**Introduction**

Canine atopic dermatitis (CAD) is a common cutaneous disease with a high incidence in Europe and the other industrialised countries [11]. A consensual definition of CAD has recently been proposed by the American College of Veterinary Dermatology Task Force on Canine Atopic Dermatitis: “A genetically-predisposed inflammatory and pruritic allergic skin disease with characteristic clinical features. It is associated most commonly with IgE antibodies to environmental allergens.” [14].

Despite its frequency in veterinary dermatology, many aspects of CAD remain to be elucidated and this disease is the object of numerous studies in different and various fields: aetiology, pathogenesis, epidemiological and clinical aspects, treatment... In relation to this, and particularly for therapeutic studies, a truly objective assessment of disease severity is important. Moreover, CAD is a chronic disease with typical flare-ups, and so the interpretation of results of studies on CAD must be done cautiously.

There is no truly objective method of assessment of the severity of atopic dermatitis in any species. In human medicine, several investigators have studied haematological and biochemical markers like eosinophilic cationic proteins without significant results [1]. An objective measure using permeability barrier function and stratum corneum hydration, with computer-assisted estimates showing the extent of disease seems interesting but still experimental [23]. Many scoring indices have been proposed in order to compensate for the lack of objective methods. More than 13 different indices have been described [5] (e.g. *Atopic Dermatitis Assessment Measure* [6], *Three Item Severity score* [24], *Six Area Six Sign Atopic Dermatitis* [3], *Eczema Area and Severity Index* [10]). These scores are based on the assess-
ment of a variable number of parameters more or less correlated with disease severity. Until now only one has been clinically validated, the SCORing Atopic Dermatitis (SCORAD) [12, 21]. It is a composite index, which associates 6 clinical criteria (erythema, lichenification, excoriation, cutaneous dryness, papules/edema, crusts/oozing), the global extension of lesions and eventually the intensity of pruritus and its influence on life quality. The SCORAD is the consensual result of a logical progression that has taken place over several years.

In veterinary medicine, a composite score has been used for the last 5 years with a small number of variations in its structure for several therapeutic studies. This system of cotation, the Canine Atopic Disease Extent and Severity Index (CADESI), is based on 3 parameters (erythema, lichenification and excoriation) and on the extension of the lesions [16]. However it is neither standardized nor validated. This lack of standardization and validation is responsible for some difficulties of interpretation and comparison.

The aim of this study was the evaluation of CADESI reproducibility between investigators from photographs and from live animals and the determination of any factors resulting in variation.

Material and methods

EVALUATION OF CADESI REPRODUCIBILITY FROM PHOTOGRAPHS

Twenty eight photographs of dogs with CAD were selected. These showed four body areas: face (seven photographs), internal pinnae (11), plantar aspect of a foot (7) or the perineum (3). All were printed on the same paper at the same size (16 cm x 10.4 cm) and a full set was submitted to 6 European College of Veterinary Dermatology (ECVD) diplomates (investigators 1 to 6).

Each investigator scored the photographs on three criteria (erythema, lichenification and excoriation). For each combination a score from 0 to 3 was given. Photograph global score was defined as the sum of these three scores given for each photograph by each investigator.

EVALUATION OF CADESI REPRODUCIBILITY WITH LIVE PATIENTS

Twenty dogs seen in consultation for CAD or adverse reactions to foods were selected. Among these 20 animals, two were seen two and three times respectively with a one-year interval. Twenty three cases were therefore included. Two investigators (A and B) participated in this study. Both investigators scored all the animals on 3 criteria (erythema, lichenification and excoriation) from a predefined list of 39 body areas. For each combination a score from 0 to 3 was given.

STATISTICAL ANALYSIS

The scores were compared in order to determinate the reproducibility of a CADESI-like index between investigators. Mean (± Standard Deviation) is used to describe the variables. The proportion for each categorical variable is described using percentages.

Correlation and agreement between investigators were tested by using the Spearman rank test (Sp; α=5%) and Cohen’s Kappa (K; α=5%) respectively [7, 9].

Results

CADESI REPRODUCIBILITY WITH PHOTOGRAPHS

Table I shows the mean and standard deviations observed for each criterion and each investigator.

<table>
<thead>
<tr>
<th>Investigator</th>
<th>Erythema</th>
<th>Lichenification</th>
<th>Excoriation</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.32(±0.62)</td>
<td>0.96(±0.95)</td>
<td>0.57(±0.74)</td>
<td>2.86±1.9</td>
</tr>
<tr>
<td>2</td>
<td>1.61(±1.07)</td>
<td>1.36(±0.91)</td>
<td>0.39(±0.57)</td>
<td>3.36±2.02</td>
</tr>
<tr>
<td>3</td>
<td>1.68(±0.90)</td>
<td>0.89(±1.1)</td>
<td>0.25(±0.52)</td>
<td>2.82±1.56</td>
</tr>
<tr>
<td>4</td>
<td>1.18(±0.86)</td>
<td>0.93(±1.18)</td>
<td>0.5(±0.69)</td>
<td>2.61±1.93</td>
</tr>
<tr>
<td>5</td>
<td>1.29(±0.76)</td>
<td>1(±0.9)</td>
<td>0.82(±0.67)</td>
<td>3.11±1.52</td>
</tr>
<tr>
<td>6</td>
<td>1.18(±0.9)</td>
<td>0.61(±0.92)</td>
<td>0.54(±0.79)</td>
<td>2.32±1.81</td>
</tr>
</tbody>
</table>

| Ratio highest/lowest | 1.42 | 2.24 | 3.29 | 1.45 |

Table I. — Criterion assessment of photographs : mean (± standard deviation) for each criterion and each investigator.

The mean score for each criterion varied between 0.51 (±0.51) for excoriation and 1.38 (±0.81) for erythema. The ratio between the lowest and highest mean scores varied from 1.42 for erythema to 3.29 for excoriation. The mean score for the total was 2.84 (±1.63). The ratio between the lowest and highest mean scores was 1.45.

The range distribution between investigators for all photographs is presented in table II. The mean value was 2.1 (±0.96) and reached a maximum of 4 for two photographs. Only one photograph was scored in the same way by each investigator.

<table>
<thead>
<tr>
<th>Range</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>%</td>
<td>3.57</td>
<td>21.43</td>
<td>42.86</td>
<td>25.00</td>
<td>7.14</td>
</tr>
</tbody>
</table>

Table II. — Maximal range between investigators for photograph global score.

The average correlation between each combination of investigators for each criterion is shown in table III, by the average Spearman rank observed. A significant correlation (p<0.05) was observed for erythema, lichenification and

<table>
<thead>
<tr>
<th>Criterion</th>
<th>Erythema</th>
<th>Lichenification</th>
<th>Excoriation</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>0.777</td>
<td>0.748</td>
<td>0.494</td>
<td>0.795</td>
</tr>
<tr>
<td>Standard Deviation</td>
<td>0.070</td>
<td>0.082</td>
<td>0.180</td>
<td>0.079</td>
</tr>
</tbody>
</table>

Table III. — Average correlation (Sp rank) for each criterion and total score for photographs.
total score for every combination of investigators. In contrast, no significant correlation (p>0.05) was observed in 27% of combinations for excoriation.

Cohen’s K coefficient, calculated for each combination of investigators from all scores, varied between 0.221 and 0.508 with a mean of 0.395 (± 0.076).

CADESI REPRODUCIBILITY BETWEEN INVESTIGATORS FOR LIVE PATIENTS

Table IV shows the mean and standard deviations observed for the three criteria and the two investigators. Average scores were close with the exception of excoriation for which a ratio of 2 between the two investigators was observed.

<table>
<thead>
<tr>
<th>Investigator</th>
<th>Erythema</th>
<th>Lichenification</th>
<th>Excoriation</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>0.80 (±0.75)</td>
<td>0.34 (±0.64)</td>
<td>0.10 (±0.34)</td>
<td>1.23 (±1.43)</td>
</tr>
<tr>
<td>B</td>
<td>0.82 (±0.60)</td>
<td>0.34 (±0.58)</td>
<td>0.20 (±0.42)</td>
<td>1.35 (±1.35)</td>
</tr>
</tbody>
</table>

Table IV. — Mean (± standard deviation) for each combination of investigator and criterion evaluated on live patients.

The correlation and agreement observed between both investigators are presented for each criterion in table V. Agreement (K) between investigators was mild for erythema, lichenification and total score, and poor for excoriation.

<table>
<thead>
<tr>
<th>Criterion</th>
<th>Erythema</th>
<th>Lichenification</th>
<th>Excoriation</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spearman rank</td>
<td>0.533</td>
<td>0.561</td>
<td>0.294</td>
<td>0.591</td>
</tr>
<tr>
<td>P value</td>
<td>&gt;0.0001</td>
<td>&gt;0.0001</td>
<td>&gt;0.0001</td>
<td>&gt;0.0001</td>
</tr>
<tr>
<td>K observed</td>
<td>0.366</td>
<td>0.385</td>
<td>0.226</td>
<td>0.313</td>
</tr>
</tbody>
</table>

Table V. — Correlation and agreement between investigators for each criterion for live patients.

When comparing scores by body areas a significant correlation (p<0.025) was observed between the two investigators for each site. The agreement between the two investigators varied markedly according to the site. The closest agreement was found for the hind limbs, lateral aspects of the forelimbs, the axillae and the thorax (K>0.5). The poorest agreement was observed for the face and internal aspects of the pinnae (K<0.25). The agreement, therefore, varied from very poor to average.

When comparing symmetrical body sites like the limbs or pinnae, a very low or no disparity was observed (Spearman rank coefficient < 0.06).

Discussion

Photography allows for the presentation of exactly the same image to a large number of investigators. In contrast, no contact with the patient and with no possibility of changing perspective increases the difficulty in being objective about some lesions. Finally, the size and photograph quality are limiting factors to good observational grading.

Observation of live patients during consultations gives investigators access to considerable information, but is difficult to organize. Thus comparison of the observations was less significant in our study, which used only two investigators. Moreover, there are important variations in lesion intensity depending on how active the patient is. This is particularly true for erythema, as effort leads to vasodilatation [13].

These two studies produced complementary information with CADESI reproducibility between investigators being tested under two different conditions.

Observation of photographs showed regular and significant differences between investigators, with some investigators clearly giving superior average scores than others. This last feature might vary depending on the criterion studied. Here, the investigator who gave the highest erythema score was also the one who gave the lowest excoriation score.

The variations in total score are also found for each photograph. In some cases there was an important difference between the lowest and the highest scores. In one third of cases, there were more than three points of difference, one third of the maximum score for each picture.

When comparing all the results from live patients, there was only a small difference between the two investigators, with the exception of excoriation. However important variations were observed in several cases for erythema, lichenification and total score. For excoriation, on a case-by-case basis, more pronounced variations were observed.

Statistic analysis of data from the photographs showed an important correlation between all investigators for erythema, lichenification and total score. By contrast, the observed correlation for excoriation was considerably lower, or non-significant in 27% of cases. This was responsible for a marked decrease in overall agreement between investigators.

Statistical analysis of the patients seen in consultation gave similar results: there were very significant correlations for all criteria and the total score (p<0.001). The correlation was of more significance lichenification and lower for excoriation. The data allowed highlighting of differences at various sites. The face was the area with the lowest correlation.

The CADESI is currently the only scoring index for atopic dermatitis severity in veterinary dermatology. It is a useful tool in that it limits subjective assessment. Unfortunately, the index is not validated. That may be why in some studies modified CADESI scores [15, 17, 18, 22] or new indexes were used for CAD severity assessment [4, 20].

CADESI can be considered as a basis for the construction of an index, which would be more reproducible between investigators and easier or faster to use. Validation will have to take account not only of reproducibility between observers, but also of repeatability over a period of time. This last point is important for clinical studies, in which results are analysed and compared on different dates by the same clinician. In fact, repeatability over time is probably more important than reproducibility between investigators for use in therapeutic studies.

Revue Méd. Vét., 2005, 156, 7, 382-385
Development of such a tool requires a real consensus on clinical criteria correlated with CAD severity - eventually including some other important lesions (alopecia, oozing, crusts...) or some historical features (pruritus intensity, changes in sleep patterns). Once these parameters have been identified and validated, a reproducible method (correlated with disease intensity) could be developed and validated.

Additionally, many important phenotypic differences exist in the canine species and can be responsible for interpretation variations. For example, as in human dermatology [2] skin pigmentation masks, erythematous lesion intensity, which may cause a significant under-estimation of this criterion. Hair density can result in incorrect interpretations on any of the criteria.

**Conclusion**

These results allow several conclusions:

- Erythema and lichenification are reproducible criteria.
- Excoriation is a less reproducible criterion.
- There are differences of scoring dependent on body areas.
- Observed differences between symmetrical areas were very low.
- The use of CADESI with photographs or with live patients showed significant differences in total scores from one clinician to another.

In the authors’ experience, some difficulties exist with the CADESI and may explain some of the problems encountered:

- The CADESI is time consuming, and so observation quality can be negatively influenced.
- There is no clear reference or precise definition for each clinical stage for each criterion as only one study gave the investigators a reference atlas [8].
- This index is not closely correlated with the disease state. Every site carries the same number of points (2.5 % of the total score in our case) although the corresponding skin surface area is extremely variable. The extremities are over-represented and some other areas such as the abdomen or limbs are under-represented. In some areas such as the face or abdomen, old, but focal, lesions could be observed. These areas are particularly difficult to score in terms of lesion intensity or extent.
- The importance of a number of sites may favour multiplication of the differences, which may lead to important variations in the total score.
- Use of the CADESI is dependent on observer training and experience. The use of reference atlases or software tools should considerably improve the reproducibility of this method, as a human study has shown [19].

**References**


