Acta Derm Venereol (Stockh) 1989; Suppl 144: 28–30

Standardized Grading of Subjects for Clinical Research Studies in Atopic Dermatitis: Workshop Report

JON M. HANIFIN

Department of Dermatology, Oregon Health Sciences University, Portland, USA

Participants at the Third International Symposium on Atopic Dermatitis met in a workshop on the final afternoon of the symposium. The overall objectives for the workshop were to (*a*) define baseline extent and severity in study populations of patients with atopic dermatitis for basic and clinical investigations; (*b*) establish parameters for grading response to therapy and (*c*) optimize clinic trial designs. Presentations by Rajka & Langeland (1) and by Seymour (2), provided useful examples of both simplified and more complex schemes to select patients for research protocols and to define baseline extent and severity of atopic dermatitis.

This report suggests a practical framework of parameters for assessing the course of disease or response to therapy. It is hoped that these suggestions will serve as basis for more uniform study designs that eliminate unnecessary variables and can be compared in confirmatory investigations in various parts of the world.

GRADING OF DERMATITIS

Extent of dermatitis at baseline

Establishing each subject's baseline condition is critical for later assessment of the course of disease or response to therapy. Participants considered that either a dermogram or a percentage estimate of body surface involvement could be used to describe the extent and severity at baseline. The previously established convention of Queille et al. (3), assigning a proportion of 30% to extent and 70% to severity of disease has shown good utility in a recent study (4) and appears to provide a reasonable formula.

Course/stability

As noted by Rajka & Langeland (1), a record is necessary of the course and stability of a patient's dermatitis over the weeks preceding the start of a study. This provides a background level of disease activity against which flares may be contrasted rather than such flares being designated adverse events. Attention to the stability of disease can also be important because unstable patients may have quite severe flares during wash-out periods preceding therapeutic trials. Course/stability can be simply estimated on a scale of 1–3 as described (1). Alternatively, descriptors in this category can include remission (greater than three months), stable, continuous dermatitis or flaring.

Intensity/severity

A wide range of parameters were considered in relation to severity or intensity of atopic dermatitis (Table I). It is evident that for short term studies of topical agents in adults, a minimal, simplified scoring system is ideal, using only the major parameters of erythema, induration and pruritus. In a study or therapeutic trial lasting longer than two weeks, it may also be desirable to add the parameter of lichenification. Likewise this parameter may be important in evaluation of prolonged systemic therapy as might the parameters of scaling and dryness. It should be realized that scaling is a manifestation of healing or resolution and can be a confounding factor. In studying pediatric populations, especially infants and young children, it may be useful to have a category for evaluating erosions which might include the additional descriptors of oozing or crusting. Some discussion was given to the term "vesiculation" but all agreed that this term is not appropriate to atopic dermatitis

Table I. *Signs and symptoms used in grading severity of dermatitis*

1. Erythema[a]
2. Induration/papulation[a]
3. Pruritis/excoriation[a]
4. Lichenification[b]
5. Scaling/dryness[b]
6. Erosion/oozing/weeping[c]

[a] Requisite basic parameters.
[b] May be helpful for systemic trials or for more extensive or long-term evaluations.
[c] Requisite for younger pediatric populations; otherwise generally not relevant to adult populations.

Table II. *Scoring of atopic dermatitis signs and symptoms in a target area*

Erythema	Induration	Pruritus
0.0 = None	0.0 = None	0.0 = None
0.5	0.5	0.5
1.0 = Mild	1.0 = Mild	1.0 = Mild
1.5	1.5	1.5
2.0 = Moderate	2.0 = Moderate	2.0 = Moderate
2.5	2.5	2.5
3.0 = Severe	3.0 = Severe	3.0 = Severe

except for perhaps studies of hand or foot lesions. Discussion was also given to the consideration of eroded, oozing, weeping, crusted lesions as being indicative of infection, but the word "infection" is not recommended for inclusion. Thus, a comprehensive list could include six grading parameters as described in Table I.

Scoring of severity-intensity

Participants came to full agreement that the preferred scoring system utilized a scale from 0–3 indicating none, mild, moderate and severe. Half steps were advised since there are clearly instances where three points alone lack sensitivity for optimal discrimination and intermediate scores are necessary (Table II). It is generally advisable to assess a selected "target" site at each evaluation point but for studies lasting two weeks or more, a global assessment should be made.

Early phase testing of topical agents

A special category of prospective testing in topical trials is the use of limited targeted areas for initial field testing. For comparison testing, symmetrical or otherwise similar sites can be selected and indicated on a dermogram. In the United States, paired comparison testing is generally used for early-phase efficacy and irritancy evaluations. It is important to stress however that because of the potentially labile nature of atopic dermatitis lesions, these studies must be more stringently designed and executed than for testing of more stable conditions such as psoriasis. The ideal patients for such studies are those with moderate disease in a very stable condition. For this reason, it is generally preferable that the spring/summer season be utilized.

Usually two comparable sites are selected, excluding hands, face and thickly lichenified lesions. The size of such test sites can be as small as 2 cm diameter but 3–4 cm diameter lesional areas are preferable. While more than two test sites may be evaluated, this can complicate the procedure. Studies are best carried out on two comparison sites using active drug versus placebo or perhaps, active versus a known agent such as a mid-strength corticosteroid. Dose-ranging studies versus placebo using three sites or dose-ranging versus a known active agent, utilizing four sites can be constructed on special patients but this generally implies a more extensive dermatitis with less stability and more likelihood of flaring.

Test sites should be carefully outlined with skin markers such as are used in radiotherapy. Markings should be reapplied as often as every one to two days depending on how quickly the lines are lost. Caution should be given that the marks might be removed by test vehicles, especially alcoholic or other liquid compounds. Also the markers must be non-irritating since this can add another variable to the evaluation of adverse effects and patient cooperation.

The grading of the small paired comparison sites relates primarily to erythema and induration (see Table I). Evaluation of pruritus can be unreliable on such small sites. Photographs, using a high-quality mobile camera with dedicated flash system and through-the-lens light metering is highly desirable both for grading and for identification of test sites. A general photograph of the large area should be used along with a close-up of the site in order to relocate areas where marks might have been lost. Photographs should be made at the beginning of study and at each visit.

DISCUSSION

Most physicians working in the field of atopic dermatitis have become cognizant of the need to establish a standardized grading system for research studies. Such systems would be applicable not only for topical and systemic drug studies but also for psychological evaluations, assessments of systemic immunotherapy, allergen avoidance regimens and others. After careful establishment of patient-baseline conditions, a standardized, comprehensive yet simple grading system should improve experimental designs and allow comparisons to be made between studies. Additionally, use only of pertinent parameters for evaluation should decrease unnecessary variables; for example, there is often a tendency in clinical studies to apply parameters that are inappropriate for atopic

dermatitis, such as those used for studies of psoriasis where scaling is a necessary descriptor. This leads to imprecision and irregular evaluations.

Workshop participants expressed good agreement regarding the wash-out period for topical systemic drugs. Clearly, systemic corticosteroids should be eliminated for a minimum four week period prior to onset of therapeutic trials and for as long as 3–6 months for studies of the hypothalamic-pituitary-adrenal axis. Other drugs may be continued through the course of therapy as feasible, but certainly for topical trials, a wash out period of topical glucocorticosteroids is necessary, at least from the target treatment area. While from a theoretical standpoint, a wash out of 2–5 weeks for topical agents might be ideal, in practical terms, such a wash-out is both unnecessary and potentially inhumane because topical corticosteroids are a mainstay of therapy and patients not infrequently flare within two days of discontinuance. Some loss of subjects is seen even with a one week wash-out period. Such problems can be obviated by selection of stable patients and testing during spring and summer seasons. Use of emollients should not be interdicted during studies since patients can be considerably harmed by dryness resulting from lack of moisturizing agents.

Conditions for designating non-specific flares of atopic dermatitis should be established before starting trials of therapeutic agents. This can obviate the confusion that might result in otherwise considering flares as adverse effects caused by the drug or vehicle and it also establishes criteria for drop-outs. Generalized flares would be unlikely to result from application of a substance to a small area, especially when the size of test areas have been determined by early toxicology information.

Early-phase testing can be important to new drug development because (*a*) there is no useful animal model of atopic dermatitis and (*b*) existing inflammation models are not necessarily realistic or apropriate to studies of atopic dermatitis. With carefully designed trials, new, potentially beneficial agents such as nonsteroidal drugs can be rapidly tested for efficacy and adverse effects.

Most of the considerations detailed in this report relate to studies of ambulatory patients. Testing of hospitalized patients can be useful, although the spontaneous clearing of atopic dermatitis during hospitalization can cause confusion. However, the hospital setting also can eliminate variables inherent in the outpatient setting.

Any scientific endeavor seeks to reduce variables and improve investigative comparisons. The enthusiastic and constructive cooperation by the participants in the workshop setting at the Third International Symposium on Atopic Dermatitis should advance clinical research in atopic dermatitis and improve development of new therapeutic agents. Hopefully the suggestions recorded here will lead to studies that are more easily interpreted across the wide range of climatic, racial and pigmentary variables.

REFERENCES

1. Rajka, G, Langeland T. Grading of the severity of atopic dermatitis. Acta Dermatovener (this issue).
2. Seymour JL, Hanifin, JM. Detection and grading of atopic dermatitis in infants for clinical research comparison: A study of morphological and microbial parameters (this issue).
3. Queille-Roussel C, Raynaud F, Saurat JH. A prospective computerized study of 500 cases of atopic dermatitis in childhood. I. Initial analysis of 250 parameters. Acta Derm Venereol (Stockh) 1985; 114: 87–92.
4. Seymour JL, Keswick BH, Hanifin JM, Jordan WP, Milligan MC. Clinical effects of diaper types on the skin of normal infants and infants with atopic dermatitis. J Am Acad Dermatol 1987; 17: 988–97.