J. Soc. Cosmetic Chemists **18** 439-453 (1967)

The clinical evaluation of antidandruff shampoos

N. J. VAN ABBÉ and PATRICIA M. DEAN*

Presented at the Symposium on "Product Testing", organised by the Society of Cosmetic ehdmists of Great Britain in Eastbourne, Sussex, on 15th November, 1966

Synopsis—Assessment of the efficacy of a treatment for dandruff demands a rigidly controlled methodology, comparable to that employed in other types of clinical testing. Experimental techniques are illustrated and discussed, with emphasis on the need for careful training of observers, who may then be able to derive meaningful results from a study of human volunteers using an antidandruff shampoo.

In a previous communication (1), the study of dandruff on volunteer human subjects was described in detail, with emphasis on methods of studying the disorder itself rather than its treatment. Shampoos containing various active constituents represent the accepted means of providing medication for dandruff and several papers (2-4) report clinical investigations of these and other means of applying the active ingredient. In recognition of the need to devise techniques of scientific validity for evaluating potential therapeutic measures, this paper is intended to review some of our own experiences critically.

As before, attention will be confined principally to the manifestation of dandruff as visible scaling, i.e. desquamation in excess of and in fragments larger than that due to a simple shedding of the horny layer of the epidermis. We are not, in the present context, interested in the causation of dandruff, although this must obviously have a bearing on the clinical aspects; our

*Beecham Toiletry Division, Brentford, London.

concern is to assess the feasibility and reliability of various ways of evaluating potential treatments.

Principles of clinical testing

Dandruff is a chronic abnormality of the scalp and, as in the case of chronic diseases generally, the prognosis for any individual sufferer at a given point in time is uncertain. In other words, when a treatment is administered to a subject, it is difficult to decide whether future progress (favourable or adverse) is due to the treatment or to spontaneous changes. The pre-treatment phase shown in *Fig. 1* shows how dandruff fluctuates spontaneously; if a course of treatment was initiated at a time when the

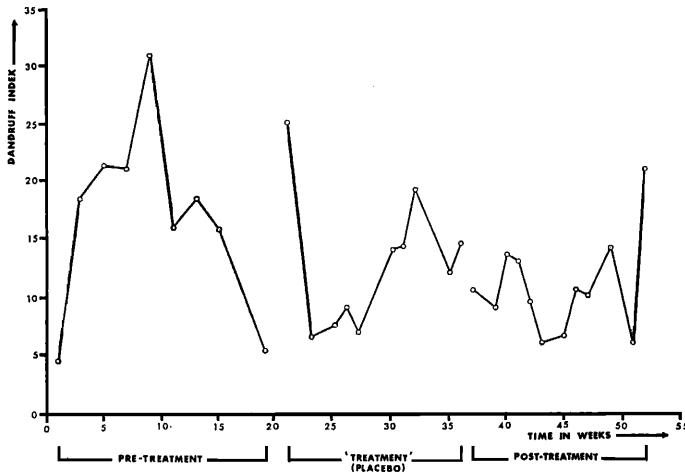

Figure 1
Results of 17 weeks' treatment with placebo shampoo on a case of moderate dandruff (Detailed method of inspection).

level was high, there would obviously be a distinct hazard that subsequent inspection would register a reduction in dandruff although this might be wholly independent of the treatment given. Fluctuation of this nature can be taken care of to some extent by using sizeable panels of subjects, so that the purely random effects tend to cancel out. However, if progress under treatment is compared only with pre-treatment levels, there is an indeterminable risk that extraneous factors may be operative during treatment, possibly having a greater influence on the dandruff levels than the treatment itself. Such an irrelevant feature could well be climatic, for an effect of this nature during dandruff treatment has been noticed previously (1).

To overcome this difficulty, it is essential to compare the progress of a treated panel with an untreated panel running concurrently.

Dandruff investigations would be simplified considerably if objective measurements were practical. A technique has, in fact, been published whereby scale samples are taken by means of a miniature vacuum cleaner and subsequently weighed (4). Since this technique may involve the disturbance of scale attached to the scalp which may influence the disorder itself and in the absence of alternatives not showing a similar disadvantage, a subjective method of evaluation is to be preferred.

Clinical trials in which subjective assessments are made always embody the risk of significant observer error and bias. Comparison between treatment and no-treatment panels tends to rule out error and bias due to the subjects themselves, especially if they are allocated to the different panels by a suitable method of randomization; but it is still vital that the observer should not be able to identify whether a subject is receiving treatment or not at the time of inspection. For various reasons it is preferable that this knowledge should also not be available to the actual subject and so the "double blind" technique should be adopted. This involves the employment of a placebo, identical in all discernible respects to the treatment but lacking the active constituents; one half of the subjects receive the treatment and the remainder have the placebo and serve as controls. Whilst it is usually possible to ensure a close resemblance between a placebo and treatment ("control" and "test") shampoo, some ingenuity of the formulator is sometimes needed to achieve this.

To make sure that the identity of test and control materials is unknown to both subjects and observers, whilst also ensuring that the observers do not know or try to guess the allocation of subjects to the various panels, it is desirable to apply several different code-letters both to the test and control products and to allocate them to the subjects by means of a randomization chart. This also helps to prevent subjects comparing notes with one another and possibly influencing their cooperation.

Elimination of bias may also be assisted by stratifying the subjects according to various criteria possibly affecting the dandruff condition, e.g. ensuring that various degrees of initial severity are equally distributed between the panels. Age and sex might well be treated similarly, though the factor next in importance to initial severity is probably the question of prior usage of a medicated shampoo.

"Cross-over" technique would theoretically help greatly to strengthen the validity of a clinical trial; that is, one panel would start with the control

product and later use the test, whilst the other panel would use the two products in reverse order. If, however, the test product has any real effect, it is vital to know how soon the dandruff level reverts to normal when treatment is discontinued, otherwise a "carry-over" effect will be operative during the control phase; for the panel starting with the test product, results during the control phase will also depend on the actual time required to regenerate dandruff by infection or other means. The post-treatment zone illustrated in *Fig. 2* shows that several weeks are required to re-establish the original pre-treatment dandruff level when an active anti-dandruff agent has been used. A practical solution might be to allow a prolonged time-lag between test and control phases, but this is likely to prolong the whole procedure of experiment to an unacceptable extent. A "cross-over" trial ignoring the points noted would tend to underestimate the efficacy of a treatment.

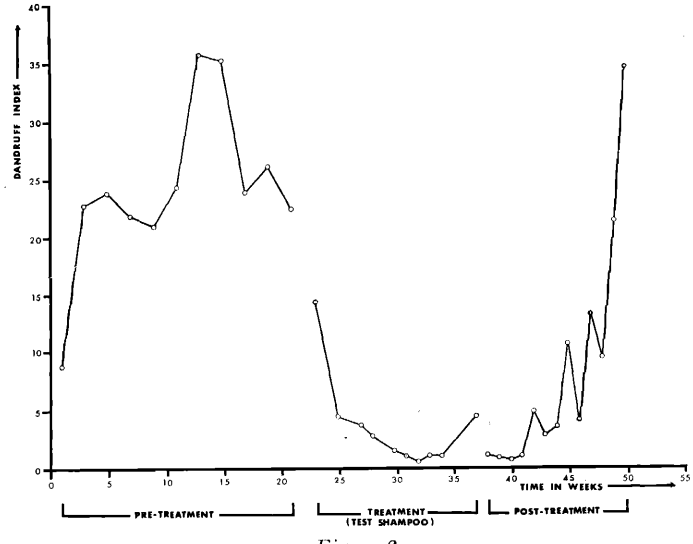

Figure 2

Results of 16 weeks' treatment with shampoo containing 2% w/w zinc bis(pyridine-2-thiol 1-oxide) on a case of moderate-severe dandruff. (Detailed method of inspection).

It is almost a truism to suggest that the statistician who will have to determine the significance of the results, should be fully consulted at the planning stage of a clinical trial. It is the authors' view that a carefully-planned and well-conducted trial on a limited panel is far more valuable than a poorly-conceived investigation carried out on a relatively vast scale; an essential feature of sound planning is to ensure that the results will be adequate and in a suitable form for statistical analysis.

Special problems of dandruff trials

One of the main considerations is the decision on what is to constitute a favourable result for an individual participant. It may, for example, be shown that any shampoo, whether medicated or not, will remove some 70% of dandruff scale during the actual process of shampooing. In the absence of effective treatment, however, dandruff will return to approximately the original level within about five days; what matters to the subject is that this recurrence should not happen, at least to the same extent. Hence an antidandruff shampoo can only be classed as effective if it significantly and consistently reduces the dandruff levels attained five days after shampooing and it is appropriate always to take measurements at this timing as far as possible. It should not be forgotten that clinical trial techniques stand or fall by the extent of cooperation achieved with the participants, and there is an ever-present hazard that some of the subjects may not really have used the test or control products at all, especially if they are unpleasant or complicated in use or deemed to be ineffective. Similarly precise timing of observations after shampooing is a target to be aimed at and it has to be hoped that minor discrepancies will cancel out between the panels.

Recruitment of subjects for dandruff studies presents surprising difficulty. Obviously someone who has no dandruff cannot show any beneficial effect, however efficacious the treatment. It is therefore necessary to recruit at least moderately severe sufferers, though it is also desirable to have some subjects who are only slightly affected in case either the test or control product demonstrates an adverse effect. It is quite impossible to ensure that a sample of subjects in a trial is fully representative of the population as a whole, if panels are to be kept within manageable proportions; for instance, the source of infection in one geographical area may differ from others and may also differ in response to particular treatments (assuming that dandruff results from infection). If the subjects for a trial are recruited from an essentially "closed" community such as a boarding school, the risk of bias due to atypical sampling may be serious. On the other hand, it may be much easier to obtain the desired level of cooperation in such a community and far simpler to make suitable arrangements for conducting inspections. Probably the most satisfactory answer in practice is to utilize several such "closed" communities for the full evaluation of a new product. Subjects who are skin patients attending hospital are not ideal for the purpose, as they may not respond in the same way as "normals".

Some published clinical trials on dandruff appear to have been restricted

to recording the overall impressions of a dermatologist, the findings being expressed in a limited range of descriptive terms such as "cleared", "improved" or "no change"(3). Brief consideration will show that such an approach (especially if it is not accompanied by a strictly "double-blind" routine) leaves much to be desired unless it is only required to distinguish all-or-none efficacy. Important factors, in our experience, are that:—

i. Training and extensive practice in the study of dandruff, making use of the concordance tests previously described (1), are pre-requisites for discriminating and reproducible assessments.

ii. Clear definitions for a gradation of clinical features are necessary, along with a definite system for examining the scalp. Observations may take the form of word descriptions at the time of examination but are preferably transcribed into pre-selected numerical values for subsequent tabulation and analysis.

iii. Intervals between inspections, during which the various treatments are used, need to be programmed on sound lines.

The technique of examining the scalp in 25 sections (1) has formed the basis of much of our experience for several years. If this is used in conjunction with a "double-blind" system of test and control panels, it is possible to compare the average trends for each panel during treatment. Since each physical examination occupies almost 30 min/subject, relatively small panels have to be employed and it becomes arguable whether a less detailed inspection of larger numbers would not be more profitable. Clearly the point could be reached where the technique of assessment became so crude that only the most glaring differences could possibly attain statistical significance; we have therefore sought to test some comparatively simple techniques and are still continuing such investigations.

Our own staff of observers, who have long experience in the detailed, 30 min method of assessing dandruff, have more recently been trained to conduct a quicker method of inspection which takes about five min. In the detailed method the scalp is partitioned into 25 imaginary sections and each section in turn is scored for severity of dandruff and proportion of area occupied by dandruff; in the rapid method the scalp is divided into four sections instead of 25. The subject is seated under a good diffused lighting and each notional quadrant is examined by parting the hair with a comb at intervals over the area. The estimated amount of scale attached to the scalp (not loose in the hair) for each quadrant is rated by the following verbal descriptions, which are later given the numerical values shown and added together to yield the index for the whole scalp.

Nil	0
Very slight	1
Slight	2
Slight to moderate	3
Moderate	4
Severe	8

As in the case of the more detailed technique, it is vital to test the concordance between observers, examining the same subjects independently at the same time, as often as possible. Typical results of such concordance tests between a pair of examiners are shown in *Table I*. Whilst these are not as good as those obtained for the more detailed method and reported earlier (1) they nevertheless do show quite good agreement and indicate that a discrepancy of more than 'one place' is unusual. Even when concordance tests show good agreement between observers it is still considered advisable that each subject in a trial should always be examined by the same observer at successive inspections; this will be specially important when some of the observers are relatively inexperienced. An observer is at no time allowed to see the subject's earlier records during an examination

Table I

Concordance between observers

		Assessment of dandruff: observer A					
		Nil	Very slight	Slight	Slight to moderate	Moderate	Severe
Assessment of dandruff observer B	Nil	21	6				
	Very slight	9	58	5	2		
	Slight	1	10	6	4		
	Slight to moderate		4	4	11		
	Moderate		1	1	4	13	3
	Severe						5

Figures represent number of scalp quadrants

and neither the observer nor the statistician know the breakdown of the product codes before the trial has been completed and analysed.

Our earlier studies to investigate dandruff without particular reference to treatment, involved either weekly or fortnightly examination of subjects over long periods of time. Similarly in clinical trials using the same detailed procedure, we have made frequent inspections over long treatment periods;

indeed, the less effective the treatment, the longer a trial needs to proceed in order to demonstrate a difference between test and control. However, with the records of earlier trials available for study and with more effective products now coming under test, it has been possible to show that the number of inspections during a trial can be greatly reduced without detriment to the validity of the main findings, providing that the panel size is not too small. To make due allowance for any cumulative effect of a shampoo treatment it must be used several times before an attempt is made to measure its efficacy. Our experience suggests that a first examination before using the product should be followed by a second examination after four weeks' use and a third examination after eight weeks. Comparison of the third versus the first reading reveals the main trend, but the second examination is a useful check on placebo effect which may be expected to be substantially the same at the second and third examinations. When using this method we prefer to use panels at least twice the size of those used for the method with more frequent examinations.

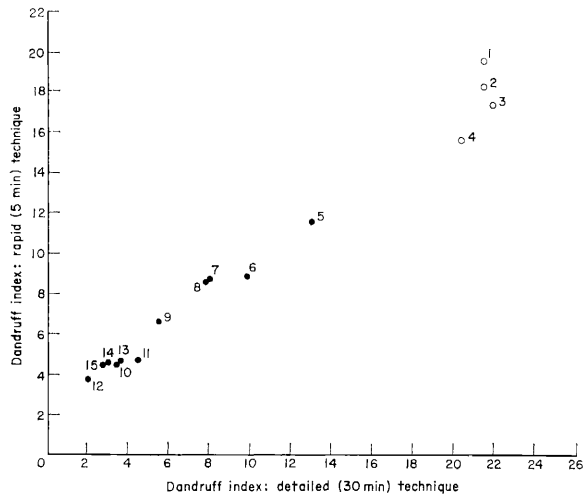

Figure 3

Comparison between "5-minute" and detailed "30-minute" technique for dandruff inspection. Numbered points represent successive weekly averages for the same 10 subjects under treatment with (1-4) a non-medicated, and (5-15) an antidandruff shampoo.

Results

Critical comparison of techniques for the assessment of dandruff ideally requires that the different methods should be used side-by-side on

the same subjects. Correspondence between the rapid (5 min) method and the original detailed 30 min method is illustrated in *Fig. 3.* For levels in the moderate to moderately-severe region, there is little doubt that the two methods are in good agreement. There is a suggestion that the rapid method may be rather more sensitive at low and high levels of dandruff, possibly because the condition is often of patchy distribution and the scores tend to be "diluted" by the zero scores for clear areas in the detailed method.

The distribution of points along the curve in *Fig. 3* reflects the efficacy of the treatment, but this is more readily examined in other ways, e.g. as

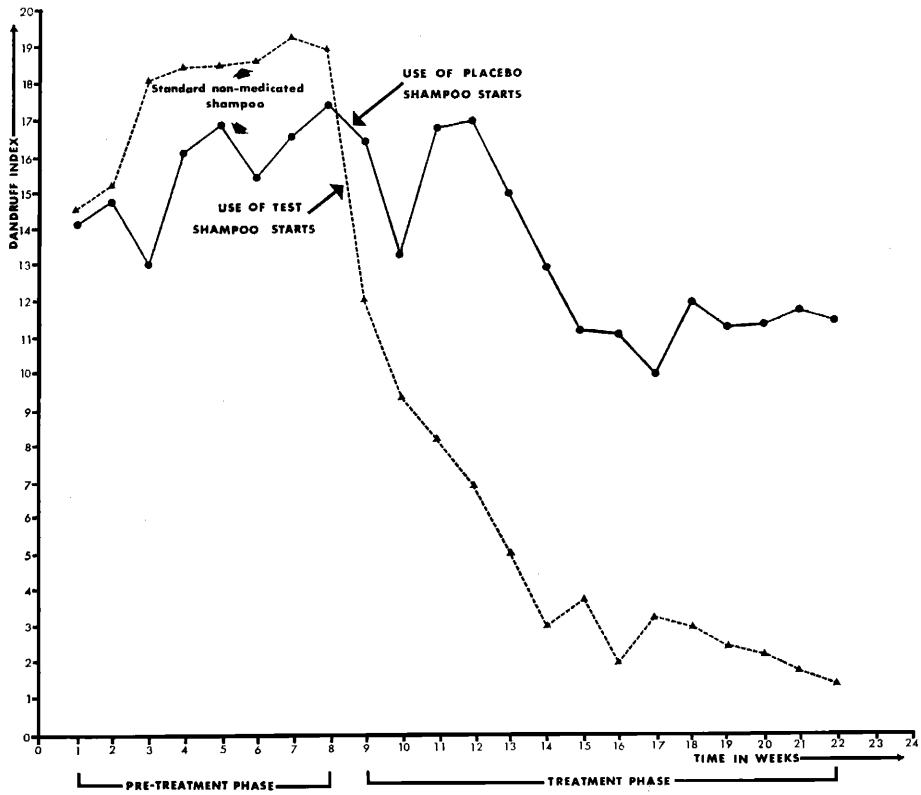

Figure 4

Trial of an anti-dandruff shampoo containing 2% w/w zinc bis(pyridine-2-thiol 1-oxide). The points represent average levels for panels of 13 subjects, assessed by the detailed method.

shown in *Fig. 4.* The frequency of examinations helps to show the speed of response to the treatment as well as the maximum benefit obtained.

Table II summarizes the data from a trial on the same shampoo, using the rapid inspection technique and the timing pattern discussed above, i.e. before treatment and after treatment for one and two months. The difference between the placebo effect and the treatment is clearly shown, though it should be remembered that expression as percentages does not necessarily accord precisely with the "true" gradations prevailing *in vivo*; the analysis of variance undoubtedly yields a more reliable basis for judging the difference between test and control.

Table II

Trial of an anti-dandruff shampoo using the rapid method of assessment and a reduced number of observations

	Placebo shampoo	Test shampoo
Average level at 1st examination for all subjects who reattended for 2nd examination	9.3	10.8
Average level at 2nd examination	7.3 (38 subjects)	5.0 (39 subjects)
% reduction	21.8%	54.3%
Average level at 1st examination for all subjects who reattended for 3rd examination	9.6	11.0
Average level at 3rd examination	7.5 (37 subjects)	3.2 (32 subjects)
% reduction	22.0%	70.9%

Table IIA

Analysis of variance of results in *Table II*

		Sum of squares	Degrees of freedom	Mean variance	Variance ratio
Examination 2v.1	Between shampoos	105	1	105	5.8*
	Within shampoos	1370	75	18	
	Total	1475	76		
Examination 3v.1	Between shampoos	313	1	313	17***
	Within shampoos	1222	67	18	
	Total	1535	68		

*Significant at 5%, ***Significant at 0.1%, confidence level

The progress made by individual subjects in a dandruff trial is well illustrated by means of a ternary diagram (*Fig. 5*) showing the cumulative effect of continued usage; this indicates in concise form not only the beneficial effect of a product but also the proportion of subjects whose dandruff level remains unchanged or increases. In this trial the product was considered to have a beneficial effect when the dandruff level after treatment was half, or less, of that before treatment.

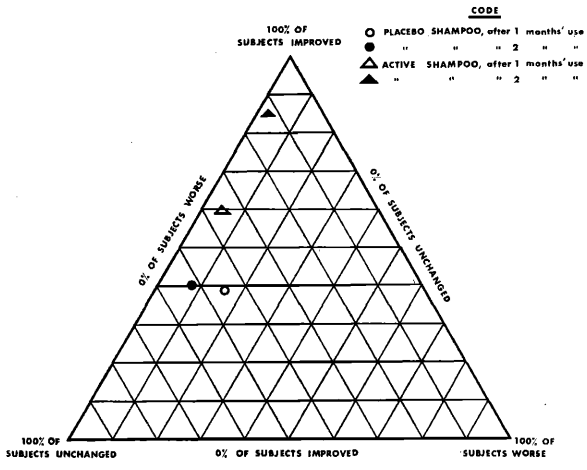

Figure 5
Proportion of subjects showing improvement, no change, or worsening of their dandruff (Rapid method of inspection).

The results shown here cannot justifiably be taken to indicate the numbers of subjects necessary in antidandruff trials generally, since this will also depend on the efficacy of the treatment under test; these numbers were nevertheless clearly adequate in the examples quoted.

Discussion

Experience has taught us that treatments to combat dandruff require to be evaluated no less stringently than medicaments for correcting other chronic disorders. Good organization and, in particular, utilizing the "double-blind" method of clinical trial are exceedingly valuable. Unfortunately, it does not seem to have been appreciated in the past that the clinical impressions of a busy consultant, undertaking a trial as an isolated experiment and with no proper controls, will scarcely do more than

confirm the obvious and confuse the more subtle differences between various treatments. Nevertheless the supervision of a dermatologist is essential to provide guidance on correct diagnosis and to safeguard the interests of the volunteer participants.

For the purpose of taking measurements, however, it is desirable to employ specially-trained observers, the reliability of whose results has been rigorously tested; such observers need not be medically qualified.

Despite the supposedly ubiquitous occurrence of dandruff, we have never found it easy to recruit large numbers of severe sufferers for trials; this may have been a blessing in disguise, insofar as it has inspired us to devise techniques applicable to the available numbers of volunteers. The clinical studies reported here should not, however, be considered to be adequate to establish fully the efficacy of the proposed formulation. It is, for example, necessary to determine whether efficacy is maintained over many months of usage or whether resistance develops; in this particular case, another experiment continuing for 12 months has, in fact, shown that reductions in dandruff levels were satisfactorily maintained but the short term studies are naturally undertaken first, primarily for screening purposes.

It is interesting to note the finding that about 10 weeks are required to re-establish starting levels of dandruff when treatment has reduced it virtually to zero. The true induction period for dandruff is probably rather less than this, however, since the treatment is likely to have exerted some "carry over" effect.

Whenever clinical trials are conducted, according to the patterns discussed here or in other ways, it is important to keep detailed records of any adverse effects (in addition to worsening of the condition under examination). These will generally be confined to mild episodes of transient erythema or itching but the supervising dermatologist will certainly wish to examine anything more than this in some detail. Comparison between test and control products on the incidence of physiological reactions should be made, to show whether irritancy is due to the active constituents or the shampoo base itself or even to establish whether the active constituents may have an anti-irritant effect.

Acknowledgements

The authors wish to thank Dr. W. G. Tillman, who supervised the work reported here as consulting dermatologist.

(*Received: 25th August 1966*)

REFERENCES

(1) Van Abbé, N. J. *J. Soc. Cosmetic Chemists* **15** 609 (1964).
(2) Spoor, H. J. *Proc. Sci. Sect. Toilet Goods Assoc.* No. 31 33 (1959).
(3) Thorne, N. *Brit. J. Clin. Pract.* **17** 357 (1963).
(4) Vander Wyk, R. W. and Roia, F. C. *J. Soc. Cosmetic Chemists*, **15** 761 (1964).

Introduction by Mr. Van Abbé

Unlike our previous paper on the subject of dandruff, we are now dealing with the clinical evaluation of treatments. It would naturally be preferable to carry out a simple laboratory test but this cannot be done with certainty and so we find ourselves in the field of human subjects and biological variation. The real difficulties arise from the fluctuating character of dandruff itself and a relatively sophisticated technique is necessary for validation of the efficiency of a medicated shampoo.

In this paper, we have dealt with the 25-area method of inspection that appeared in the last paper and with a more rapid technique where the scalp is divided into only four imaginary areas. Contrary to expectation, the quadrant method appears to be rather more sensitive. There is a general parallellism or, in fact, linearity between the two techniques, over most of the range of moderate to severe dandruff but at the level of high dandruff there is a greater spread out on the rapid method and the displacement from zero at the lower end suggests that the rapid method is also more sensitive at low levels of dandruff. We have distinguished between weekly or fortnightly examinations over a fairly long time and a programme of only three examinations, comprising before treatment, after one month's treatment and after two months' treatment. One way of representing the data obtained from the abbreviated programme of inspection is in the form of a ternary diagram which shows not only improvement of the group but changes in the direction of "worse" or "no change".

DISCUSSION

MR. C. PUGH: When one looks at the ternary diagram, one can see very clearly the proportion of people getting benefit. To my mind this is a most important quantity to examine in trying to make a better shampoo. In *Fig. 5* one can see that the placebo shows virtually no movement, but the test shampoo is moving firmly towards 100% getting better. If the trial had been continued for another month, the treated panel might have reached the apex, although this is unlikely. I would like to stress that this seems to be an excellent method of presentation giving far more valuable information than percentage improvement on an "average head".

MR. VAN ABBÉ: Although curves such as those represented in *Fig. 4* appear to show clearly what is happening week by week, they represent panel averages and they do not show the proportion getting worse or the numbers unchanged. For our ternary diagram, we have arbitrarily chosen a level of 50% improvement or 100% worsening as the level for illustration. However, we could have chosen other levels and shown the appropriate proportions. The actual points on the diagram would differ but the conclusions regarding superiority of treatment-v-placebo would not materially change.

MRS. D. L. WEDDERBURN: You could, perhaps, have omitted a control group in this trial owing to your employment of control and then test and then back on control again. Does your use of a placebo mean that you have found seasonal variations in

dandruff and that you regard it as necessary to have a control group in parallel because the incidence of dandruff fluctuates at different times of the year?

MR. VAN ABBÉ: The dandruff level of untreated subjects does not seem to be dependent on the season but we did see a sign in our earlier work that under active treatment there was a seasonal influence. Other factors might be involved too, such as holidays, examinations and various psycho-somatic influences that could make it difficult to draw conclusions without a placebo.

MISS DEAN: I should like to add that although we carry out concordance tests between examiners, this does not safeguard against all the examiners drifting in the same direction over a period of time. If we run a control panel, the observers do not know which product is control and which is treatment, so it does help to safeguard against any drift of this nature. Of course, using a placebo does put some people off; they realise they are not getting better, and sometimes a proportion of our subjects get discouraged and tend to default if we keep them too long on no-treatment.

MR. K. M. GODFREY: Is the time required to re-establish dandruff affected by residual absorption of the active ingredients?

MR. VAN ABBÉ: Possibly. Zinc omadine (the active ingredient referred to in the paper) is only soluble in water to the extent of about 6 ppm. Nonetheless it is soluble to this extent and may be absorbed or adsorbed; this may contribute towards its efficacy, for some of it may be taken up during treatment and not be eluted for some time afterwards.

MRS. S. M. LUDFORD: Do you use people again for assessment once they have recovered to their pre-treatment level; if so, do they respond differently from new subjects? Do you observe any other effects of antidandruff products or do you only look for effect on dandruff?

MISS DEAN: Occasionally we do use subjects again. We allow a considerable recovery period and, when using them again, we take great care to stratify them between the groups. We have supplemented the clinical examinations by looking for effects on *Pityrosporum* and bacterial counts on the scalp. We pass an applicator through the subject's hair and take plate counts of micro-organisms. Sometimes we find a reduction but so far we have not really demonstrated any definite correlation with dandruff scaling.

MR. VAN ABBÉ: We also record itching and erythema, again without showing any distinct correlations.

MR. K. V. CURRY: Is the response of the moderate/severe group of subjects similar to the response of the "very slight" group? I have a feeling that the majority of people with dandruff have the second type which is really a social nuisance rather than a scalp disorder.

MISS DEAN: On the whole it is easier to show a significant reduction on people who have severe dandruff and for the bulk of our panels we prefer to use such subjects. I think the proportion showing a reduction who start with slight dandruff is nearly as great as for those who start with a severe level but, of course, it is not such a great reduction.

MR. R. CLARK: Is your objection to the use of vacuum cleaner technique purely theoretical or is it based on practical experiments?

MR. VAN ABBÉ: We have carried out some experiments. We did not like the *Hair Vac* (as used by Vander Wyk) in its original state but even when this had been modified by us, we still felt that it was impossible to know whether the course of the condition was being influenced. This is the main difficulty, for one just cannot know and therefore it does not seem to be a desirable technique.

MR. R. CLARK: Did you run a "vacuum cleaner" trial in conjunction with your own subjective method of assessment?

MR. VAN ABBÉ: We have not done so yet, although we have considered the possibility.