

CONCISE COMMUNICATION

Reliability testing of the Six Area, Six Sign Atopic Dermatitis severity score

C.R.CHARMAN, A.J.VENN* AND H.C.WILLIAMS

Department of Dermatology, Queen's Medical Centre, Nottingham NG7 2UH, U.K.

*University of Nottingham, Nottingham NG7 2RD, U.K.

Accepted for publication 21 October 2001

Summary

Background The Six Area, Six Sign Atopic Dermatitis (SASSAD) severity score is an objective atopic eczema severity index designed principally to assess response to treatment in therapeutic trials. Validity has been demonstrated in single and multicentre clinical trials, although data on the reliability of the index have not previously been published.

Objectives To assess inter- and intraobserver variability of the SASSAD index.

Methods Six observers with experience in the assessment of atopic eczema were each asked to score disease severity in six patients with moderate to severe atopic eczema using the SASSAD index. Repeat observations were carried out on randomly selected patients by each observer to estimate intraobserver variation.

Results The interobserver variation in total SASSAD scores for each patient ranged from 7 to 30 (median 15.5) out of a maximum possible score of 108. The intraclass correlation coefficient r_i for the total scores among all six observers was quite high at 0.70, although interobserver agreement for individual components of the index was poor to moderate. The maximum recorded intraobserver variation in total SASSAD score in any of the examined patients was 8 units.

Conclusions As with many other tested atopic eczema scoring indices, the SASSAD index is subject to significant interobserver variation, reflecting the difficulties in reliably assessing eczema severity objectively.

Key words: atopic eczema, reliability, SASSAD, severity score

The Six Area, Six Sign Atopic Dermatitis (SASSAD) severity score¹ is an atopic eczema severity index that has evolved from an early attempt to assess the outcome of a non-sedating antihistamine in atopic dermatitis.² It has been widely used in both single and multicentre studies,^{3–7} and has been recommended in some recent reviews of atopic eczema outcome measures.^{8,9} The score comprises assessment of six clinical signs of disease intensity (erythema, exudation, excoriation, dryness, cracking and lichenification) measured at six defined body sites (arms, hands, legs, feet, trunk, head and neck), each on a scale of 0–3, giving a maximum score of 108. Although the index has been validated against global assessments of disease severity, pruritus, sleep loss and topical steroid requirements,⁹ no data on the reliability

of the index (i.e. its ability to produce consistent results if repeated) have been published to date.

Patients and methods

Six patients aged 3–35 years were enrolled in the study; all satisfied the U.K. Working Party's diagnostic criteria for atopic eczema¹⁰ and had moderate to severe disease. The six observers were all experienced in the assessment of atopic eczema, and comprised two dermatology consultants and a dermatology specialist registrar, in addition to three dermatology research fellows who were all carrying out atopic eczema clinical trials at the time of the study.

Written details of the SASSAD scoring system were given to observers in the week before the study, with a further briefing 15 min before the study. Observers examined all patients in a predetermined random order

Correspondence: Dr Carolyn R. Charman.
E-mail: carolyn.charman@nottingham.ac.uk

with no time limit to the examination. At the end of the examination period each observer was asked to re-examine the second patient they had seen without prior warning or availability of previous scores.

Statistics

The kappa statistic (κ) was used as a measure of chance-corrected agreement between the six observers for individual components of the index. Values of κ can range from -1 to $+1$; 0 indicates chance agreement only, 1 represents perfect agreement, and negative values represent less agreement than expected by chance alone. κ -values of <0.20 , 0.21 – 0.40 , 0.41 – 0.60 and >0.61 represent poor, fair, moderate and good agreement, respectively.¹¹ The intraclass correlation coefficient was used to measure interobserver reliability for total scores, using the random effects model.¹²

Results

Total scores for the six patients are shown in Table 1. The interobserver variation in total SASSAD scores for each patient ranged from 7 to 30 (median 15.5). The intraclass correlation coefficient r_i for the total scores for all six observers was 0.70, representing good agreement. In four patients two different observers gave exactly the same score, although in none of these cases were all individual elements of the score graded identically (Table 2). Interobserver agreement for individual elements of the score (Table 3) ranged from less agreement than would be expected by chance alone to moderate agreement only. The maximum intraobserver variation recorded was 8 of 108 points.

Table 1. Six Area, Six Sign Atopic Dermatitis total scores in six patients with atopic eczema as scored by six different observers. The figures in brackets indicate the repeat scores when each patient was re-examined by one observer at the end of the study

	Consultant 1	Consultant 2	SpR	Research Fellow 1	Research Fellow 2	Research Fellow 3	Median score	Range
Patient 1 Age 12 years White European	32	44	43	44	47	53 (52)	44	32–53
Patient 2 Age 3 years Afro-Caribbean	10 (10)	15	25	28	17	16	16.5	10–28
Patient 3 Age 5 years White European	41	40 (32)	53	42	40	46	41.5	40–53
Patient 4 Age 26 years White European	33	44	57	47 (52)	35	63	45.5	33–63
Patient 5 Age 16 years White European	33	30	35	27	27 (24)	38	31.5	27–38
Patient 6 Age 35 years White European	26	33	32 (34)	33	29	31	31.5	26–33

SpR, Specialist Registrar.

Table 2. Six Area, Six Sign Atopic Dermatitis score components in patients given an identical score by two different observers

	Patient 1		Patient 3		Patient 5		Patient 6	
	Observer 2	Observer 4	Observer 2	Observer 5	Observer 4	Observer 5	Observer 2	Observer 4
Head and neck	9	8	4	6	1	4	6	4
Trunk	7	5	5	6	0	0	5	5
Hands	6	7	11	7	8	8	7	8
Feet	7	8	8	6	6	7	4	4
Arms	10	7	6	8	4	4	6	8
Legs	5	9	6	7	8	4	5	4
Total	44	44	40	40	27	27	33	33

Table 3. κ scores for six observers scoring six patients with atopic eczema using the Six Area, Six Sign Atopic Dermatitis index. κ values are given for the six clinical signs scored in six body areas

	Head and neck		Hands		Arms		Trunk		Feet		Legs	
	κ	95% CI	κ	95% CI	κ	95% CI	κ	95% CI	κ	95% CI	κ	95% CI
Erythema	0.18	0.00–0.36	–0.03	–0.20–0.20	0.08	–0.09–0.26	0.25	0.06–0.44	0.14	0.00–0.28	0.09	–0.10–0.29
Exudation	0.46	0.08–0.84	0.08	–0.22–0.38	0.41	0.13–0.70	0.36	–0.26–0.98	0.25	0.02–0.47	0.24	–0.15–0.63
Excoriation	0.06	–0.14–0.26	0.08	–0.16–0.32	0.11	–0.13–0.34	0.19	0.02–0.36	0.14	–0.01–0.29	0.27	0.13–0.41
Dryness	0.19	–0.02–0.41	0.02	0.00–0.05	0.18	–0.03–0.39	0.21	0.04–0.38	0.09	–0.08–0.26	0.20	–0.01–0.41
Cracking	0.07	–0.21–0.35	0.48	0.46–0.50	0.19	–0.09–0.48	0.07	–0.78–0.92	0.36	0.14–0.58	0.19	–0.08–0.45
Lichenification	–0.01	–0.23–0.20	0.06	–0.11–0.23	0.01	–0.20–0.22	0.17	–0.03–0.37	0.18	0.03–0.33	0.04	–0.13–0.21

CI, confidence interval.

Discussion

This study demonstrates the difficulties involved in consistently scoring atopic eczema severity using an objective physician-rated scoring system. Initial inspection of total SASSAD scores suggests reasonable overall agreement between observers, although further analysis reveals wide interobserver variation in scoring for separate components of the index. Lichenification and dryness were the most difficult features to assess reliably, and may be less useful for short-term clinical trials as they tend to respond slowly to treatment. Dryness may be affected by the timing of topical therapies, although this was minimized in the study by asking patients not to apply therapy prior to the examination. Erythema, which showed only poor to fair agreement, is frequently ill-defined and may be difficult to assess in black skin where it is seen as darkening rather than redness. Exudation and cracking showed the best interobserver agreement in this study, although for areas where most observers scored the sign as absent, the resulting 95% confidence intervals were wide.

The SASSAD score is not alone in demonstrating problems with reliability. Lichenification assessments have shown significant interobserver variation using the SCORAD index (12 observers and 19 patients)¹³ and only reasonable agreement using the Eczema Area and Severity Index (EASI; 15 observers and 20 patients).¹⁴ Assessments of erythema and excoriation have also shown statistically significant interobserver variation using the SCORAD index (two observers and 34 patients),¹⁵ with only moderate agreement in more recent reliability studies using the SCORAD index (three observers and 20 patients)¹⁶ and EASI.¹⁴ The assessment of excoriation has also shown significant interobserver variation using Costa's scoring system (two observers and 34 patients).¹⁵

This study provides the first published attempt to assess the reliability of the SASSAD index. We accept that we have included relatively small numbers of patients, but we attempted to include those whose disease severity was representative of patients recruited into clinical trials. Intraobserver agreement was generally good, although memory recall bias cannot be excluded, as the time period between examinations was relatively short (60–90 min). This was chosen for practical purposes and patient compliance. However, each observer recorded over 200 measurements with four possible grades for each measurement, making it difficult to remember individual scores without prior warning.

Relatively little attention has been devoted to the problems of objectively assessing skin disease, although the increasing number of clinical trials and systematic reviews being performed in dermatology is likely to demand a more rational use of validated and reliable scoring systems. Although increasing emphasis is being placed on patients' symptoms and quality of life, most clinical trials still use objective physician-rated scores as their primary outcome measure and, when analysing and comparing results, it is important that the observer variation demonstrated in this and other studies is borne in mind when such clinical measurements are interpreted.

Acknowledgments

This work was carried out as part of a Health Services Research Training Fellowship funded by the Trent NHS Executive.

References

- Berth-Jones J. Six Area, Six Sign Atopic Dermatitis (SASSAD) severity score: a simple system for monitoring disease activity in atopic dermatitis. *Br J Dermatol* 1996; **135** (Suppl. 48): 25–30.

- 2 Berth-Jones J, Graham-Brown RAC. Failure of terfenadine in relieving the pruritus of atopic dermatitis. *Br J Dermatol* 1989; **121**: 635–7.
- 3 Harper JL, Ahmed I, Barclay G *et al.* Cyclosporin for severe childhood atopic dermatitis: short course versus continuous therapy. *Br J Dermatol* 2000; **142**: 52–8.
- 4 Zurbriggen B, Wuthrich B, Cachelin AB *et al.* Comparison of two formulations of cyclosporin A in the treatment of severe atopic dermatitis. A double-blind, single-centre, cross-over pilot study. *Dermatology* 1999; **198**: 56–60.
- 5 Tan BB, Weald D, Strickland I, Friedmann PS. Double-blind controlled trial of effect of house dust-mite allergen avoidance on atopic dermatitis. *Lancet* 1996; **347**: 15–18.
- 6 Berth-Jones J, Finlay AY, Zaki I *et al.* Cyclosporin in severe childhood atopic dermatitis: a multicentre study. *J Am Acad Dermatol* 1996; **34**: 1016–21.
- 7 Granlund H, Erkkö P, Sinisalo M, Reitamo S. Cyclosporin in atopic dermatitis: time to relapse and effect of intermittent therapy. *Br J Dermatol* 1995; **132**: 106–12.
- 8 Finlay AY. Measurement of disease activity and outcome in atopic dermatitis. *Br J Dermatol* 1996; **135**: 509–15.
- 9 Charman C, Williams HC. Outcome measures of disease severity in atopic eczema. *Arch Dermatol* 2000; **136**: 763–9.
- 10 Williams HC, Burney PGJ, Strachan D, Hay RJ. The U.K. Working Party's diagnostic criteria for atopic dermatitis. *Br J Dermatol* 1994; **131**: 383–416.
- 11 Fleiss JL. Measuring nominal scale agreement among many raters. *Psychol Bull* 1971; **76**: 378–82.
- 12 Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. *Psychol Bull* 1979; **86**: 420–8.
- 13 Kunz B, Oranje AP, Labreze L *et al.* Clinical validation and guidelines for the SCORAD index: consensus report of the European Task Force on Atopic Dermatitis. *Dermatology* 1997; **195**: 10–19.
- 14 Hanifin JM, Thurston M, Omoto M *et al.* The Eczema Area and Severity Index (EASI): assessment of reliability in atopic dermatitis. *Exp Dermatol* 2001; **10**: 11–18.
- 15 Sprickelman AB, Tupker RA, Burgerhof H *et al.* Severity scoring of atopic dermatitis: a comparison of three scoring systems. *Allergy* 1997; **52**: 944–9.
- 16 Wolkerstorfer A, de Waard van der Spek FB, Glazenburg EJ *et al.* Scoring the severity of atopic dermatitis: three item severity score as a rough system for daily practice and as a pre-screening tool for studies. *Acta Derm Venereol (Stockh)* 1999; **79**: 356–9.