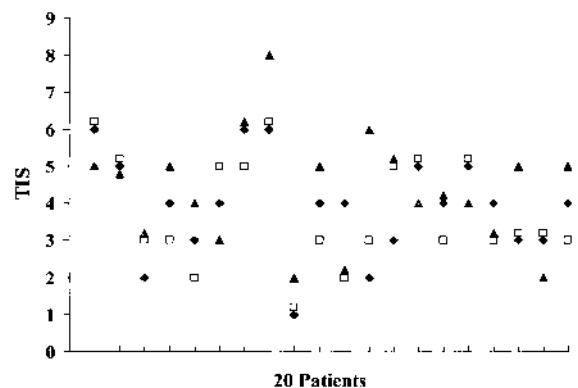


Fig. 3. TIS in 20 children by 3 different physicians (◆ expert dermatologist, ▲ dermatologist, □ non-dermatologist). Inter-observer agreement was fair: $\kappa=0.58$.

Table I. Inter-observer agreement (*kappa*) in 20 children by 3 different physicians

	$\sigma^2_{\text{patient}}^*$	$\sigma^2_{\text{total}}^\dagger$	Kappa ‡
Objective SCORAD	115.88	141.32	0.82
TIS	1.22	2.11	0.58
Extent	627.64	679.72	0.92
Lichenification	0.9	1.19	0.76
Dryness	0.51	0.73	0.7
Oozing/crust	0.39	0.69	0.57
Excoriation	0.23	0.42	0.56
Erythema	0.24	0.45	0.52
Oedema/papulation	0.25	0.6	0.41

* $\sigma^2_{\text{patient}}$ = between-patient variance.
 $^\dagger \sigma^2_{\text{total}}$ = total variance.
 ‡ Kappa represents inter-observer agreement. A kappa ≤ 0.4 represents poor agreement, kappa between 0.4 and 0.75 represents fair agreement and kappa > 0.75 represents excellent agreement.

for extent and the lowest for oedema/papulation. Table I shows the between-patient variance, the total variance and the interobserver agreement (κ) for each of the scored items. Generally, a higher intensity of AD did not increase the variation between observers. An exception was the scoring of lichenification where the interobserver variation increased with the objective SCORAD (Rank Spearman's $r_s = 0.56$ $p < 0.05$).

DISCUSSION

The TIS score is a simple scoring system for AD, which is quick and easy to perform.

In the present study we observed a high correlation of the TIS score with the objective SCORAD and a fair inter-observer agreement between physicians in the TIS score. Of the 6 intensity items scored in the objective SCORAD, excoriation was the 1 that clearly correlated best with the objective SCORAD. This reflects the clinical observation regarding pruritus as a prominent symptom of AD. Next to excoriation, oedema/papulation also showed a good correlation with the objective SCORAD. This supports our choice of including these 2 items in the TIS score. Dryness was the only item that showed poor correlation with the objective SCORAD. While we found a high correlation between the TIS score and the objective SCORAD, it is clear from Fig. 1 that there is still so much variation left that the TIS score cannot accurately predict the objective SCORAD for all individual patients. This means that the TIS score and the objective SCORAD cannot be used interchangeably in individual patients.

In contrast to earlier studies in which considerable inter-observer variation was reported (11, 16), we demonstrate excellent inter-observer agreement between the 3 physicians using the objective SCORAD. We believe that this is a result of both training in scoring AD and the fact that the physicians involved already used the objective SCORAD for scoring AD in clinical routine.

A scoring system to be used in daily clinical routine should be as simple as possible, but this may result in a less sensitive and accurate system. During the course of time, a few systems have been developed with simplicity in mind (16, 17). Of these

scoring systems, only the Basic Clinical Scoring System (BCSS) (16) is as simple and quick to perform as the TIS score. In the BCSS, the extent of AD is scored by evaluating the number of sites involved (5 sites are scored). Each site is scored 0 (no lesion) or 1 (lesion). This system was demonstrated to have an excellent inter-observer agreement, but showed poor agreement with the SCORAD index. The drawback of this system is that it presumably provides a poor reflection of therapeutic interventions. In fact, when patients with AD are treated, the involved sites will not clear completely but improve in extent and intensity without a change of the BCSS. Therefore the BCSS is probably not suited for the follow-up of patients and for evaluating the efficacy of therapy. Although sensitivity to change in individual patients has never been addressed in any study it is known from clinical practice that the items represented in the TIS score are among the first to improve under treatment. The choice of the items used in the TIS score was based on the following criteria:

- 1. The items should be relevant for all age groups.
- 2. If two items are highly correlated only one is scored.
- 3. The items should reflect disease severity and should be independent of other interfering factors.
- 4. The items should be subject to change and improve when AD improves.
- 5. No combination of objective signs with subjective symptoms.

Dryness, xerosis and scaling are characteristic features of AD. They are used in many scoring systems but were not included in the TIS score as they largely depend on when the emollient was last applied. Lichenification which is also used in many scoring systems is not relevant in the very young children as it does not occur before the age of 2 years in Caucasians. Furthermore, lichenification responds rather slowly to therapy and is therefore not suited when evaluating the short-term effects of therapy. Oozing is very typical in infants, but is rare in older children. Moreover, oozing is closely linked to erythema and is therefore already represented by erythema in the TIS score.

Subjective symptoms like pruritus or sleep-loss are strongly influenced by psychological factors and can cause large variations. Therefore, these symptoms were not included in the TIS score. However, these symptoms remain important as an indicator of the quality of life and may serve as a separate measurement tool for follow-up (11). For a complete measurement of the quality of life, DLQI (18) for adults and CDLQI (19) for children are appropriate techniques. Results of earlier studies indicate that the objective SCORAD can also be used by non-dermatologists if they are trained to score AD (12). This is also true for the TIS score as in our study we did not find a difference between the scores by a dermatologist and a non-dermatologist compared with those by an expert dermatologist.

In conclusion, we demonstrate that the TIS score is a reliable and simple scoring system for AD with a fair inter-observer agreement. It is particularly suitable for general practice, for routine clinical use and for screening purposes in clinical trials. However, for research purposes, when a sensitive method is required, the objective SCORAD offers a more detailed and comprehensive assessment of AD with an excellent inter-observer agreement.

REFERENCES

- Schultz Larsen F. The epidemiology of atopic dermatitis. *Monogr Allergy* 1993; 31: 9–28.
- Rook AR, Savin JA, Wilkinson DS. The prevalence, incidence and ecology of diseases of the skin. In: Rook AR, Wilkinson DS, Ebling FJG, et al., editors. *Textbook of dermatology*. 4th edn. London: Blackwell, 1986: 40.
- Mc Henry PM, Williams HC, Bingham EA. Management of atopic eczema. *BMJ* 1995; 310: 843–847.
- Przybilla B, Eberlein-König B, Ruëff F. Practical management of atopic dermatitis. *Lancet* 1994; 343: 1342–1346.
- Finlay AY. Measurement of disease activity and outcome in atopic dermatitis. *Br J Dermatol* 1996; 135: 509–515.
- Hanifin JM. Standardised grading of subjects for clinical research studies in atopic dermatitis: Workshop report. *Acta Derm Venereol* 1989; Suppl 144: 13–14.
- Bahmer FA, Schäfer J, Schubert HJ. Quantification of the extent and the severity of atopic dermatitis: The ADASI score. *Arch Dermatol* 1991; 127: 1,239.
- Sowden JM, Berth-Jones J, Ross JS, Motley RJ, Marks R, Finlay AY, et al. Double-blind, controlled cross-over study of cyclosporin in adults with severe refractory atopic dermatitis. *Lancet* 1991; 338: 137–140.
- Harper JI, Mason UA, White TR, Stoughton RC, Hobbs JR. A double-blind placebo-controlled study of thymostimulin (TP-1) for the treatment of atopic dermatitis. *Br J Dermatol* 1991; 125: 368–372.
- European Task Force on Atopic Dermatitis. Severity scoring of atopic dermatitis: the SCORAD index. *Dermatology* 1993; 186: 23–31.
- Kunz B, Oranje AP, Labreze L, Stalder JF, Ring J, Taieb A. Clinical validation and guidelines for the SCORAD index: Consensus report of the European Task Force on Atopic Dermatitis. *Dermatology* 1997; 195: 10–19.
- 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.
- Sampson H. Pathogenesis of eczema. *Clin Exp Allergy* 1990; 20: 459–467.
- Williams HC, Burney PGJ, Hay RJ, Archer CB, Shipley MJ, Hunter JJ, et al. The UK working party's diagnostic criteria for atopic dermatitis. *Br J Dermatol* 1994; 131: 383–416.
- Cohen J. A coefficient of agreement for nominal scales. *Educ Psychol Meas* 1960; 20: 37–46.
- Sprickelman AB, Tupker RA, Burgerhof A, Schouten JP, Brand PL, Heymans HS. Severity scoring of atopic dermatitis: a comparison of three scoring systems. *Allergy* 1997; 52: 944–949.
- Costa C, Rilliet A, Nicolet M, Saurat JH. Scoring atopic dermatitis: The simpler the better? *Acta Derm Venereol* 1989; Suppl 69: 41–45.
- Finlay AY, Khan GK. Dermatology Life Quality index (DLQI): a simple practical measure for routine clinical use. *Clin Exp Dermatol* 1994; 19: 210–216.
- Lewis-Jones MS, Finlay AY. The Children's Dermatology Life Quality index (CDLQI): initial validation and practical use. *Br J Dermatol* 1995; 132: 942–949.