CLINICAL EVALUATION IN ACNE

S. Lidén[1], M.D., K. Göransson[1], M.D., & L. Odsell[2]

[1]*Department of Dermatology, University of Umeå, Umeå* [2]*AB Draco-Tika, Dermatology Group, Lund*

The clinical assessment of the effect of treatment in acne vulgaris is difficult. In most instances there is only partial improvement and it is seldom possible to obtain a complete cure. Several different methods have been used to determine the degree of improvement. The overall estimate of the doctor and/or the patient is seldom used as sole criterium nowadays. Photographs taken before and after the treatment period are often used, but it seems as if few investigations rely entirely upon this method. Counting of the lesions in a more or less sophisticated way has become increasingly common during the last decade.

In 1974 the first observations of the Uppsala group on the beneficial effects of zinc in acne were reported (1). The present work was planned as an independent study of this interesting observation. It was designed in a way which also permitted a methodological study.

MATERIAL AND METHODS

The investigational procedure is seen in Tables 1—6.

The statistical evaluations of the results have been made by using Student's t-test, Wilcoxon's rank sum test and χ^2-test.

Table 1. *Material and methods.*

Patients with acne of grade I—III
Study performed October 1974 — January 1976; pause May — September 1975
59 patients — 5 drop outs
Double-blind, randomized design
Duration of treatment 6 weeks

Table 2. *Material and methods.*

	Zinc sulphate 0.6 g daily	Placebo
Sex	19 males, 8 females	13 males, 14 females
Mean age	19.5 years	21.1 years
Range	14—38 years	16—30 years

Table 3. *Material and methods.*

Assessment at 0, 3, and 6 weeks
— opinion of the patient
— counting of the lesions (same investigator)
— colour photographs

Table 4. *Definitions of type of acne lesions and their severity index. A score representing the total "acne load", is obtained by multiplying the number of lesions of each type by the severity index.*

Severity index	Definition
1/2	Non-inflamed comedones, open and closed. (No erythema).
1	Comedones with surrounding erythema, superficial pustules in which the visible pus has a diameter of 2 mm at the most and with no, or little erythema.
2	Pustules with a diameter exceeding 2 mm or pustules with a significant erythema.
3	Deep infiltrates with or without pustules: isolated cysts.

Table 5. *Material and methods*

Assessment of photographs
— step 1: 37 patients (Sept., 1975)
— step 2: 37 + 17 patients (May, 1976)
Two investigators
Independent, double-blind evaluation

Table 6. *Results of assessment of colour transparencies of 37 patients. Examination of the same slides twice with an interval of nine months.*

Investigator A:	equal assessment 20/37
Investigator B:	equal assessment 25/37
Confidence limits (95%):	A 0.38—0.70
	B 0.53—0.83

RESULTS

The therapeutic effects of zinc versus placebo are seen in Fig. 1 and Table 7. After both treatments there was a statistically significant reduction in total number of lesions and the overall score. However, no statistically significant reduction was observed between the third and sixth week in the placebo-group, whereas the reduction in the zinc-treatment group was significant after both the first and the second 3-week period.

The results of treatment achieved in the two treatment groups was compared by a Wilcoxon's rank sum test (Table 7). No difference was found after 3 weeks of treatment. After 6 weeks of treatment there was a significant difference in favour of the zinc group in the total number of lesions, in score, and the percentage change in score. The effect of zinc was more pronounced on inflammatory lesions in grade 2—3 acne than on comedonal lesions. An analysis of the patients' subjective evaluation after 6 weeks of treatment showed no statistically significant differences.

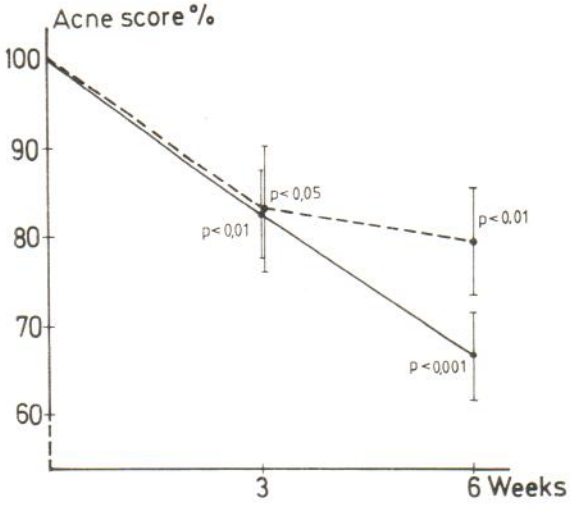

Fig. 1. Per cent change in acne score ± standard error of the mean (vertical lines) and p-values after treatment for 3 and 6 weeks with placebo (— — —) or zinc (————).

Table 7. *Differences between treatments with oral zinc and placebo. The figures indicate the p-values obtained by Wilcoxon's rank sum test.*

	Acne grades 1/2	1	2-3	Total number	Score	% Change score
3 weeks	0.84	0.16	0.53	0.65	0.69	0.99
6 weeks	0.33	0.10	0.08	0.03*	0.013*	0.007**

* = $p < 0.05$
** = $p < 0.01$

EVALUATION OF PHOTOGRAPHS

No difference between the effect of the two types of treatment was found when evaluating photographic records. The overall results from the second evaluation, comprising all the 54 patients, is shown in Fig. 2.

The pictures of the first 37 patients were evaluated twice, with nine months in between the evaluations. A comparison between the two evaluations showed great differences in the results. One of us made equal assessments in 20 patients of the 37 and the other in 25 patients of the 37. Hence, the probability of equal assessments are estimated as $p_1 = 20/37 = 0.54$ and $p_2 = 0.68$. Approximate 95% confidence intervals are (0.38; 0.70) and (0.53; 0.83), respectively.

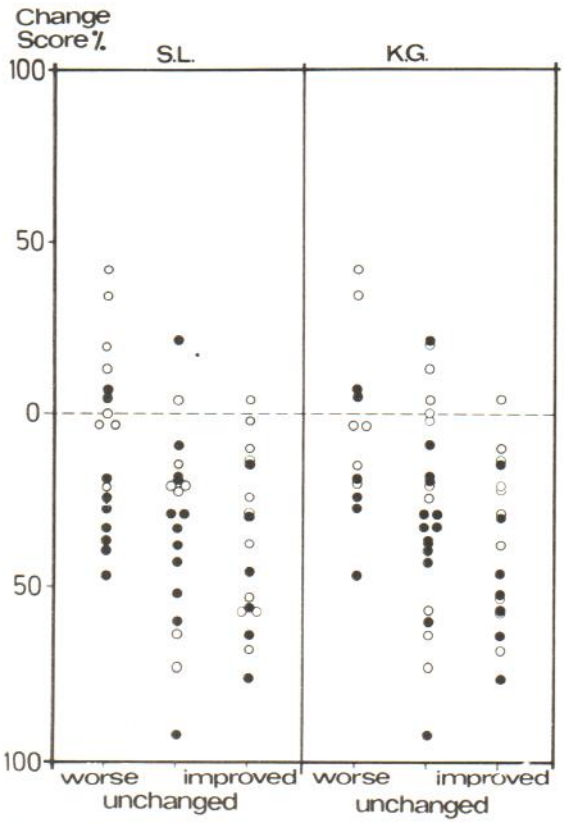

Fig. 2. The results of the evaluation of colour photographs by two independent observers, compared with the acne score obtained by lesion counting after 6 weeks of treatment. ○, Placebo; ●, Zinc sulphate.

DISCUSSION

To date the effect of zinc on acne vulgaris has been described in four papers (1, 2, 3, 4). A varying degree of beneficial effect has been found in all but the paper by Weismann et al. (2). The weak but statistically significant effect found in the present work after 6 weeks of treatment seems to be comparable with the results obtained by Hillström et al. (4).

Neither the patients' subjective evaluation nor the photographic assessment were sensitive enough to reveal a positive effect from zinc treatment. The very low reproducibility of the photographic assessment might be surprising. We consider it as further support for lesion counting methods. While this method is subjective, the investigator is forced to make up his mind on every single lesion and this is done when the conditions for a correct grading are most favourable, i.e. when the patient can be examined in surgery.

SUMMARY

Various methods for clinical evaluation of the effect of treatment of acne vulgaris have been studied. Counting of acne lesions was found to be more sensitive than the opinion of the patients and the assessment of colour photographs. The correlation between the results obtained by lesion counts and evaluation by photographs was low, as was also the reproducibility of photographic assessment.

This methodological study was performed during a 6 week period, using a double-blind technique for comparing the effect of 0.6 g of oral zinc sulphate daily versus placebo in 54 patients. With zinc treatment acne improved with about one third as rated with a score system. Compared to placebo this result is slightly better ($p < 0.05$). Inflammatory acne of grades 2—3 responded more favourably than comedonal acne to zinc treatment.

REFERENCES

1. Michaëlsson, G., Juhlin, L., Vahlquist, A.: Effects of oral zinc and Vitamin A in acne.
   Arch Dermatol 1977; 113:31.
2. Weismann, K., Wadskov, S., Söndergaard, J.: Oral zinc sulphate therapy for acne vulgaris.
   Acta Derm Venereol (Stockh) 1977; 57:357.
3. Michaëlsson, G., Juhlin, L., Ljunghall, K.: A double-blind study of the effect of zinc and oxytetracycline in acne vulgaris.
   Br J Dermatol 1977; 97:561.
4. Hillström, L. et al.: Comparison of oral treatment with zinc sulphate and placebo in acne vulgaris.
   Br J Dermatol 1977; 97:679.

DISCUSSION

Strauss, Iowa: I would like to take the chairs' prerogative to make a comment. First of all, in reference to the material that Prof. Lidén presented, I feel that the global assessment of acne, which essentially is what is done with photographs, has limited significance. A patient with three huge cystic lesions on her face may only have these three lesions and will be classified as having grade 4 acne. Similarly, the patient who has 200 comedones, all of which are non-inflamed lesions, probably will look as bad and may end up being called a grade 4 acne subjectively. The evaluator, therefore needs excellent guidelines for global assessment. Furthermore, photographic interpretation is of value only if close-ups are available and all equipment for projecting the slides is identical, particularly if before- and after-photographs are projected simultaneously. I think that lesion counting is much more accurate.

Plewig, Munich: Prof. Lidén, I do not understand why you used a counting system and then multiplied by the grading system. Would it not have been better if you had just counted the individual types of lesions, such as the number of comedones or the number of inflammatory lesions and then compare the pre-and post-treatment figures for every patient? How does your data look if it is handled in this way?

Lidén, Umeå: You mean, if we had split up the material into the different severity classes?

Plewig, Munich: No. You have pooled the inflammatory and non-inflammatory lesions and multiplied the number by the Pillsbury-grading system. I have objections to this kind of evaluation. Why did you not count the lesions in every patient: open comedones, closed comedones, papules, pustules, and cysts and evaluate what happened to these numbers over a 6-week period? If your therapeutic agent is effective for inflammatory acne, the number of inflammatory lesions should decrease. On the other hand, if treatment is effective for comedones, the number of comedones should decrease.

Lidén, Umeå: We have partially done that, as you may have noticed in one of the tables. The comedones

did not respond as well as the more severe inflammatory lesions. We feel, however, that you get a better picture of the total acne load if you multiply the individual counts. This is simply another way of making a global assessment and expressing it in figures.

Plewig, Munich: But you see, the problem is that what you call severe acne in Umeå may be moderate acne in Leeds. We need a standard method of evaluating acne and this is a major problem in today's research on acne. I think the best way is still to count individual lesions, keeping the classes of lesions separate.

Leyden, Philadelphia: I shall make the same comment as Dr. Plewig has made. It seems to me that by picking your number as 1/2, 1, 2, or 3 you can greatly influence the severity score. As already pointed out, there are great difficulties in evaluating photographs, but I thought the pictures were better than the severity scores for the same reason that Dr. Plewig has pointed out. You have picked a number to multiply your total count by and if you happen to have a couple of large lesions at one point, then that number is going to be much higher than the next time if three lesions have disappeared. If the lesions happen to be large ones, you are going to have a 50 or 60% change, when actually all you may have had was approximately a 5% change in total number of inflammatory lesions. I would not use that multiplication factor either — unless there was some way of standardizing the numbers in a way that we all can agree on.

Zachariae, Aarhus: I just want to ask you how you feel about seasonal variations? They might influence the data in your study, depending on when it was done. You did not mention this?

Lidén, Umeå: We interrupted the study during summer time. We worked between October 1974 and May 1975 and then from October 1975 to January 1976.

Cunliffe, Leeds: I think the topic of clinical grading in acne again could keep us here to midnight. But I would agree with what several people have said that really it is quite crazy that in 1978 there is no relatively unified way of assessing acne lesions. There is no doubt that counting the various types of lesions such as whiteheads, blackheads, papules, and pustules is something which we all understand. I think that there is something to be said, Dr. Strauss, for the use of the over-all grade. I think this is worthwhile doing as well as the counting, even if it is of value only in your own clinic. When we see patients in the clinic there is no time to count the individual lesions. For the last seven or eight years we have been assigning a grade to each patient on a 0—10 scale simply by looking at them. We have found that within the same person you do get an 85% reproducibility and it is only in 1% that in fact you are more than two grades away. I do not think that that is bad for a patient evaluation technique. In the clinic we can measure blood pressure and the pulse. Why is it that we cannot try to develop a simple system for grading? Not just in our day-to-day management of acne patients, but in clinical trials.

I would also like to ask Dr. Lidén what is his reproducibility of his technique? It he takes 10 patients and scores them in the morning and again in the afternoon, or has a colleague also scored them? What is the coefficient of variation? Also, do you make two control observations for every trial that you are doing? I think you should. Another point is that I understood you to say that you have done just 6-week studies. I think that, because of the menstrual cycle, you can get into trouble. We have been counting lesions during the menstrual cycle and we do find quite significant variations, which is not surprising. This type of variation could influence your data in a 6-week study.

Lidén, Umeå: In answer to your first question, we have not checked the reproducibility in the way you asked. In answer to your second question I agree that six weeks is too short a period. This duration of treatment was chosen because we got the information at the start of the study that treatment with zinc had a quite obvious effect already at four weeks irrespective of the sex of the patients.

Strauss, Iowa: I would like to answer Dr. Cunliffe's question of what you do in practise and what you do in clinical trials. What you do when you see a patient and you write your own personal chart is not the question. We are talking about clinical trials and the minute you start using the over-all global classification, which has all sorts of problems associated with it, you are entering a source of error. We are talking about clinical trials and with clinical trials I think global evaluation has limited usefulness.

Cunliffe, Leeds: Well, on this point we disagree. I wonder if anybody else has any thoughts on this?

Shuster, Newcastle-upon-Tyne: I would like very much to agree with Dr. Cunliffe. I think that at present we have not developed a good method for assessing clinical improvement in acne. People just do papule and pustule counts, because they are easy to do. What has not been shown is that these counts necessarily correlate with improvement. Until we have a method that can be unambiguously associated with improvement the only criterium that is worthwhile in acne is a gross improvement. All other levels of improvements are to be ignored in the present state of acne technology. Instead we now seem to be discussing how many angels can stand on a pin's head. So, as far as clinical trials are concerned, the methodology is such that detection of marked improvement is all that is acceptable.

Strauss, Iowa: Dr. Shuster, when you look at gross changes you are then going to limit the usefulness of method to observing the changes that can be seen with the most active agents such as oestrogens and X-ray. Many of the drugs used in acne therapy are useful, even though they may not produce as dramatic changes.

Plewig, Munich: I agree with Dr. Strauss because there are very few experts who can really discern even a weak drug.

Coming back to Dr. Lidén's work, the problem is that with the placebo you had more than a 20 to 25% improvement over six weeks. We have seen the same. If you wait 15 weeks instead of 6 weeks, how much improvement will you get with a placebo? Is it a linear response amounting to 50%—60% improvement? Comparisons are always necessary.

Lidén, Umeå: What you are referring to might be caused by the so called order effect. We have tried to assess this by examining if the counts at the first visit were distributed differently at the start, in the middle, or at the end of the study. There were no such differences as far as we could discern.

Juhlin, Uppsala: Yesterday we were looking at some patients together with Dr. Cunliffe. There was one point which came out which I think is quite essential. We look at the patients and count the number of papules and comedones. When Dr. Cunliffe looks at patients he palpates the papules with his fingers. That means that he gets about twice as many papules as we did when we performed the study. It is a small detail, but it is extremely important. If we are going to compare research from various centres, we must use the same method.

The other point I want to make is that I disagree with Dr. Leyden about the score in acne. I think that it has some use. If you have some terrible lesions on your nose and some up on the forehead, and they disappear. It is a hundred per cent good effect to me. The patients look terrible with just a few lesions sometimes and therefore they should have a high score on these lesions.

Shuster, Newcastle: I am not actually disagreeing with Dr. Strauss on the vital need to be able to detect very small changes. It is quite clear that we must be able to measure the small changes that are produced by some drugs. What I am saying is that the technology for measurement of clinical effects is so poor that we cannot use it to measure those small effects, desirable though it is. I think the first step is to go back to the clinical lab and design better methods for detecting these small changes. It is quite obvious to me that the methods we are using such as lesion counts, photography, and over-all grading are not yet good enough. That is why I believe that in the present state of the art we can only accept a gross change.

Leyden, Philadelphia: One thing that has not been discussed and is something that I think is important in the clinical trial is the question of what are the alternative reasons besides a drug failure for patients getting worse during a clinical trial? As Dr. Cunliffe has already mentioned, you have to be careful about the menstrual period. If you happen to see the patient at the end of a study and it is three days before her menstrual period, she may look a lot worse than she will a week later. Therefore, in spite of the fact that she may have had a significant improvement over-all, because you see her at that one particular time at the end of the study, she will be graded as a treatment failure. You could have people on systemic antibiotics and having a full antibiotic effect as far as its ability to suppress *P. acnes* but who are not doing well because of other reasons such as some undue stress. If you see students around exam-time they usually are worse. There are lots of reasons why people can be doing poorly at particular times that are unrelated to the drug or the placebo that is being used. One of the ways of getting around that is something that Dr. Plewig and I have had some experience with in the past and that is if you have the capability and time to see

people on a very frequent basis. When we first started our work together we used to see patients three times a week. Some patients were also commonly seen twice a week and once a week was the longest interval between visits. If patients in a study are seen very frequently, then you can get better feeling for some of the other factors that are involved.

Michaëlsson, Uppsala: I agree with Dr. Plewig. It is very essential to count the various types of lesions, but I think that is not enough. We should perhaps subdivide the various types of lesions. There are so many types of paules and pustules that it is not enough to just count the papules and pustules without dividing them into several sub-groups.