ISSN 0037-9832

VOL 29 NO 7 JULY 1978

Journal of the Society of Cosmetic Chemists

Contents

Now her skin has a new best friend:

New Loramine F.

The Vitamin F derived alkylolamide

To our consistently successful range of alkylolamides for shampoos, detergents and cosmetics, we now add an impressive newcomer: Loramine F.

What has Loramine F got that our other alkylolamides have not? Two words, Linoleic Acid. This essential Fatty Acid, frequently termed Vitamin F, is held to be highly beneficial in maintaining the health of skin and hair.

Loramine F offers the formulating chemist in a cosmetically acceptable form the benefits of Vitamin F plus the advantages of an alkylolamide to improve the 'skin friendly' properties of the other ingredients in your products.

Pioneers in Alkylolamides

For over a quarter of a century, we at Dutton & Reinisch have been pioneers in this country of a wide range of alkylolamides. We show below the most widely used part of our range. In addition we can offer alkylolamides to your requirements.

Loramine F	linoleic diethanolamide
Loramine C212	coconut monoethanolamide
Loramine CD Loramine DC212 S Loramine DL240	coconut diethanolamide
Loramine IPP240	coconut monoisopropanolamide
Loramine L203	lauric monoethanolamide
Loramine DL203 Loramine DL203 S	lauric diethanolamide
Loramine IPL203	lauric monoisopropanolamide
Loramine DLMS	lauric myristic diethanolamide
Loramine S280	stearic monoethanolamide
Loramine DO280 Loramine DO280 S	C18 unsaturated diethanolamide
Loramine IPE280	C18 unsaturated monoisopropanolamide
Speciality alkylolamides include:	
Loramine OM101	lipstick dye solvent
Loramine U185 Loramine DU185	speciality fungicides

For samples and further information just call

Dutton & Reinisch Ltd.

Specialists in Surfactant Chemistry

London Sales: Crown House, London Road, Morden, Surrey SM4 5DU Tel: 01-543 3335 Telex: 23254
Works: Flimby, Maryport, Cumberland Tel: Maryport 3333 Telex: 64217

Germany: REWO Chemische Werke GmbH, D 6497 Steinau I Postfach 60 Tel: 06663 5071 Telex: 0493589
Spain: REWO Chimica S.A., Gran Via 17 Poligono Industrial, Montmelo Barcelona Tel: 568 1595 0442 1845 Telex: 52368
USA: REWO Chemicals Inc., Airport International Plaza, 125 Wilbur Drive, Bohemia L.I., N.Y. 11716 Tel: 516 567 0700 Telex: 960238

Journal of the Society of Cosmetic Chemists

This edition is published for

THE SOCIETY OF COSMETIC CHEMISTS
OF GREAT BRITAIN

by Blackwell Scientific Publications Ltd, Osney Mead, Oxford OX2 0EL
Hon. Editor: J. M. Blakeway
Roure Bertrand Dupont S.A. 55, Voie des Bans, 95100 Argenteuil (France)

VOL. 29 JULY 1978 No. 7

GENERAL NOTICES

Publication dates: The 'Journal of the Society of Cosmetic Chemists' is published on the 5th of each month.

Five issues for the Society of Cosmetic Chemists of Great Britain
56 Kingsway London WC2B 6DX.

Seven issues by the Society of Cosmetic Chemists
50 East 41 Street, New York, N.Y. 10017, U.S.A.

Issue No.	*Publication Date*	*Country of Origin*
1	January	Great Britain
2	February	U.S.A.
3	March	U.S.A.
4	April	Great Britain
5	May	U.S.A.
6	June	Great Britain
7	July	Great Britain
8	August	U.S.A.
9	September	U.S.A.
10	October	Great Britain
11	November	U.S.A.
12	December	U.S.A.

Advertisements: All enquiries regarding advertisements in the British Editions of the Journal should be addressed to Blackwell Scientific Publications, Osney Mead, Oxford OX2 0EL.

Subscriptions: All members of the Society of Cosmetic Chemists of Great Britain receive one copy of each edition free. Further copies at non-member rates. Industrial and non-member subscribers: £40; overseas £48. If payments are made by bank transfer, all charges shall be at the remitter's expense.

Missing numbers: **Journals are despatched at Printed Paper rate. Claims for missing numbers can be entertained only from subscribers in the country of origin of the particular issue, and must be made within 30 days from date of issue.** Members and subscribers are urged to give notice of change of address to the Publications Offices.

Responsibility for statements published: The Society of Cosmetic Chemists of Great Britain and its Hon. Editor assume no responsibility for statements or opinions advanced by contributors to this Journal.

Lectures: The Society shall have the right of first publication of any lecture or address delivered before it, but does not undertake to publish any given matter.

Manuscripts: These should be in accordance with the '*Directions for the preparation of manuscripts*', copies of which are available from the Society of Cosmetic Chemists of Great Britain. 56 Kingsway, London WC2B 6DX.

more of a bind... less of a drag

syncrowax

Syncrowax
A comprehensive range of waxes designed specifically for use in cosmetics.

Syncrowaxes in sticks can
reduce drag, bleeding, sweating and blooming
increase stability, strength, gloss and ease of application.

Syncrowaxes in creams can
produce intense white creams
confer stability, texture, emollience and water repellency.

Syncrowax BB series are also excellent beeswax alternatives in a variety of systems.

Complete the coupon for further details

Please send me further details of
Syncrowaxes for cosmetics ☐ ☐ Syncrowax BB

name

position

company

address/tel

Croda Chemicals Ltd

Cowick Hall
Snaith Goole
North Humberside DN14 9AA
Tel (0405) 860551 Telex 57601

SYNOPSES FOR CARD INDEXES

The following synopses can be cut out and mounted on 127 × 76 mm index cards for reference without mutilating the pages of the Journal.

A rapid hot-room procedure for testing the performance of antiperspirants: D. C. CULLUM. *Journal of the Society of Cosmetic Chemists* **29** 399–412 (1978)

Synopsis—Previously described hot-room methods for the evaluation of antiperspirant efficacy either take an inconveniently long time to complete or permit only a low rate of throughput of subjects. A protocol is described by which a test can be completed in 3 days. It is a shortened form of an established method, and it is shown that the shorter version satisfies the basic requirements of a valid evaluation procedure. The nature of these requirements is also discussed.

Evaluating the performance of antiperspirants: W. B. DAVIS and A. M. REES-JONES. *Journal of the Society of Cosmetic Chemists* **29** 413–422 (1978)

Synopsis—Particular attention is paid to two methods of performance evaluation (thermography and hygrometry) that allow antiperspirant activity measurements to be made under controlled conditions that do not interfere with the normal operation of the glands or the cooling caused by the evaporation of sweat at, or near, the skin surface. Thermography involves imaging the axillary vault region from the long-wave radiations emitted by the warm skin surface and uses the cooling effect of evaporation to map out areas where water is evaporating from the surface. Hygrometry involves relative humidity measurements of ambient air passing over the skin surface; sweat evaporating increases the relative humidity of the current of air. The results obtained from these methods are compared with those derived from conventional gravimetric tests on the back and the axillae in sweat is collected in occluded absorbent pads.

Toothbrush wear, brushing forces and cleaning performance: B. R. PUGH. *Journal of the Society of Cosmetic Chemists* **29** 423–431 (1978)

Synopsis—How long a brush should last is largely determined by the magnitude of the forces used during toothbrushing. Other factors are also implicated. These may include the method of brushing, the geometry of the teeth and whether the individual pre-wets the toothbrush in hot or cold water, thus affecting the mechanical properties of the filaments. Forces measured during toothbrushing range from 4 to 20 N independent of the manual strength or sex of the individual Those individuals who consistently brush with forces below 6 N do not wear out brushes. Toothbrushes having various states of wear have been examined for their cleaning effectiveness *in vitro*. No serious loss in ability to remove cosmetic stains from teeth could be observed until the brushes were severely worn. For the majority of the population, this is unlikely to lead to a noticeable increase in cosmetic stains on their teeth.

Techniques for the evaluation of emollients and keratolytics: R. MARKS. *Journal of the Society of Cosmetic Chemists* **29** 433–440 (1978)

Established and new technologies have been evaluated to assess the efficacy of emollients and keratolytics. The newer techniques include the use of (a) an instrument designed to measure *in vivo* intracorneal cohesion, (b) an instrument to assess the point penetrability of the stratum corneum *in vivo* and (c) an instrument designed to apply a standard stimulus in order to quantitate the number of squamae that are released *in vivo*. The established methods used have included surface contour analysis of skin surface replicas and morphological assessments using macrophotography and scanning electron microscopy. The keratolytics so far evaluated have included preparations of salicylic acid and urea and the emollients evaluated have included three commercially available preparations. The results thus far show that keratolytics are difficult to evaluate in entirely normal skin but that morphological assessments are best in normal skin and physical measurements are of more help in abnormally scaly skin. Emollients have proved much easier to evaluate and their effects can be detected by all the techniques described.

Evaluation of mechanical stresses set up in lipstick during application: R. G. DREW. *Journal of the Society of Cosmetic Chemists* **29** 441–446 (1978)

Synopsis—A device is described which measures the stresses set up when lipstick is applied; this device is used to assess the application properties of two ranges of commercially available lipsticks. The values obtained facilitate definition of the strength requirements of a lipstick. Knowledge of the in-use stresses will enable formulators to avoid fracture problems during lipstick application. A trial using ten commercially available lipsticks indicates that the application force is related to the ease of colouring of the lips by the applied lipstick. Application forces were greater when the lipstick was applied by older subjects.

INDEX TO ADVERTISERS

Sunburn protection
Eusolex®
oil-soluble and water soluble filters for cosmetics

Offers and technical information on request

Cosmetic ingredients

MERCK

THE QUALITY
YOU CAN TRUST

E. Merck, Darmstadt/Federal Republic of Germany

A 34 EÜ

J. Soc. Cosmet. Chem. **29** 399–412 (1978)

A rapid hot-room procedure for testing the performance of antiperspirants

D. C. CULLUM *Unilever Research Laboratory at Isleworth, 455 London Road, Isleworth, Middlesex, U.K.*

Presented at the Symposium on 'Product Evaluation', 18 April 1978, at Eastbourne

Synopsis

Previously described hot-room methods for the evaluation of **antiperspirant** efficacy either take an inconveniently long time to complete or permit only a low rate of throughput of subjects. A protocol is described by which a test can be completed in 3 days. It is a shortened form of an established method, and it is shown that the shorter version satisfies the basic requirements of a valid evaluation procedure. The nature of these requirements is also discussed.

INTRODUCTION

During the last 20 years or so, hot-room methods for testing the performance of antiperspirants have found increasing use and are now generally accepted; indeed, the OTC Antiperspirant Panel of the FDA demands a specified level of performance in either a hot-room test or an ambient temperature test for a product to be recognised as an antiperspirant (1).

The various published procedures all suffer from one disadvantage, namely that the rate of throughput of test subjects is inconveniently low, particularly for a product development programme in which the decision on what to test next day may well depend on the result of the last test. For example, the method proposed by Fredell and Read (2) required 4 weeks of daily measurements for one test. All methods involving a cross-over of treatments require at least 4 weeks, of which the middle 2 are normally a rest period during which the effects of the products administered in the first week gradually wear off. The method of Fredell and Longfellow (3) could be completed in 4 days, but hot-room sittings had to be repeated daily, so that the number of subjects was limited. The same applies to the method used at Hilltop Research Inc. (4). The method of Wooding and Finkelstein (5) involved only one hot-room sitting but required daily application of product for 5 days, so that only one test per week could be done, and the maximum number of subjects was limited to that number which the hot-room facilities could deal with in 1 day.

We required a method which would enable us to complete a test within a week, using a sufficiently large number of subjects to yield acceptably narrow confidence intervals as well as to give reasonable assurance that they resembled the population at large. Fairly arbitrarily this number was set at fifty, although we often work with fewer.

Experience had shown that with some products no appreciable antiperspirant activity was shown in a hot-room test immediately following application, while maximum

0037–9832/78/0700–0399 $02.00

effectiveness was not reached until after at least five consecutive daily treatments. It was decided that three consecutive daily treatments, followed by a hot-room sitting on the third day, would be the best compromise between the need for repeated application of products and the need to do as many hot-room sittings as possible in a week, since it would permit sittings on Wednesday, Thursday and Friday in a normal working week. The procedure was thus essentially a modification of the 'SSEM' method of Wooding and Finkelstein. It is described in detail later. The chief purpose of this paper is first to define the basic requirements of a valid performance evaluation method and then to show that, in its modified form, our 3-day test satisfies these requirements.

PRINCIPLES

THE PURPOSE OF THE TEST

It ought to be self-evident that the first thing to do in devising any kind of test is to define precisely the information the test is required to produce. In product development, performance tests normally serve one of three purposes: to provide support for advertising claims; to assess the extent to which users will be able to perceive that the product works; and to tell product development scientists whether they are moving in the right direction.

In practice, 'the right direction' will usually be defined in terms of either advertising support or user perception, so there are really only two purposes. It is unlikely, though not impossible, that one fixed procedure will serve both of these. The hot-room stimulus and the methods used for assessing sweat output are artificial, and the only purpose such a test can serve is to rank products in order of efficacy under an arbitrary set of conditions. For reasons to be discussed later this will not necessarily be the order of efficacy perceived by users.

MODE OF APPLICATION OF PRODUCTS

In recommending a test procedure we may prefer on the one hand complete control over all the variables, which does not resemble what happens in real life, or on the other hand complete realism, which will entail considerable variation in modes of application of the product, applied stimuli, and parameters and criteria of efficacy. Usually we shall strike some kind of compromise between these extremes.

The question of how best to apply the products is a particularly difficult one, yet it is surprising how little attention it has received in the published literature. Its supreme importance lies in the fact that when two different products are compared, either in separate tests against a common control or directly in one test, the result in the sense of a decision as to which is better may be dictated by the amounts applied. Although the hot-room test is artificial and can achieve no more than a ranking of products, it is surely implicit that we are interested in the ranking order under conditions of actual use. The problem is most easily seen when a comparison is made between products of different types, for example a roll-on and an aerosol. It we apply eight strokes of the roll-on and a three-second spray with the aerosol, perhaps the aerosol will appear better. If we give twelve strokes and a two-second spray, perhaps the roll-on will appear better. Indeed, such a comparison has no discoverable meaning unless the products are applied in a manner which closely resembles the mode of application in real use.

An obvious solution to the problem is to find out by experiment the mean weight of each product used by a large assembly of subjects, and to apply that weight to each experimental axilla. This is not very practicable, since it would have to be done for every product tested. It is also scientifically wrong. The amount of roll-on used in real life is markedly influenced by the topography of the user's axillae, and this can vary very considerably. The mean amount can be grossly excessive for subjects with small, deep axillae, and not nearly enough for subjects with broad, flat axillae. The same is probably true, perhaps to a lesser extent, for aerosols. The application of the mean amount to all axillae is therefore no more realistic than the application of any other constant amount.

Another difficulty is that although the ratio of the mean amounts of the two products may be known, it is not justifiable to assume that all users would apply them in that ratio; one subject might use a lot of roll-on and a little aerosol while another might do the opposite.

The same considerations apply even when the two products are of the same type. Aerosols differ in their non-volatile content, discharge rate, coldness and other characteristics, and real users do not use a constant spray time nor a constant discharged weight. The use of either of these in a test method carries the risk of giving an unfair advantage to one or other of the products.

The most realistic procedure would be to allow the subjects to apply the products themselves. This carries the obvious risk of increased scatter in the results and a consequent need for larger numbers of subjects. On the other hand, any other procedure carries the risk of generating ranking orders different from those experienced in actual use.

The method to be described employs a two-second spray applied by a member of the laboratory staff, when aerosol products are being tested. This is justified by the admittedly tenuous arguments that it is what most other laboratories do and it is what the manufacturers of most aerosol products recommend on the pack. Also from our own crude estimates it seems to be a reasonable approximation to the normal practice of a majority of users. For the infrequent experiments in which products of different types are compared, we have adopted the practice of allowing the subjects to apply the products themselves. So far we have too little experience to be able to weigh the advantages and disadvantages of doing so.

What must always be borne in mind is that the result found in an individual test is not necessarily the same as that which would have been found had the test been conducted under different conditions. When we say, 'Product A was $x\%$ better than product B', we always imply, 'under the particular conditions of our test'.

CHOICE OF PARAMETER TO BE MEASURED

The primary measurement is, of course, the number of milligrams of sweat secreted by an axilla during a specified collection period. The parameter of efficacy which this is used to assess may, however, be one of several things. In making a choice it is essential to bear in mind the function of a control group or substrate in any experiment: the function of the control substrate is to provide the best possible estimate of how the test substrate would have responded if it had received the control treatment instead of the test treatment. It follows that the two substrates (groups) must be alike with respect to the parameter to be measured. This means that they must respond in the same way when they both receive no treatment and when they both receive the same treatment. The parameter

selected must also be as constant as possible, otherwise the random variations which must inevitably occur may be so great as to make it impossible to fulfil these conditions.

The parameters to be considered include but are not necessarily limited to the following: sweat weight from an individual axilla; sweat weight from an individual subject; ratio of sweat weights from right and left axillae of an individual subject ('R/L ratio'); mean sweat weight, geometric or arithmetic, for a group of axillae or subjects; and mean R/L ratio, geometric or arithmetic, for a group of subjects.

The best parameter to use is the one which shows the least variation. In practice the choice may be influenced by other constraints.

It is essential to appreciate that none of these parameters is necessarily among those which the user employs to assess efficacy. For example, common sense suggests that in normal use the consumer can observe only inefficacy. That is, if she perceives that she is not sweating, she cannot tell whether her antiperspirant is working or whether she would not have been sweating if she had not used it. If she perceives that she is sweating, however, she is likely to interpret that fact as evidence that her antiperspirant is not working. Secondly, she will observe that she is sweating through some signal such as wet patches on her clothing. The connection between presence or absence of wet patches on clothing and variations in the number of milligrams of sweat on absorbent pads may be a tenuous one. The user's judgement may also be influenced by the various sensory impressions she gains when she applies the product, and by the characteristics and persistence of the perfume. There is therefore no *a priori* reason for supposing that antiperspirant efficacy as measured in the hot-room will correlate with efficacy as perceived by the user, because the parameters employed are necessarily different and may be quite unrelated.

PROTOCOL

The protocol, that is the manner in which the method of measurement is to be applied, is determined by the parameter to be measured, and by any ancillary information which may be required. For example, antiperspirants based on aluminium chlorhydrate often show only a modest effect in tests conducted up to 24 h after the first application (6). With daily or twice-daily application, the effect increases and reaches a plateau after a period which may be as long as 14 days. If it is desired to observe this build-up, repeated hot-room sittings will be necessary, but for routine testing of development products it may be adequate to take only one observation. In this case the result will be influenced by the number and frequency of product applications preceding the test, and a test done after three daily applications will be neither better nor worse than one done after eight twice-daily applications; it will merely be different. Cross-over designs have been advocated to minimise the effect of side bias. If this type of design is used it is necessary to allow at least 2 weeks to elapse between the two halves of the test to permit the effect of the first set of treatments to disappear. Provided the number of subjects is large enough, application of each treatment to equal numbers of left and right sides will minimise side bias to an acceptable extent without the need for a cross-over. Just how many subjects are enough depends not only on the level of statistical discrimination required but also on the need to use a sample reasonably representative of the population at large. It might be possible to devise a procedure which would discriminate significantly between two products whose efficacies differed by only 5% with the use of say twelve subjects. Twelve subjects, however, would be unlikely to constitute a reliably repre-

sentative sample, in which case repeating the test on another group might give a very different but equally precise result. The sample should be big enough not only to give acceptably narrow confidence limits, but also to give reasonable certainty that the result obtained is typical of the population at large. The taking of a single observation on each subject obviously maximises the sample size for a given number of hot-room sittings.

THE METHOD

Subjects. Female subjects are used exclusively, because male subjects are not available in substantial numbers during normal working hours, and because a majority of antiperspirant users are women. Before being accepted on to the test panel, each subject undergoes a screening test to determine whether her sweat output in the absence of antiperspirant treatment is sufficient for her to be useful. The lower limit of acceptability has been arbitrarily set at 100 mg per axilla over a 20 min collection period. This test is preceded by a 2 week period during which the subjects are supplied with an alcoholic deodorant product. Once they have been accepted, panellists receive a continuous supply of the same product for home use.

Frequency of participation. No subject is used more frequently than once every third week. This is to permit the effect of test products to disappear.

Protocol. Subjects attend for three consecutive days. They are treated with the products once daily and on the third day take part in a hot-room test. No cross-over is used. Normally up to fifty subjects are used for each test.

Products. Two products are used per test, designated the test and control product respectively. The control product may or may not be without antiperspirant effect, i.e. a true placebo.

Product application. The products are applied by a member of the hot-room personnel. For aerosols a 2 sec spray is given from a distance of 6 inches. Solutions are applied by pump-spray. With roll-on products the subjects are normally asked to apply the product themselves. Alternatively an attempt is made to apply 200 mg of product per axilla, but the size and topography of individual axillae cause considerable variation in the amount actually deposited. Half the subjects, selected randomly, receive the test product on the left side and the control product on the right, while the other half receive them the other way round.

Hot-room procedure. Before each subject enters the hot-room, a pair of unweighed absorbent cotton pads (the 'A' pads) are placed in her axillae. She then sits in the hot-room for 40 min to allow her sweating rate to approach equilibrium. The A pads are removed, the axillae are blotted with tissues and a weighed pair of pads (the 'B' pads) are placed in position. After 20 min these are removed and put into airtight polythene jars. The axillae are again dried with tissues and a second weighed pair of pads (the 'C' pads) are applied. After a further 20 min these are removed and put into airtight polythene jars. During the 80-minute hot-room sitting subjects may have hot or cold drinks on request, and are permitted any activity compatible with retaining the absorbent pads in the axillae. At the end of the session, subjects shower and go home.

Hot-room conditions. The hot-room is maintained at $40 \pm 1°C$ and $40 \pm 2\%$ relative humidity. Under these conditions a mean sweat weight of 400–500 mg per pad can be consistently obtained in the absence of antiperspirant treatment.

Analysis of data. The results are analysed by computer, which calculates the value of

$$100\left(1 - \frac{\text{geometric mean sweat weight per pad, test product}}{\text{geometric mean sweat weight per pad, control product}}\right)$$

and the significance of the difference between this value and zero, together with certain more detailed information. The value of a similar function using arithmetic means, corrected for side bias, is also calculated, at present manually. In general there are only small differences between the percentage reductions obtained from geometric and arithmetic means. Substantial differences indicate some peculiarity in the distributions of sweat weights.

DEMONSTRATION OF THE VALIDITY OF THE METHOD

PLACEBO *v.* PLACEBO EXPERIMENTS

It is first necessary to show that the method does not yield spurious effects in the absence of antiperspirant treatment, i.e. that the test and control arrays of axillae behave in the same way when both are treated with a placebo.

The preliminary work involved a test on sixteen subjects who were treated daily with an alcoholic deodorant on both axillae. They took part in hot-room sittings daily for 5 consecutive days. After an interval of 2 weeks the whole test was repeated. Two subjects failed to attend on the first day of the first week, and one of them failed to attend on the first day of the second week. The individual pad weights are those illustrated in *Fig.* 1. These 10 days may be regarded as ten individual mini-tests, and were analysed as follows. Using an electronic random number generator, the subjects were randomly assigned to Group 1 or Group 2. This was done separately for each of the 10 days. The subjects of Group 1 were deemed to have received the test product on the left axilla and the control product on the right axilla. The subjects of Group 2 were deemed to have received them the other way round. The ratio of the geometric mean sweat weights from 'test' and 'control' axillae was then calculated. A value of, say, 0·95 meant that the 'test' product caused a 5% reduction in sweat output. The results are shown in *Table I.*

Table I. Calibration test of placebo *v.* placebo

Week 1		Week 2	
Day	Ratio $\frac{\text{GM 'test'}}{\text{GM 'control'}}$	Day	Ratio $\frac{\text{GM 'test'}}{\text{GM 'control'}}$
1	0·884	1	1·041
2	0·915	2	0·983
3	1·078	3	1·037
4	1·026	4	0·923
5	1·009	5	1·109

The geometric mean of these ratios is 0·998 and their arithmetic mean is 1·000. This is to some extent fortuitous; if the random allocation to groups had turned out differently the results would have been different. In other words if the analysis were repeated the results could be either better or worse. In general a deviation from unity of at least $\pm 0{\cdot}15$ would have been necessary for the difference between 'test' and 'control' means to be significant at the 5% level.

A more satisfactory experiment was carried out later on twenty-eight subjects, who were randomly allocated to Groups 1 and 2 before the test. The mean sweat weight from the 'test' axillae was 4% greater than that from the 'control' axillae. This difference did not approach significance ($t=0{\cdot}83$, $P<0{\cdot}25$).

EXPERIMENTS WITH THE SAME ACTIVE ANTIPERSPIRANT ON ALL AXILLAE

It is equally necessary to show that the two arrays of axillae respond to the same extent when both are treated with the same active antiperspirant, i.e. that no spurious effect is observed when test and control sides receive the same active treatment. This has been necessary on four occasions in normal product development programmes. The results are shown in *Table II*. The same product was used on all axillae in the first two tests, and a different product in the second two.

In no test did the difference between the two mean weights approach significance.

Table II. Calibration test of active control *v.* active control

Test No.	No. of subjects	GM wt. 'test' (g)	GM wt. 'control' (g)	% difference	*t*	*P*
23A	49	0·3780	0·3940	−4	1·05	>0·25
75	24	0·3568	0·3566	0	0·01	>0·25
82	22	0·2936	0·2800	+5	0·53	>0·25
102	26	0·1864	0·1952	−5	0·80	>0·25

SYSTEMATIC BIAS AND ELIMINATION OF THE 'SIDES' EFFECT

It is also necessary to show that the method is free from systematic bias. This is self-evident from the experimental design: half the subjects receive the test product on the left side and the control product on the right, and the other half receive them the other way round. Since subjects are assigned at random to the two treatment groups, any error resulting from unbalanced side effects will be random. In mathematical terms, the parameter determined is

$$\frac{\text{Geometric mean weight, test product}}{\text{Geometric mean weight, control product}} = \sqrt{\frac{T_L \; T_R}{C_L \; C_R}}$$

where T_L, geometric mean weight from left axillae treated with test product; T_R, geometric mean weight from right axillae treated with test product; C_L, geometric mean weight from left axillae treated with control product; and C_R, geometric mean weight from right axillae treated with control product.

If we consider in isolation the group which receives the test treatment on the left side, the ratio T_L/C_R is a biased estimate because left and right sides do not normally yield the same weights of sweat in the absence of treatment. We can, however, determine the

right : left ratio R_1 (at least hypothetically) and apply a correction, so that the 'true' (i.e. unbiased) ratio T_1/C_1 is given by

$$\frac{T_1}{C_1} = \frac{R_1 T_L}{C_R}, \quad \text{or} \quad \frac{T_L}{C_R} = \frac{T_1}{C_1 R_1}$$

Similarly

$$\frac{T_R}{C_L} = \frac{T_2 R_2}{C_2}$$

the correction factor R_1 appearing in the numerator because the products were applied the other way round. Combining and substituting,

$$\sqrt{\frac{T_L T_R}{C_L C_R}} = \sqrt{\frac{T_1 T_2 R_2}{C_1 C_2 R_1}}$$

That is, our experimentally found test : control ratio is biased to the extent of $\sqrt{R_2/R_1}$. Since the two groups are randomly selected, R_2/R_1 will vary randomly about a mean value of unity and there will be no systematic bias in the observed result.

It is another matter whether the error resulting from the 'sides' effect is reduced to an acceptable degree by the allocation of products as described, without cross-over. Let us consider the placebo *v.* placebo experiments described above, in which the same anti-perspirant was applied to 'test' and 'control' axillae. In these cases the true effect was by definition zero, and any observed effect is an error.

The observed percentage effect is given by

$$100\left(1-\sqrt{\frac{T_L T_R}{C_L C_R}}\right) = 100\left(1-\sqrt{\frac{T_1 T_2 R_2}{C_1 C_2 R_1}}\right)$$

Since the unbiased ratio $\frac{T_1 T_2}{C_1 C_2} = 1$ by definition, this reduces to

$$100\left(1-\sqrt{\frac{R_2}{R_1}}\right)$$

The observed deviations from zero effect are thus direct measures of $100\sqrt{(R_2/R_1)}$. Only in a few of the mini-tests on sixteen subjects did the error resulting from the 'sides' effect exceed 5%, and the method of allocating products to sides may be considered to minimise this effect to an acceptable degree.

POSITIVE EFFECTS AND REPRODUCIBILITY

It is of course also necessary to show that the method succeeds in demonstrating a difference between products when there is one, and that the value found for this difference is reasonably reproducible.

This has been done in four sets of experiments. In the first three, an active anti-perspirant was tested against a placebo (an alcoholic deodorant). In the other three pairs of tests, different active antiperspirants were tested against another antiperspirant as control. The control was the same in all three pairs. The results are shown in *Table III.*

In all cases the replication of the result is very satisfactory. The near-constancy of the mean control weight in the experiments with large numbers of subjects suggests that these groups of subjects were representative of the population at large.

Table III. Results of testing an antiperspirant *v.* placebo (treatment 1) and three antiperspirants against one antiperspirant control (treatments 2, 3 and 4)

Test No.	Treatment No.	No. of subjects	GM wt., test (g)	GM wt., control (g)	% difference
23E	1	48	0·2590	0·4240	−39
92	1	28	0·2156	0·3348	−35
96	1	22	0·2257	0·4079	−45
39	2	34	0·1624	0·1925	−16
100	2	42	0·2622	0·3185	−18
110	3	57	0·2096	0·2823	−26
112	3	80	0·2317	0·3058	−24
37	4	46	0·1524	0·2138	−29
68	4	56	0·2279	0·3080	−26

PRECISION

Finally it is very desirable to have some idea of the intrinsic precision of the procedure, i.e. to be able to estimate the number of subjects required to give a confidence interval of prescribed width. Since the test and control populations of sweat weights are approximately normally distributed, we can express the relation between the 95% confidence interval Δ and the number of subjects n in the form of the equation of Student's t:

$$t = \frac{\Delta\sqrt{n}}{S} \doteq 4, \text{ whence } \Delta \doteq \frac{4S}{\sqrt{n}}$$

where S is the pooled standard deviation expressed as the coefficient of variation. This is not at all rigorous, but it leads us to expect an inverse relation between the width of the 95% confidence interval and the square root of the number of subjects. By empirical analysis of the results of sixty-nine tests, we arrive at the relation

$$n=(120/\text{required } 95\% \text{ confidence interval})^2$$

or

$$95\% \text{ confidence interval}=120/\sqrt{n}.$$

Table IV shows how well this works in practice. 'Predicted n' is the number of subjects calculated from the actual mean confidence interval, and 'Predicted C.I.' is the width of the 95% confidence interval calculated from the actual mean number of subjects.

Table IV. Relation between number of subjects and width of 95% confidence interval

No. of subjects	No. of tests	Mean n	Mean C.I. (%)	Predicted n	Predicted C.I. (%)
20–29	37	24·8	24·6	24	24·1
30–39	8	33·0	20·7	34	20·9
40–49	10	42·8	19·1	39	18·3
50–59	9	52·7	15·7	58	16·5
60–69	1	68·0	12·6	91	14·6
70–79	2	75·0	12·9	87	13·9
80–89	2	84·0	11·7	105	13·1

It can be seen that the relation works well up to sixty subjects, but for larger numbers tends to overestimate. This may be due to the very small number of tests on more than sixty subjects. The accuracy of prediction in individual tests is subject to considerable variation, however, depending as it does on how lucky one is with respect to the distribution of sweat weights in individual groups of subjects.

BACKGROUND DATA AND THEIR IMPLICATIONS

The greater part of the necessary background information came from an experiment, described earlier, in which sixteen women, who had used only an alcoholic deodorant for the preceding 2 weeks, took part in hot-room sittings on 5 consecutive days, during which they received one daily treatment with the same alcoholic deodorant on both axillae. After an interval of 2 weeks they returned and repeated the same experiment. Since all the axillae had received the same treatment it was legitimate to call any arbitrarily selected group the test group and the remainder the control group.

DISTRIBUTION OF SWEAT WEIGHTS AND THEIR LOGARITHMS

Since each subject yielded four pad weights, the maximum possible number of single observations was $16 \times 10 \times 4 = 640$. In the event hot-room appointments were missed on three occasions, so that the actual number of pad weights recorded was 628. *Figs* 1 *and* 2 show histograms of the actual weights and their logarithms to the base 10.

Clearly, the histogram of weights is positive skewed, but equally clearly, the histogram of log weights is negatively skewed. Thus, whilst it is not rigorously correct to apply to the population of weights statistical tests which depend on a Gaussian distribution, it is not rigorously correct to apply them to the population of log weights either. In practice neither population deviates very seriously from normality, and the inevitable errors will be of no great consequence whichever population is used, provided always that the distributions obtained in a particular experiment are of reasonably smooth and regular shapes. In practice, in a test on thirty or forty subjects, yielding twice that number of pad weights for each treatment, the distributions are quite often very irregular and in such

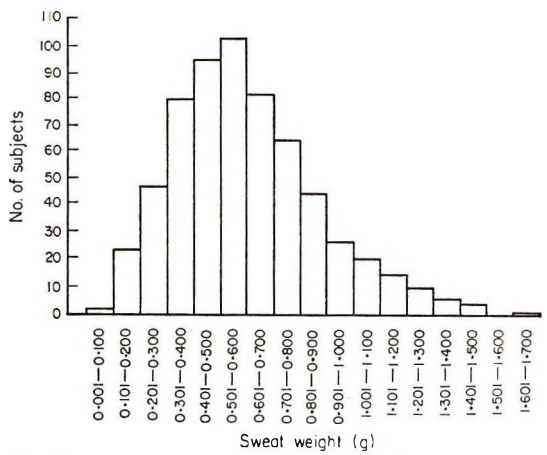


Figure 1. Distribution of sweat weights. Sixteen subjects × ten sittings × four pads per sitting; three subject-sittings missed; total observations, 682.

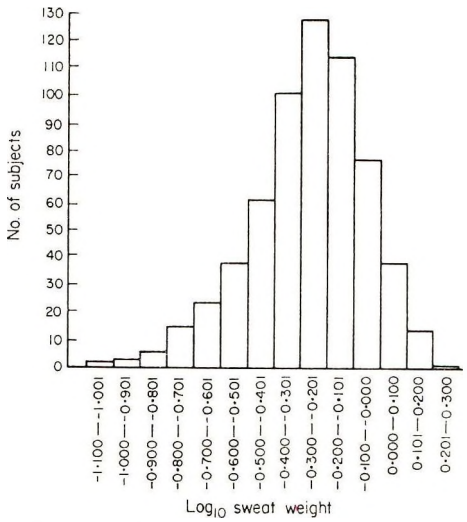


Figure 2. Distribution of $\log_{10}$ sweat weights, original data as for Fig. 1.

cases the sample standard deviation will necessarily be an unreliable estimate of the population standard deviation, regardless of whether the actual weights or their logarithms are used.

Over a large number of tests we have found that the percentage reductions in sweat weight calculated from geometric and arithmetic means rarely differ by more than 2 or 3%, and decisions based on the results would be the same whichever value was used. Major differences have been few, and have always proved to indicate a peculiar distribution of either test or control sweat weights.

It would seem prudent to consider the practical advantages as well as the mathematical rigour of using arithmetic or geometric means. Wooding (7) has shown that the distribution of log weights deviates less from normality than the distribution of the weights themselves, and has argued that it is therefore more correct to use geometric means. With present levels of product performance there is probably no great advantage or disadvantage in doing so. If the time ever comes, however, when products will reduce the sweat weight in at least a few axillae to zero, the use of geometric means will produce a problem. Since the geometric mean of any array of numbers which includes zero is zero, adherence to the use of geometric means would then either lead to the manifestly absurd conclusion that 100% reduction had been achieved when only one test axilla yielded zero weight, or require rejection of the very results which most strikingly demonstrated the success of the product. This may be in the far distant future, but nevertheless is food for thought.

RATIOS

The same data were used to study the variation between subjects and within subjects of the ratio of sweat weights from right and left sides. The object was to discover whether this ratio was sufficiently constant in the absence of antiperspirant treatment to serve as a parameter for the measurement of efficacy. Presentation of full individual results would involve tables containing several hundred figures and it is hoped that the following digest will sufficiently illustrate the main points.

First the ratio of weights from right and left axillae was calculated for each individual pair of pads. To assess the between-subject variation, the mean ratio was calculated for B and C pads separately for each of the 10 days of the test, together with its standard deviation. To simplify assessment of constancy the coefficient of variation (CV), i.e. the standard deviation expressed as a percentage of the mean, was also calculated. The results are summarised in *Table V.*

Table V. Between-subject variation in R : L ratio

		Day 1		Day 2		Day 3		Day 4		Day 5	
		B	C	B	C	B	C	B	C	B	C
Week	Mean	1·11	1·04	1·20	1·00	1·12	1·06	1·21	1·19	1·07	1·15
1	S.D.	0·30	0·23	0·45	0·24	0·29	0·28	0·29	0·36	0·18	0·20
	CV	27%	22%	38%	24%	26%	26%	24%	30%	17%	18%
Week	Mean	1·08	1·04	1·05	1·09	1·06	1·10	1·07	1·12	1·13	1·08
2	S.D.	0·35	0·20	0·27	0·27	0·27	0·20	0·24	0·26	0·21	0·26
	CV	32%	19%	26%	25%	25%	19%	22%	23%	18%	24%

The mean coefficient of variation was 25% in Week 1 and 23% in Week 2. The lowest individual value of the R : L ratio was 0·57 and the highest 2·39.

To assess the within-subject variation the mean ratio, with its standard deviation and coefficient of variation, was calculated for the five B and five C pads separately for each week. For these figures the mean coefficient of variation was 15% in Week 1 and 12% in Week 2. The variation within one subject on different occasions is therefore appreciably less than the variation between subjects on one occasion. In practice, however, one uses the mean value for the B and C pads together, and for simplicity *Table VI* shows only the R : L ratio for the combined B and C pad weights.

Table VI. Within-subject variation in R : L ratio

Subject	Week 1					Week 2				
No.	Day 1	Day 2	Day 3	Day 4	Day 5	Day 1	Day 2	Day 3	Day 4	Day 5
53	1·08	1·12	1·12	1·00	1·21	1·13	1·13	1·28	1·21	1·21
38	0·72	0·99	1·29	1·89	1·03	0·77	0·77	0·90	1·04	0·84
339	0·95	0·90	0·83	0·99	1·04	0·94	0·81	0·85	0·89	0·93
42	1·18	0·91	1·22	1·32	1·32	0·97	1·21	0·72	0·87	1·09
350	—	1·03	1·20	1·08	1·04	—	1·78	1·48	1·76	1·75
208	0·83	0·76	0·76	0·83	0·79	0·78	0·80	0·85	0·77	0·77
163	0·91	0·98	1·04	1·05	1·13	1·01	1·14	1·25	1·31	1·16
338	0·86	1·68	0·95	1·28	1·00	0·91	0·88	0·93	0·92	1·20
194	1·13	1·36	1·51	1·41	1·27	1·80	0·96	1·07	1·03	1·11
275	1·12	1·28	0·96	1·61	1·24	1·06	1·25	1·22	1·05	0·96
7	1·33	1·15	1·38	1·39	1·41	1·17	1·24	1·37	1·37	1·25
138	1·04	1·06	1·05	0·99	1·14	1·01	0·97	1·02	1·10	1·07
248	1·78	1·75	1·64	1·47	1·38	1·40	1·27	1·23	1·27	1·20
80	1·14	1·05	1·06	1·28	1·00	1·08	0·93	1·05	1·02	0·93
285	0·97	0·79	0·74	0·75	0·96	1·17	1·06	1·16	1·00	1·10
322	—	0·80	0·75	0·91	0·81	0·78	0·97	0·95	1·02	1·15

It is not the purpose of this paper to enter into contention with other evaluation scientists; two conclusions are, however, inescapable if these sixteen women are accepted

as being reasonably representative of the population at large. The first is that since ninety-nine of the 157 ratios given in *Table VI* differ from unity by 10% or more, in either direction, the ratio of weights from treated and untreated sides cannot be regarded as a satisfactory estimate of the efficacy of a product in an individual subject. The draft report of the OTC panel (1) advocates the use of the proportion of subjects in which this very parameter has a value of 0·80 or less as a measure of the efficacy of an antiperspirant. It is obvious that, given a reasonably large number of subjects, this proportion will increase as the efficacy of the treatment increases. Nevertheless it cannot be too strongly emphasised that the ratio of weights from treated and untreated sides does not, by itself, give any indication of the efficacy of a product in an individual subject.

Majors and Wild (4) recognise this fact and in their method of analysis they seek to correct for the error by measuring the change in the ratio from the value in the absence of any treatment to the value when one side is treated with the test product, i.e. the no-treatment value in an individual is used as a control. Inspection of *Table VI* shows that the value of the ratio on one day is an unreliable estimate of its value on a subsequent day. In other words the 'pre-test' ratio is not a good estimate of what the 'post-test' ratio would be if the treatment had no effect, i.e. it does not adequately fulfil the role of a control. The second conclusion, therefore, is that the comparison of pre- and post-test ratios is not a valid means of assessing the efficacy of a product in an individual subject. In the absence of a real antiperspirant effect, however, one would expect half the post-test ratios to be greater than the corresponding pre-test ratios and the other half to be less. A significant shift from a 50 : 50 split is therefore valid evidence of an effect.

SUMMARY AND CONCLUSIONS

The basic criteria which an evaluation procedure must satisfy are these. It must not show spurious effects in the total absence of an effect, i.e. the test and control substrates must behave in the same way when neither receives any treatment. It must not show spurious differences in the presence of a constant effect, i.e. the test and control substrates must behave in the same way when both receive the same active treatment. It must succeed in showing an effect, and must do so with acceptable reproducibility, when test and control substrates receive treatments of different efficacies. It must be free from systematic error or bias. The test conditions and protocol must be designed in such a way that the results contain the desired information.

We have shown that a three-day adaptation of the 'SSEM' method of Wooding and Finkelstein satisfies these criteria and that the elimination of bias has been satisfactorily accomplished without the need for a cross-over design.

Because the simple ratio of sweat weights from test and control sides is an unreliable index of the efficacy of a product in an individual subject, we have avoided the use of the parameter advocated by the FDA's OTC Panel, i.e. the proportion of subjects in which the per cent effect exceeds 20%. In every other detail of protocol and experimental procedure, however, the method fulfils the requirements of that body, as set out in its Draft Report (1).

In other respects the method is neither better nor worse than other hot-room procedures. It is certainly not in any sense definitive; it yields one particular kind of information, namely the efficacy of one product relative to another after three consecutive once-daily applications and under a particular set of experimental conditions. A greater or smaller number of applications, or an increased frequency of application, or different

experimental conditions, will yield different results in terms of percentage reduction in sweat weight. All would seem to be equally valid; they are simply different kinds of information. All hot-room methods serve to rank products in order, *under their own particular experimental conditions*, and all involve the tacit assumption that the ranking order in real use will be substantially the same.

ACKNOWLEDGMENTS

The author wishes to thank Mrs D. M. Minton and Miss L. Taylor for carrying out the hot-room tests, and various members of the Computing and Statistics Section, Unilever Research Laboratory, Isleworth, for much help and advice as well as routine analysis of results.

REFERENCES

1 Tentative Findings of the OTC Antiperspirant Panel: Information Copy, Draft Report ep. 130 (see particularly pp. 168 and 168B).
2 Fredell, W. G. and Read, R. R., *Proc. Sci. Sect., Toilet Goods. Ass.* No. 15, 23 (1951).
3 Fredell, W. G. and Longfellow, J. Report on evaluating antiperspirant and deodorant products. *J. Soc. Cosmet. Chem.* **9** 108 (1958).
4 Majors, P. A. and Wild, J. E. The evaluation of antiperspirant efficacy: influence of certain variables. *J. Soc. Cosmet. Chem.* **25** 139 (1974).
5 Wooding, W. M. and Finkelstein, O. A critical comparison of two procedures for antiperspirant evaluation. *J. Soc. Cosmet. Chem.* **26** 255 (1975).
6 Tentative findings of the OTC Antiperspirant Panel: Information Copy, Draft Report, p. 144.
7 Wooding, W. M. Interpretation of gravimetric axillar antiperspirant data. *Proc. Joint Conf. Cosmet. Sci.* 91 (1968).

J. Soc. Cosmet. Chem. **29** 413–422 (1978)

Evaluating the performance of antiperspirants

W. B. DAVIS *and* A. M. REES-JONES *Beecham Products Research Department, Leatherhead, Surrey, England*

Presented at the Symposium on 'Product Evaluation' 18 April 1978, at Eastbourne

Synopsis

Particular attention is paid to two methods of performance evaluation (thermography and hygrometry) that allow **antiperspirant** activity measurements to be made under controlled conditions that do not interfere with the normal operation of the glands or the cooling caused by the evaporation of **sweat** at, or near, the skin surface. **Thermography** involves imaging the axillary vault region from the long-wave radiations emitted by the warm skin surface and uses the cooling effect of evaporation to map out areas where water is evaporating from the surface. **Hygrometry** involves relative humidity measurements of ambient air passing over the skin surface: sweat evaporating increases the relative humidity of the current of air. The results obtained from these methods are compared with those derived from conventional gravimetric tests on the back and the axillae in which sweat is collected in occluded absorbent pads.

INTRODUCTION

Thermo-regulation in man is a complex efficient system in which the thermal energy released as a result of the chemical and physical activities within the body is used, in conjunction with local and central temperature control systems, to maintain the body and blood temperature near 37°C (*Fig.* 1).

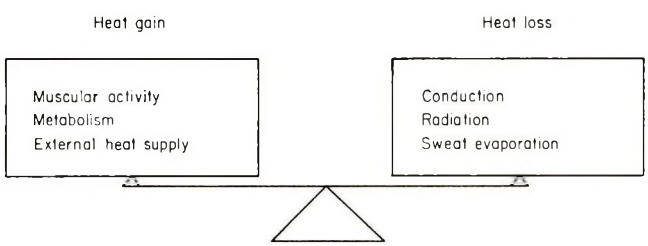


Figure 1. The thermal balance.

Heat is gained from within the body as a result of metabolic activities. Sources such as the heart and liver are relatively constant whereas heat generated by skeletal muscular activity produces a variable contribution. Heat production under resting conditions is approximately 4·2 kJ/kg of body weight per hour (1 kcal/kg/h). During physical activities this heat production rate increases ten-fold.

Heat is gained from the environment by radiation and conduction. Heat is lost by conduction and evaporation in the lungs and by radiation, conduction and evaporation at the skin surface. In a temperate climate a resting, clothed person loses approximately half of his heat by radiation and conduction and half by evaporation.

0037–9832/78/0700–0413 $02.00

Clothes are intended primarily as a means of adjusting the balance of heat gain and loss. Civilisation and fashion often interfere with this freedom to adjust our clothing, therefore products such as antiperspirants and deodorants become necessary.

Sweating is a mechanism by which the body can involuntarily lose heat by the evaporation of sweat in a short time in order to dissipate heat that would otherwise cause a rise in blood temperature. For every gram of sweat that is evaporated up to 2 kJ is extracted from the environment at the skin surface. This heat extraction occurs only when the sweat evaporates, which is why a slightly elevated ambient temperature such as 38°C, can be almost unbearable when the relative humidity is near 100%, but pleasant and even exhilarating when the humidity is low.

The sources of copious, watery sweat are the eccrine glands (atrichial) situated 1–3 mm below the skin surface, connected to the surface via spiral ducts in the epidermis. The eccrine glands are distributed generally over the body surface but are more dense on the palms and soles. The eccrine sweat glands in the axillary vault can be activated as a result of emotional or thermal stress. The sweat produced in the axillae can be particularly troublesome because sweat evaporation is limited except when the arms are raised, thus liquid sweat may accumulate, wetting surrounding clothing and inducing feelings of discomfort.

To reduce the output of sweat in the axillae manufacturers provide solutions of aluminium salts, in easily used dispensers.

The mode of action of aluminium salts has not yet been established unequivocally. Three possible modes of action involve (a) blockage of the ducts, (b) increased permeability of the duct wall, and (c) reduced activity of the gland.

In measuring the efficacy of antiperspirants it is not necessary to know how or where the product is working, provided that the parameters used to define efficacy are realistic in terms of user perception. If, however, research leads to the mode of action being elucidated, subsequent improvements in efficacy may be anticipated.

The efficacy of an antiperspirant is best defined as the percentage reduction in the rate of sweating in the axilla that may be achieved after a realistic application or series of applications of the test product. There are three basic methods by which we measure efficacy: (1) gravimetric, tared pads collect sweat for a fixed period; (2) hygrometric, the increased humidity of air passing over the skin is recorded at a fixed flow rate. (3) thermographic, the extent of localised lack of cooling when an antiperspirant treatment is working is imaged and quantified using a scanning camera sensitive to infrared radiations.

It is the aim of cosmetic chemists to produce a product that will consistently achieve a temporary decrease of the sweat rate in the axillae in the order of 20–60% reduction. Such treatment would reduce the formation of liquid sweat on the skin surface without the risk of undesirable side effects.

Although it is possible to achieve this efficacy in the majority of subjects, some do not respond well to treatment and some are even induced to sweat more. This type of reaction to a product is described as 'properspirant'.

SWEAT INDUCED BY THERMAL AND EMOTIONAL STRESS

In order to determine the efficacy of a product and its consistency in action from subject to subject it is necessary to standardise the experiment without making it unrealistic or non-representative of the whole population

Panels should consist of at least twenty subjects (for adequate statistical analysis of data), randomly selected, who are screened as being fit to withstand thermal stresses. According to Majors and Wild (1) only about 1% of volunteer subjects are found to be unsuitable. Kuno (2) showed that eccrine sweating can be affected locally and centrally by the magnitude of the thermal stress. It can also be induced by emotional stimuli such as anger, frustration, embarrassment and even the mental stress associated with mental arithmetic as shown in the work of Quatrale, Stower and Felger (3). In order to prevent emotional sweating causing excess scatter in our thermal stress data, we impose the following restrictions on the test.

(1) Subjects are asked to sit quietly in a waiting room held at 23°C for 1 h after they have changed into shorts and suntop plus a towelling robe. During this stabilisation period they are allowed one cup of hot coffee and are provided with magazines to read.

(2) Groups of subjects are thermally stressed without their towelling robes whilst sitting in a warm room at an air temperature of $38{\cdot}5 \pm 1$°C and relative humidity $35 \pm 5\%$.

(3) Subjects are asked to keep both feet on the floor and maintain an upright posture. They are asked to avoid discussing emotive subjects.

(4) Liberal applications of antiperspirant product are applied to the whole of the axillary region by experienced staff, at least 1 h before subjects are thermally stressed.

Shelley and Hurley (4) showed that even hyperhidrotic subjects could be treated effectively if an antiperspirant treatment was applied before going to bed and occluded overnight. What we suggest happens during the night is that at some stage the sweat glands are inoperative, giving the active, available ions the opportunity to diffuse down the sweat ducts to the zone of action. If applied when the subject was sweating the active ingredient might be flushed away; to avoid scatter in antiperspirant efficacy tests, applications should only be made when the glands in the area are inactive and at least 15 min should be allowed for the product to dry.

It is not clear whether or not the sweat glands responsible for thermal regulation in the axillae are the same ones that respond to emotional stimuli, neither is it known whether an effective antiperspirant causes (a) all glands to operate at a reduced rate, (b) all glands to operate for only a part of the time, or (c) some glands to cease functioning temporarily. Antiperspirants may cause a combination of (a), (b) and/or (c) to occur.

Transient physiological factors make sweating difficult to control precisely. An example of this is the fact that the body thermostat setting changes quite markedly over a 24 h cycle. In addition, subjects who are suffering from an infection, such as the common cold, can have a high blood temperature without sweating, for a day or so before other symptoms appear.

GRAVIMETRIC DETERMINATIONS OF ANTIPERSPIRANT EFFICACY

Having defined potential sources of variation in the sweat rate in the axilla and ways of limiting their effect there are certain precautions that must be taken to ensure that useful data are obtained from gravimetric sweat collection procedures.

Majors and Wild (1) showed that the right axilla (particularly if he is right-handed) in an individual may consistently sweat more than the left. Therefore it is necessary to establish a subject's right : left ratio before a test product is applied to one axilla. It is also accepted that even under well controlled temperature and humidity conditions a

Table I. Seasonal effect on sweat collected under fixed hot-room conditions from the back

	Mean amount of sweat collected from 4 cm² (mg)	
Subject	Summer	Winter
1	46	15
2	172	82
3	113	11
4	63	26
5	94	18
6	76	10
Mean	94	27

subject may sweat more one day than another. *Table I* compares summer and winter sweat rates under fixed hot-room conditions and fixed stabilisation conditions. To overcome these problems the percentage efficacy of a product is typically quantified on an individual subject in the following way.

(1) Subjects are asked to refrain from using antiperspirants for 2 weeks immediately before the test week.

(2) On days 1 and 2 of the test week (control days 1 and 2) subjects stabilise in the waiting room for 1 h and then sit in the hot-room for 80 min during which time weighed absorbent pads are applied to the axillae. Two separate collections are made, each of 20 min duration after an initial warm-up period of 40 min.

(3) For the efficacy determinations the product is applied to one axilla at the end of the second control day and then 1 h before and 15 min after the collection periods on test days 3 and 4 and 1 h before collection on day 5.

(4) The pads are weighed before and after sweat collections so as to establish the right to left sweat weight ratio* for the control and test days. The sweat ratio is derived from control days 1 and 2 by dividing the sweat collected from the assigned test axilla by that from the control axilla. From the ratios of sweat collected during test days 3, 4 and 5, the percentage efficacy of the product is obtained.

The equation used to define efficacy is

$$\% \text{ reduction in sweat rate} = 100\left(1 - \frac{\text{test ratio}}{\text{control ratio}}\right)$$

Specific problems arising from this type of gravimetric test are mainly those concerned with (a) thorough coverage of the treated area by the product, (b) keeping the weighed pad in place during the collection period and (c) ensuring that transfer of the solids content of the applied product does not interfere with the measured efficacy.

Gravimetric tests in the axillae are unavoidably long because subjects refrain from using antiperspirants for two weeks before the start of a trial and between the essential cross-over stages of such trials. In order to screen a larger number of experimental formulations a back-screening method was adopted and developed. This method involves ten, 2 cm square sites arranged as two strips from below the shoulder blade to the waist.

* By using sweat rate ratios the scatter caused by day to day variations in sweat rate is reduced.

Test products are applied in a randomised pattern on six sites, constant throughout the test week, and four control sites situated as indicated in *Fig.* 2.

0·1 ml of test product is applied from a syringe to an area 3 cm square (so that the area treated overlaps the collection area) and is allowed to dry for 1 h in a room at 23°C. At the end of this period a dried 2 cm square pad made from surgical pad gauze* backed by a 5 cm square sheet of occlusive, adhesive tape† is fixed over the collection site.

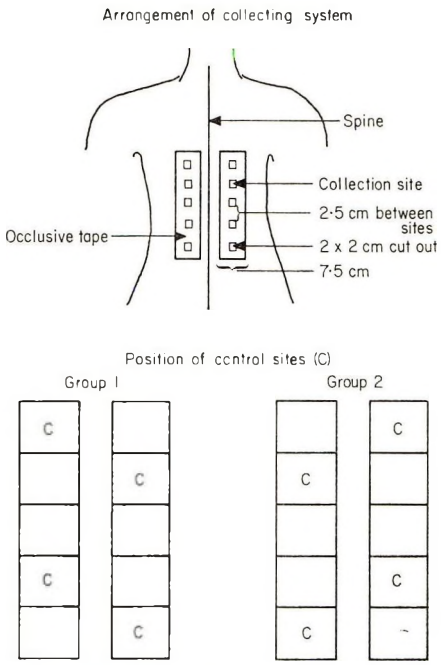


Figure 2. Gravimetric sweat collection on the back

The subjects then enter the hot-room (38·5°C 35% RH) for a period of 45 min. At the end of ths period the patches are removed and weighed. They are weighed 24 h later after drying in a desiccator. Under these conditions control site patches typically collect up to 200 mg of sweat. This compares with up to 2,500 mg over 40 min over the whole axillary vault.

The efficacy of products can be numerically higher from back screening tests than axillary tests. Nevertheless the back screening test has the advantage that six products can be compared with each other directly and it is unlikely that a usable formulation will be rejected as a result of a back screening test.

HYGROMETRY

As mentioned earlier the cooling effect of sweating is brought about by the evaporation of water at or below the skin surface, thus as ambient air flows over the skin its moisture content (relative humidity) is increased. The product of (the increased water content) × (the air flow rate) is the rate at which sweat is evaporating at the skin surface. Provided that sweat droplets are not forming on the surface the rate of evaporation equals the rate at which sweat is emerging from the sweat ducts.

* Surgipad, Johnson & Johnson Ltd, Slough, England.

† Sleek tape, S & N Southalls Ltd, Welwyn Garden City England

In order to measure the rate of sweating, cylindrical cells (*Fig.* 3) are fixed to treated and control sites on the backs of subjects. 'Ambient' air (38·5°C 35% RH) is drawn from the warm room and pumped into the cells at the rate of 150 ml per min through twelve 1·5 mm diameter holes situated concentrically inside the cells. The cells are held in close contact with the skin by 3 cm wide strips of elastic thus preventing gross leakage. The outgoing air from each cell is directed through a humidity sensor*. The changes in capacitance of the sensor (related to humidity changes) are electronically processed and amplified before being recorded directly† as percentage relative humidity. The air flow rate (150 ml/min) was found to be sufficiently fast to prevent the accumulation of liquid sweat on the test area.

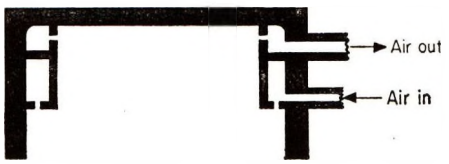


Figure 3. Hygrometric sweat measurement cell. The perspex cell directs air onto the sweating skin surface via a concentric ring of holes. The outgoing air has an elevated water content as a result of sweat evaporation.

If any leakage occurs at the cell/skin boundary the air escaping will have picked up as much moisture as monitored outgoing air. Thus the product of the humidity rise and the ingoing flow rate are the relevant measurements. The outgoing flow rate is used to check for gross leakage.

On entering the hot-room subjects were found to differ in their response to the thermal stress applied mainly in that the time required for sweating to increase to a consistent rate ranged from 10 to 30 min. The plateau values themselves varied from person to person and from day to day for each person.

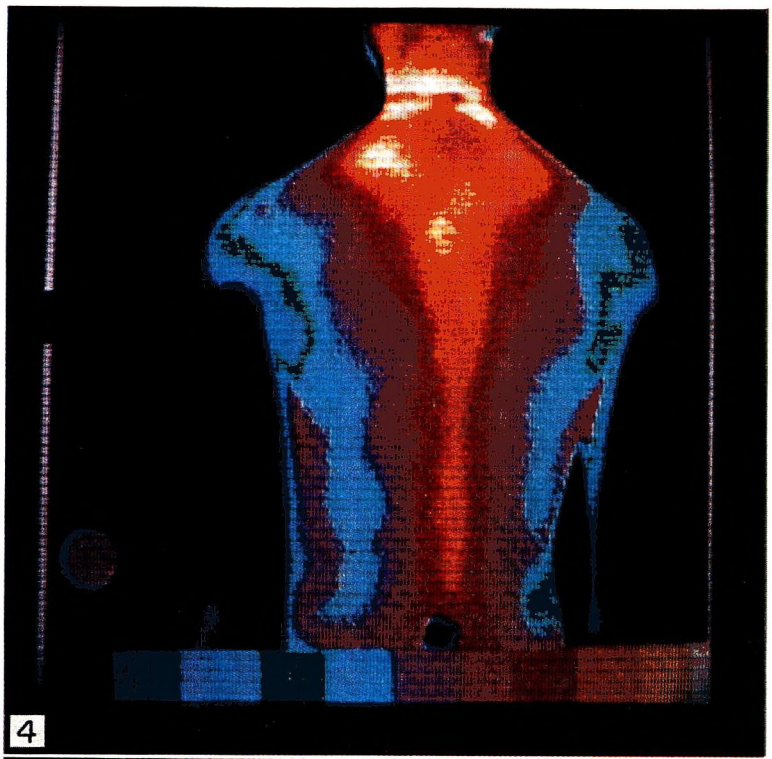
THERMOGRAPHIC MEASUREMENTS

A sweat rate equal to 50 mg/min from the axillary vault draws 120 J/min from the skin surface to evaporate the sweat. When the room temperature is at or near skin temperature radiation causes no heat loss or gain, thus all the heat generated by the body should be lost by evaporation from the skin and lungs to maintain equilibrium. Under these conditions the skin surface is the coldest part of the environment.

Localised cooling of the skin by evaporation contributes to thermal regulation, therefore it is reasonable to assume that there is less cooling of the skin over areas where an effective antiperspirant has been applied, i.e. the skin temperature is higher. Thermography, is in effect, the process of recording variations in intensity of long wavelength emissions from a surface in a mode of action similar to that of a visible wavelength television system. Hot areas emit more energy in the sensitive range of the instrument than cold areas so are thus displayed on a television screen as brighter areas in monochrome systems, or as a particular hue in colour systems displaying temperature variations over the skin surface as a map (*Figs.* 4–7).

* HP4 humidity & temperature probe, Lee Dickens Ltd, Kettering, England.

† Speedomax W Multipoint Recorder, Leeds & Northrup Ltd, Birmingham, England.



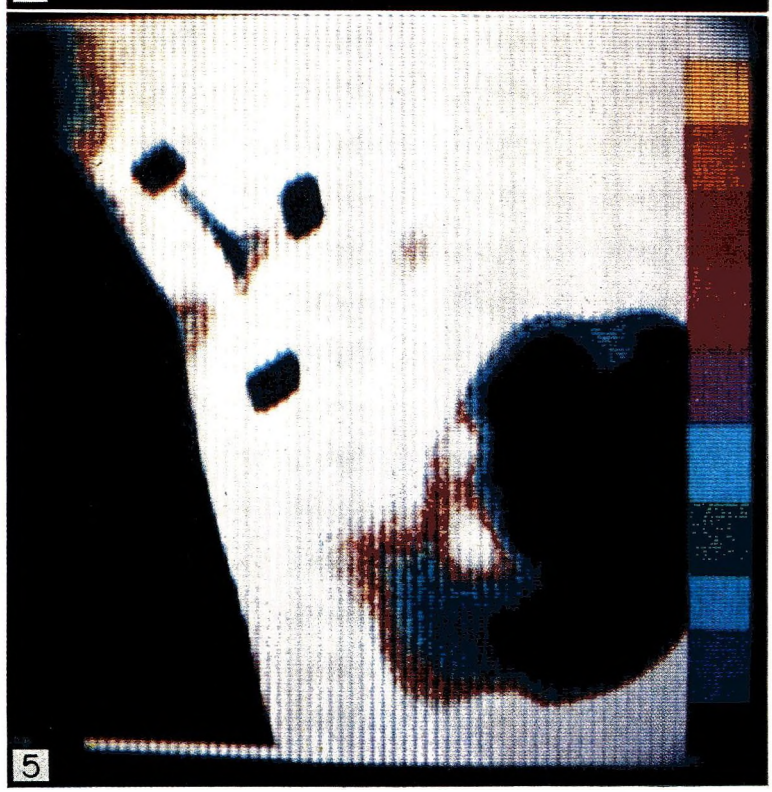


Figure 4. A thermogram of a non-sweating untreated back. A change of hue corresponds to 0·5° C.

Figure 5. A thermogram of a thermally stressed back treated with antiperspirant. A cross of water (vertically) and product applied (horizontally) 20 cm below the shoulder is delineated by black square markers. There is a hot (green) strip of antiperspirant treated skin. 5° C scale length.

Facing page 418

Figure 6. A thermogram of applied product evaporating in the axilla taken 10 min after application. The cool (coloured) cross is fading as the applications dry out. The surrounding areas appear white because they are hotter than the temperature range (1° C) covered by the ten colours.

Figure 7. A thermogram of an axilla sweating during mental stimulation under ambient conditions. 5° C scale length. The normally hot central axillary region is 4° C cooler than the surrounding skin.

Facing page 419

The sensitivity of this type of instrument is such that it can display temperature differences as small as 0·1°C as distinct colour or intensity changes. Typical magnification ranges are available to image either a whole body or a portion of a finger on a 10 × 10 cm square screen. From the thermographic work of Parke and Reece (5) the axilla is seen to be one of the hottest areas of skin.

Figure 4 is a thermogram of a non-sweating untreated back, *Figure* 5 is a thermally stressed (sweating) subject showing a relatively hot strip of antiperspirant treated skin 15 cm wide 3 cm high, 20 cm below his shoulder. (A green indentation into a cooler [blue-black] area). *Figure* 6 shows the cooling effect, caused for a period exceeding 10 min, by aqueous products applied in cross-form to the axillary vault; *Figure* 7 shows 4°C local cooling caused by mentally stimulated axillary sweating under ambient conditions.

PREDICTIONS

The sweat gland, duct and orifice are diagrammatically displayed in *Fig.* 8. For simplicity the duct is drawn straight and parallel sided, with the gland comprising the bottom element. Using this model of the system and assuming it to be full but not operating when antiperspirant is applied, calculations by a geometrical sub-division method of the diffusive flux of aluminium ions down the duct indicated that several hundred ions could have reached a gland, 2 mm from the skin surface, in 20 min. As the applied product is drying (5–10 min) aluminium ions would diffuse into and down the duct from a source at the skin surface that is becoming more concentrated.

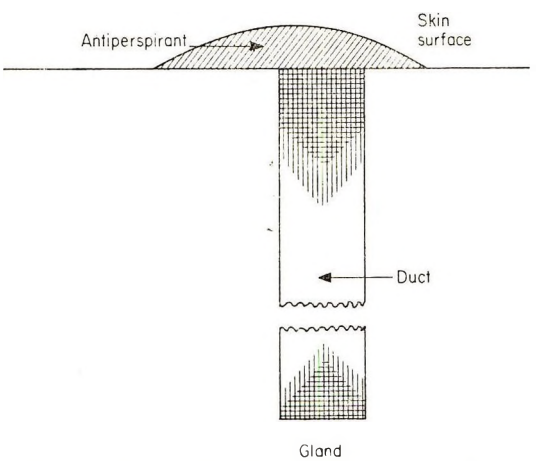


Figure 8. Geometrically subdivided sweat duct.

Although we have not, as yet, accumulated conclusive evidence of the whereabouts of the antiperspirant activity (gland, duct or orifice) we do know that within 40 min of applying a solution of aluminium chlorhydrate (5/6 basic) antiperspirant activity is achieved even if the treated area is washed to remove the applied antiperspirant.

Table II. The penetration of aluminium ions ($Al[H_2O]^{3+}_6$) from 20% aqueous solution after 20 min*

Distance from skin surface	Concentration of aluminium (ions/μm^3)	
(mm)	Mild sweating	No sweating
0·2	4×10^7	8×10^8
0·6	4×10^3	4×10^8
2·0	$< 1 \times 10^{-2}$	4×10^4

* For these geometrical subdivision calculations a Diffusion Coefficient of $7{\cdot}66 \times 10^{-7}$ cm^2 s^{-1} was assumed and 'mild' sweating was 15 $\mu g/cm^2$/min.

Calculations of diffusive penetration showed that if sweat was being produced by the glands at a rate equal to the normal rate of trans-epidermal moisture loss then the flow rate up the ducts was sufficient to prevent aluminium ions from reaching the gland (*Table II*). Thus it could be predicted that single applications of antiperspirants to subjects were unlikely to cause reproducible activity and that repeated applications of antiperspirant should improve the efficacy unless the subject was consistently sweating when the application was made.

EFFECT OF DOSAGE AND NUMBER OF APPLICATIONS ON ANTIPERSPIRANT EFFICACY

As the applied product dries the concentration of active ingredient increases, therefore, it was not surprising that with ample doses of expertly applied products the initial concentrations of aluminium chlorhydrate within a certain range, did not significantly affect the efficacy measured (*Table III*). In tests involving repeated applications to the axillae over 3 or more days, however, it is common to observe an increased efficacy after 2 or 3 days that may be the result of treating glands that were missed during the first application (*Table IV*).

Table III. Hot-room, back, antiperspirant efficacies (percentage reductions) for solutions of aluminium chlorhydrate in water

Trial	Al.Chl. content (%)	0	2	5	10	15	20	25
1	Sweat	4		39	44		32	27
2	reductions (%)		25	29	24	29	30	37

Having stated that antiperspirants can be effective within 40 min of application mention must be made of properspirancy. Aluminium chloride in water was sometimes observed to cause an increase in the sweat rate in treated axillae tested 1 h after application yet it was as effective as aluminium chlorhydrate 22 h after application.

Table IV. Comparison of efficacy after two, four and six applications, hot-room axilla, ten subjects

Treatment	Mean % reduction		
	Two applications	Four applications	Six applications
X	26	27	35
Y	25	34	37
Z	24	36	39

A student *t*-test for paired data shows that the efficacies of formulations Y and Z are significantly greater after four applications of product than after two applications.

In order to measure the speed of action of an antiperspirant or its durability over short periods it is necessary to decrease the period over which measurements are made. The gravimetric methods require at least a 20 min collection period in order to achieve acceptable reproducibility; hygrometry measurements can take less than 2 min (*Fig.* 9). The extreme of short response time has been achieved in our thermography work which takes 1 sec to map the surface temperature of the axilla. The rate of repeats is dictated by the photographic process not the infra-red detector system.

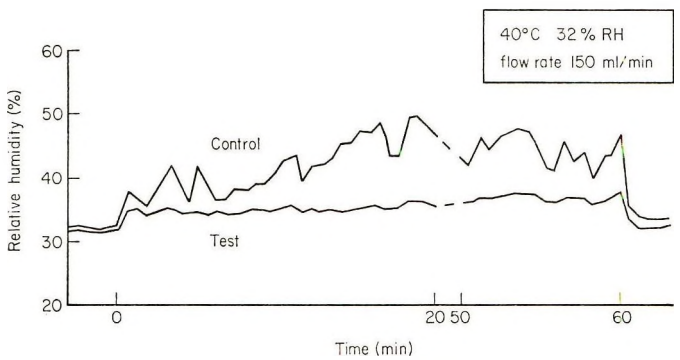


Figure 9. Hygrometric measurements.

DISCUSSION AND CONCLUSIONS

Almost any variation in an experimental procedure or physiological aspect can cause a change in the recorded antiperspirant efficacy, thus it is not surprising that different methods give different answers, or that efficacy measured on the back is greater than that measured in the axillae. Whenever antiperspirant efficacy is quoted the basic method must be defined, the degree of emotional and thermal stress must be specified and the number, timing and dosage of product application must be specified. Both the hygrometric and thermographic methods enable us to generate information that is unobtainable by established gravimetric methods, i.e. observations of short term effects and the spatial distribution of the cooling effect of sweat evaporation.

ACKNOWLEDGMENTS

The authors wish to acknowledge Mrs P. M. Baxter, Miss J. Winchester and Dr R. J. Jackson of Beecham Products Applied Research Laboratories, Leatherhead and Dr R.

Woodrough of St Bartholomew's Hospital, London for the considerable contribution toward the work and methods referred to in this paper.

REFERENCES

1 Majors, P. A. and Wild, J. E. The evaluation of antiperspirant efficacy: influence of certain variables. *J. Soc. Cosmet. Chem.* **25** 139 (1974).
2 Kuno, Y. Charles C. Thomas, Illinois. *Human Perspiration* (1956).
3 Quatrale, R. P., Stoner, K. L. and Felger, C. B. A method for the study of emotional sweating. *J. Soc. Cosmet. Chem.* **28** 91 (1977).
4 Shelley, W. B. and Hurley, M. J. Studies on topical antiperspirant control of axillary hyperhidrosis. *Acta Derm.-Vener.* **55** 241 (1975).
5 Park, W. M. and Reece, B. L. *Fundamental Aspects of Medical Thermography* (1976) British Institute of Radiology, London.

J. Soc. Cosmet. Chem. **29** 423–431 (1978)

Toothbrush wear, brushing forces and cleaning performance

B. R. PUGH *Unilever Research, Isleworth Laboratories, Middlesex, England*

Presented at the Symposium on 'Product Evaluation', 18 April 1978, at Eastbourne

Synopsis

How long a brush should last is largely determined by the magnitude of the **forces** used during toothbrushing. Other factors are also implicated. These may include the method of brushing, the geometry of the teeth and whether the individual pre-wets the **toothbrush** in hot or cold water, thus affecting the mechanical properties of the filaments. Forces measured during toothbrushing range from 4 to 20 N independent of the manual strength or sex of the individual. Those individuals who consistently brush with forces below 6 N do not wear out brushes. Toothbrushes having various states of **wear** have been examined for their **cleaning** effectiveness *in vitro*. No serious loss in ability to remove cosmetic stains from teeth could be observed until the brushes were severely worn. For the majority of the population, this is unlikely to lead to a noticeable increase in cosmetic stains on their teeth.

INTRODUCTION

Toothbrushes are widely used for keeping teeth free from soft deposits of dental plaque and food particles. They are also used with toothpastes for removing visible stains. Such stains occur as a result of the absorption of coloured materials from tobacco and some foods or drinks into the proteinaceous film (of salivary origin) covering tooth enamel *in vivo* (1). While the removal of plaque depends largely on the dexterity and motivation of the individual, cosmetic stains cannot be removed by the action of a toothbrush and water alone. A toothpaste is required for this task (2, 3).

The effects of toothbrush design and toothbrushing technique on performance have been reported (4, 5). It is surprising, however, that since many people habitually use toothbrushes that are badly worn (6), little attempt has been made to examine this factor. McKendrick, McHugh and Barbenel (7) reported that they could find no correlations between brush wear and either the Oral Hygiene Index or the Periodontal Index. Bergstrom (8) found that wear was correlated with the Hygiene Index. No studies have reported the effect of brush condition on the ability of a dentifrice to remove cosmetic stains.

How much a toothbrush has worn probably depends on the mechanical properties of the filaments, the forces applied during brushing and the manner by which the teeth are brushed. Heath and Wilson (9) have reviewed the literature on the measurement of the forces applied to a brush during toothbrushing and concluded that only a limited number of variables had been examined. In some cases subjects were asked to brush in a specified manner that was different from their normal brushing habits, and thus the forces recorded were probably atypical. Most studies were designed to establish force

0037–9832/78/0700–0423 $02.00

patterns that could give guidance in designing laboratory brushing machines for measuring the abrasivity of toothpastes. None were concerned with the effect of brushing forces on the wear of a brush.

In the investigation to be described here, the importance of brush condition on cosmetic cleaning was examined. Brushing forces were also examined to establish their role in the process of brush senescence.

Methods for measuring cosmetic cleaning *in vivo* are both time-consuming and tedious (3, 10). Simple laboratory tests using plastics, paints, metals, etc. are generally unsatisfactory owing to the difficulty of matching their tribological properties to those for naturally stained protein films *in vivo*. Alternatively, naturally stained extracted teeth can be used. The optical reflectances of such teeth change only slowly when they are brushed with a slurry of a toothpaste and there are wide variations between the individual rates of change. These problems can be overcome by (a) constructing mosaics containing ten to fifteen teeth, (b) comparing the rates of increase in reflectance for two treatments on each mosaic (whether a reference and a test toothpaste or a reference and worn toothbrush) and (c) absorbing a black dye into the protein film in order to magnify the changes in reflectance.

Several authors measured the forces applied during toothbrushing using brushes with specially constructed handles containing the transducer elements (9). A more direct approach of cementing strain gauges onto the handles of commercial brushes is preferred as the bulky handles used in the former approach could conceivably have inhibited subjects from brushing naturally.

Bergstrom (8) assessed the condition of toothbrushes by measuring the angle of permanent deformation of individual filaments. This approach was rejected by us because it could not be correlated with the perceived amount of wear. Instead, a simple scale has been developed in which samples of worn commercial brushes were ranked in order of the severity of wear.

EXPERIMENTAL

BRUSH WEAR

A large number of worn commercial brushes were obtained from people who had used them normally at home. These brushes were Wisdom, Oral B 40 and Gibbs short-head varieties. For each type of brush, six were selected to cover a wide range of senescence. Two new brushes of each type were added to the numbers, making twenty-four brushes in total.

Fifteen judges were selected at random and each independently was asked to sort the brushes into groups which represented *perceptually important* differences in patterns of wear, without restriction on either the number of groups or the number of brushes within a group. In addition, the judges assigned percentage scores to the severity of wear for each group based on the arbitrary notation that 100% represented a completely worn out brush (in their estimation).

BRUSH WEAR AND BRUSHING FORCES *in vivo*

Forty-seven panellists were recruited from office and laboratory personnel. Each was given a Gibbs short-head medium brush and asked to use it normally at home. At intervals over an eight week period, panellists returned their brushes for inspection.

Subsequently each panellist attended a panel room for measurement of the forces applied during brushing. Each was given a Gibbs short-head medium brush fitted with a strain gauge, a supply of toothpaste, and then asked to brush normally. After the panellist had been brushing for 5–10 sec the gauge recorder was switched on.

The signals from the strain gauge were amplified before recording using a high speed UV galvanometer These traces were then used to measure the peak heights of the force pulses. The brushing rate was also measured by counting the number of pulses per unit time and converting this to pulses per minute.

CLEANING *in vitro*

Freshly extracted molars and pre-molars, which had at least one surface free from dental restorations or caries, were dyed black using a mixture of 4-methylcatechol and *p*-phenylenediamine. After cutting, these were mounted in plastic rings 4 cm in external diameter to form mosaics containing ten to twelve teeth closely packed together.

A specially constructed reflectometer was used to measure the luminance of the mosaics. The light source was a ring fluorescent tube 12 inches in diameter mounted behind an hemispherical bowl made of white plastic. A mosaic placed in a circular hole in front of this bowl reflected light into an optical system, placed on the opposite side of the bowl. This condensed the light onto a photo-detector. The optics were designed so that extraneous light from surfaces other than the mosaic did not enter the detector. A digital voltmeter was used to record the output from the photo-detector.

Samples of Gibbs brushes having various states of wear were obtained. Six mosaics were used to examine each brush in the following manner: a standard slurry of toothpaste was made containing 33% w/w Pepsodent (Elida Gibbs) in distilled water. The reflectance of the first mosaic was taken, it was then placed in a laboratory brushing machine and brushed for twenty strokes using the worn brush. A further reflectance measurement was made. This procedure was repeated for four more successive increments of twenty strokes, reflectance measurements taken after each increment, to give a total of 100 strokes. The entire operation was then repeated with a new brush in place of the worn brush. This procedure was carried out on three mosaics. A further three mosaics were then tested but on these occasions the new brush was used first.

Slopes for the regression of reflectance versus brush strokes were calculated. For each mosaic, the ratio of the slopes of the worn and new brushes was calculated using the arbitrary notation that the value for the new brush was 1·00. All six ratios were then used to calculate a mean cleaning score for that brush.

MECHANICAL PROPERTIES

Three types of Gibbs short-head brushes with filaments having diameters 0·20 mm, 0·25 mm and 0·33 mm were tested (these corresponded to soft, medium and hard brushes). In turn each new brush was mounted onto a pivoted arm so that the tips of the filaments rested on the surface of a flat, rigid table capable of being moved horizontally. Various weights (up to 2 kg) were placed on the arm above the head of the brush, the table moved through a distance of 2–3 cm then after 45 seconds the decrease in the height of a fixed mark on the head of the brush above the surface of the table was measured using a travelling microscope. Once the table had been moved, the filaments attained an equilibrium position and further movement of the table (within a few seconds) did not affect the curvature of the filaments.

The time interval between applying the load and measurement of the height was kept constant at 45 seconds for all the experiments. For other intervals of time, the measured heights were found to be different owing to the time dependent nature of the stress-strain curves for polymers such as nylon.

Data obtained in these experiments were converted into strain intensities firstly, by calculating theoretically the radius of curvature of a filament corresponding to the observed change in height and secondly, calculating the maximum strain intensity (which occurred in the outer shell of the filaments) using the relationship $d/2R$. R and d are the radius of curvature of the bent filament and its diameter respectively. Strain intensity was plotted against the applied force for each brush.

RESULTS

The samples of worn and new brushes examined by the fifteen judges could be formed into eight groups ranging from new brushes to extremely worn brushes. Thus it was possible to construct a reference scale that comprised a sample of a Gibbs, Wisdom and Oral B brush in each of the eight groups. These were mounted in groups of three on black plastic tiles for easy display. The eight groups were labelled 0–7, 0 representing new brushes and 7 representing ones that were very badly worn. It was found that new brushes had been placed in categories 0, 1 and 2. More than one category was used because even with new brushes, some filaments were found to have been misaligned during manufacture and this could be mistaken for a slight degree of wear.

The mean percentage scores for wear given to these brushes by the judges do not relate linearly to the category numbers as shown by *Fig.* 1. Only the percentage scores were used, therefore, in any statistical analysis.

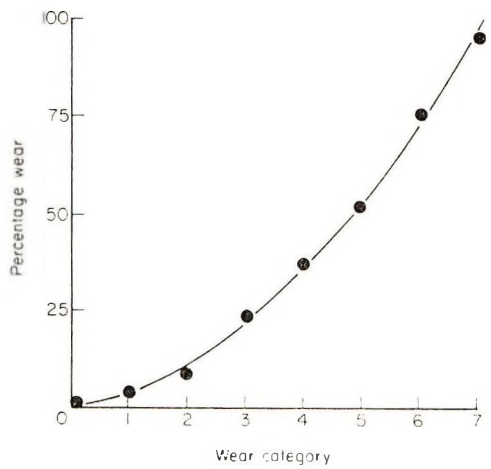


Figure 1. Percentage wear scores for the eight groups of brushes used in the standard wear scale.

The fifteen judges did not concur in their assessments of the original brushes. A Cluster Analysis technique revealed a sharp change in both the F ratio statistic (*Table I*) and the root-mean square deviation (*Fig.* 2) in the region $n=3$, thus suggesting that the judges comprised three different groups: Group 1 was made up of those judges who tended to give high scores to any worn brush, Group 3 those giving low scores, while Group 2 were intermediate between the two types (*Fig.* 3).

Table I. F ratios for *m* versus *n* clusters. Pooled data for all types of commercial brushes examined

Clusters *n* / *m*	1	2	3	4	5	6
2	3·60 (0·1%)					
3	3·77 (0·1%)	2·89 (0·1%)				
4	3·47 (0·1%)	2·49 (0·1%)	1·74 (5%)			
5	3·34 (0·1%)	2·38 (0·1%)	1·77 (0·5%)	1·66 (5%)		
6	3·16 (0·1%)	2·24 (0·1%)	1·70 (0·5%)	1·55 (2·5%)	1·36 (NS)	

First column, F ratio; second column, level of significance.

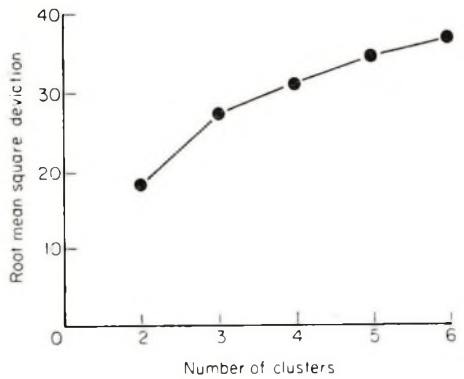


Figure 2. Root mean square deviation of observations from their cluster centres. Pooled data for all types of commercial brushes.

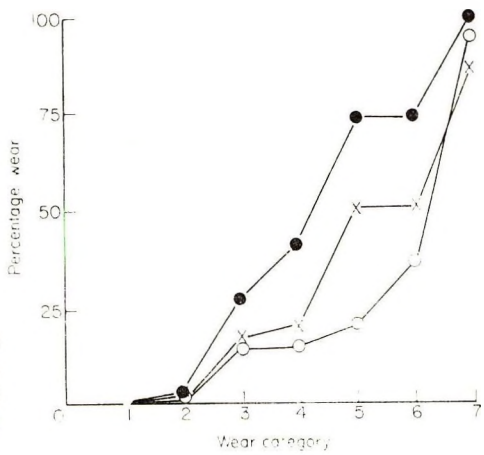


Figure 3. Fifteen judges sub-grouped into three clusters according to their assessment scores. Data for Gibbs brushes only. Cluster 1 (5 judges) ●——●; Cluster 2 (7 judges) ×——×; Cluster 3 (3 judges) ○——○.

Absolute values for the condition of a brush will be biased unless the panel of judges is balanced with respect to the different types of judges. Comparison of brushes against the reference scale of brushes, however, minimises inter-judge differences and no appreciable bias has been found in data obtained using this approach.

Brush wear data *in vivo* for the forty-seven panellists were found to give curves for percentage wear versus the number of uses that were asymptotic with respect to the time axis. Logarithm transformations of these plots gave linear regression lines, the

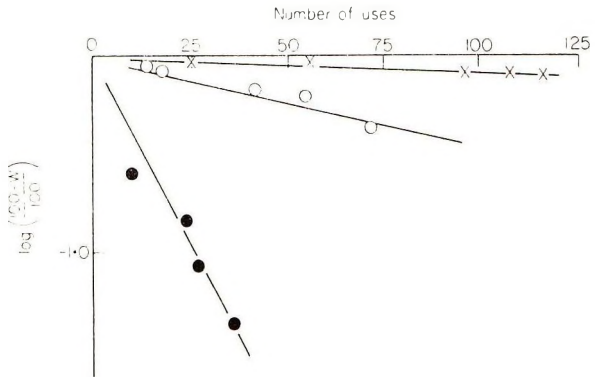


Figure 4. Logarithm transformation of the percentage wear (w) versus the number of times a brush had been used. All the subjects used a Gibbs short-head medium brush, but these data are typical of other commerical brushes. Other studies that we have carried out clearly demonstrate the fact that the rate of wear of a brush is almost entirely dependent on the subject. Those subjects who wear out their brushes quickly do so irrespective of the nature of the brush. ×——× Subject C ($k = -8 \times 10^{-4}\, t^{-1}$); ●——● (Subject B ($k = -44 \times 10^{-4} t^{-1}$); ○——○ Subject A ($k = -383 \times 10^{-4}\, t^{-1}$).

slopes (k) for which gave a measure of the rate of wear. *Figure* 4 shows such lines for three panellists who caused extreme and intermediate levels of brush wear.

Subject A had severely worn out his brush after twenty uses (i.e. 10 days). He would be classified as a 'heavy wearer' and the specific wear rate k for his performance was $-383 \times 10^{-4}\, t^{-1}$. Subject C was classified as a 'low wearer' and even after 8 weeks had not caused any appreciable wear to the brush ($k = -8 \times 10^{-4}\, t^{-1}$). Subject B was intermediate between these extremes ($k = -44 \times 10^{-4}\, t^{-1}$).

The subjects recruited for this investigation were selected to deliberately include a larger than normal proportion of those classed as 'heavy wearers'. Our experience is that such subjects do not comprise more than 10% of any typical population.

Specific wear rates calculated for the forty-seven panellists correlated with the average forces they used during brushing (*Fig.* 5). The data, however, were found to be

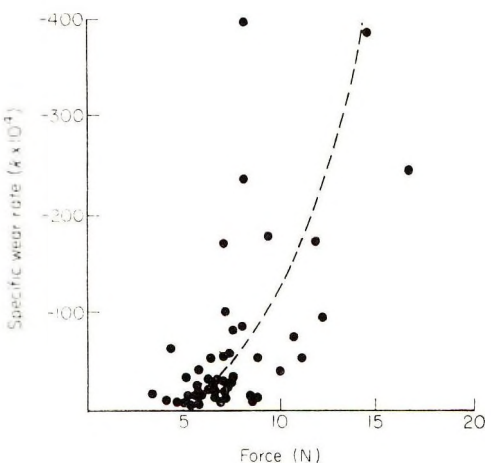


Figure 5. Correlation between the specific wear rates and brushing forces for Gibbs short-head medium brushes.

extremely scattered (r=0·50). Several subjects brushed with an average force of 9–10 N yet their specific wear rates varied between -9 and $-240 \times 10^{-4}\ t^{-1}$.

Classification of the panellists into three arbitrary groups defined by high, medium and low specific wear rates (*Table II*) shows a marked relationship between wear category and the mean brushing forces calculated for these groups. Therefore, brushing forces are implicitly involved in the ageing process.

Table II. Classification of subjects by specific wear rates

Classification		Mean force (N)	Standard error
High	wear scale 6, 7 after use for 4 weeks. ($k \geqslant 84 \times 10^{-4}\ t^{-1}$)	10·2	0·69
Medium:	wear scale 4, 5 after use for 4 weeks. ($28{\cdot}5 \times 10^{-4}\ t^{-1} \leqslant k < 84 \times 10^{-4}\ t^{-1}$)	7·7	0·51
Low:	wear scale 3 after use for 4 weeks ($k < 28{\cdot}5 \times 10^{-4}\ t^{-1}$)	6·4	0·40

Figure 6 shows typical force-strain curves for Gibbs short-head brushes. Strain intensity increased rapidly with loads above 5–6 N for the medium brush type (filament diameter, d, 0·25 mm). The hard brush (d, 0·33 mm) required loads of 12–13 N to produce an equivalent strain intensity. This correlates with the observation that a subject generally will wear out a medium brush more quickly than a hard brush. The apparent anomaly that the soft brush required a greater force than the medium brush to produce the same strain intensity arises because the former contains greater numbers of filaments per tuft, so that the force per filament will be less.

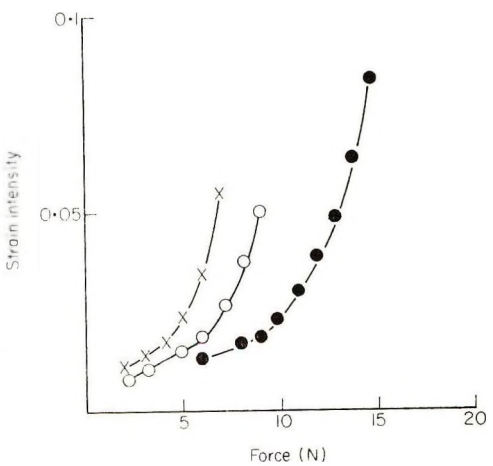


Figure 6. Strain intensity calculated as $d/2R$ (where d = filament diameter and R is the radius of curvature of the bent filament) versus the applied load. Gibbs short-head brushes containing three different diameter filaments. ●——● d, 0·33 mm; ×——× d, 0·25 mm; ○——○ d, 0·20 mm.

Figures 5 and 6 suggest that a minimum force of 5–6 N is required before appreciable wear of a Gibbs short-head medium brush would occur. Above this critical value, the filaments tend to collapse and the resultant strain intensities lead to rapid, permanent deformation of the nylon filaments.

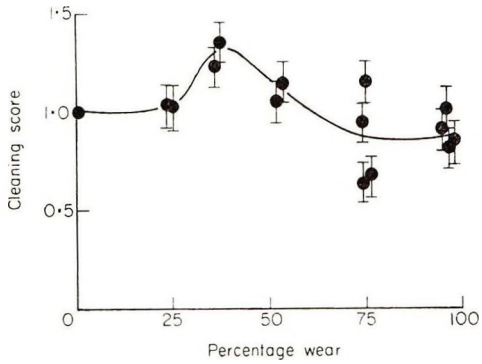


Figure 7. Cleaning scores *in vitro* for various worn brushes using Pepsodent (Elida Gibbs) as the reference toothpaste.

The ability of a toothpaste to remove stains from extracted, dyed teeth was not found to be drastically affected by the condition of the brush (*Fig.* 7). Slight wear of a brush (wear category 4) produced a significant increase ($P=0{\cdot}05$) in cleaning scores compared with the new brush. Further deterioration of the brush reduced the cleaning score. Even at wear category 7, however, at which a brush would be considered to be too worn for normal use, cleaning scores were reduced by only 10–15% (see *Table III*).

Table III. Summary of cleaning scores for various worn brushes (new brush score, 1·00)

Brush condition	Cleaning score	Number of brushes	Number of mosiacs per brush
3 (24·4% wear)	1·03	2	6
4 (36·3% wear)	1·28	2	6
5 (52·6% wear)	1·10	2	6
6 (75·5% wear)	0·86	4	6
7 (95·6% wear)	0·89	4	6

DISCUSSION

Despite the fact that the panellists in this study were made up largely of those who gave high and medium wear scores, the average force used to brush for the whole group was virtually the same as the figures quoted for normal populations by other workers. This average (7·4 N) did include subjects whose average brushing force was as low as 3 N and others who exceeded 17 N during brushing.

The brushing rate was found to be approximately 260 strokes/min with extreme values of 340 and 177 strokes/min. These figures are in excellent agreement with Robinson's value of 275 strokes/min (11).

The magnitude of the forces used during brushing have been implicated in the process of wear. The large variance observed within the correlation, however, strongly suggests other factors may be involved. Observation of the panellists' brushing habits did not reveal obvious clues For example, while some panellists pre-wetted their brushes in warm water, which will lead to a loss in mechanical resilience of the filaments, this did not relate in any simple way with the ability to wear out toothbrushes. Speculatively, the other major factor may be the manner of brushing, though again direct evidence from actual brushing habits was not forthcoming.

The condition of the brush does not drastically affect the ability of a toothpaste to remove cosmetic stains from extracted teeth. Some deterioration was observed for severely worn brushes. Our experience of evaluating the cleaning performances of toothpastes *in vivo* (3, 10) strongly suggests that this effect would not lead to any serious increase in the severity of stains on the teeth for the great majority of a population.

The function of a toothbrush is to remove dental plaque and food particles as well as help remove cosmetic stains and it does not follow that while the condition of the brush does not unduly influence the ability of a dentrifrice to keep teeth free from stains, that similar relationships hold for dental plaque and other soft deposits. This may be particularly true for dental plaque retained in the spaces between the teeth. Filaments of a badly worn brush may well be unable to penetrate into these regions, with important consequences in terms of gingival health.

REFERENCES

1 Manly, R. S. A structureless recurrent deposit on teeth. *J. Dent. Res.* **22** 479 (1943).
2 Kitchin, P. C. and Robinson, H. B. G. How abrasive need a toothpaste be? *J. Dent Res.* **27** 501 (1948)
3 Bull, W. H., Callender, R. M., Pugh, B. R. and Wood, G. D. The abrasion and cleaning properties of dentifrices. *Br. Dent. J.* **125** 331 (1968).
4 *World Workshop in Periodontics* (1966) The University of Michigan Press, Ann Arbor.
5 Frandsen, A. (ed.) *Oral Hygiene* (1972) Aalborg Stiftsbogtrykkeri, Denmark.
6 A survey into the state of the nation's toothbrushes: A British Dental Health Foundation Survey. (1975) British Dental Health Foundation, London.
7 McKendrick, A. J. W., McHugh, W. D. and Barbenel, L. M. H. Toothbrush age and wear. An analysis. *Br. Dent. J.* **130** 66 (1971).
8 Bergstrom, J. Wear and hygiene status of toothbrushes in relation to some social background factors. *Sven. Tandlak, Tidskr.* **66** 383 (1973).
9 Heath, J. R. and Wilson, H. J. Forces and rates observed during *in vivo* toothbrushing. *Biomed Eng.* **9** 61 (1974).
10 Wilkinson, J. B. and Pugh, B. R. Toothpastes: cleaning and abrasion. *J. Soc. Cosmet. Chem.* **21** 595 (1970).
11 Robinson, H. B. G. Toothbrushing habits of 405 persons. *J. Am. Dent. Ass.***33** 1112 (1946).

J. Soc. Cosmet. Chem. **29** 433–440 (1978)

Techniques for the evaluation of emollients and keratolytics

R. MARKS *Department of Medicine, Welsh National School of Medicine, Heath Park, Cardiff, Wales*

Presented at the Symposium on 'Product Evaluation', 18 April 1958, at Eastbourne

Synopsis

Established and new technologies were evaluated to assess the efficacy of **emollients** and **keratolytics**. The newer techniques included the use of (a) an instrument designed to measure *in vivo* intracorneal cohesion, (b) an instrument to assess the point penetrability of the stratum corneum *in vivo* and (c) an instrument designed to apply a standard stimulus in order to quantitate the number of squames' that are released *in vivo*. The established methods used included surface contour analysis of skin surface replicas and morphological assessments using macrophotography and scanning electron microscopy. The keratolytics so far evaluated included preparations of **salicylic acid** and **urea** and the emollients evaluated included three commercially available preparations. The results thus far showed that keratolytics are difficult to evaluate in entirely normal skin but that morphological assessments are best in abnormal skin and physical measurements are of more help in abnormally scaly skin. Emollients have proved much easier to evaluate and their effects can be detected by all the techniques described.

Emollients and keratolytics account for a large proportion of the topical applications prescribed by dermatologists and a not inappreciable proportion of products that are directly available to the public for skin care It is, therefore, astonishing that comparatively little attention has been paid to methods for the assessment of the effectiveness of these products. This has been in part due to the intrinsic difficulties in quantitating events taking place at the skin surface, in part due to lack of understanding of how emollients and keratolytics work and in part due to the empirical approach previously adopted by dermatologists and cosmetic scientists. The subject is now, however, of considerable widespread interest and the author would point out the excellent review of Quattrone and Laden (1) which concentrated on *in vitro* methods for evaluation of emollients. It is the purpose of this paper to discuss several of the techniques available for assessment, to point out those *in vivo* techniques that appear to be most useful and at the same time to document the author's experience in this subject.

TERMS AND DEFINITIONS

It is worth while at the outset to define some terms. The term 'emollient' implies (from the Latin derivation) a material designed to soften the skin. The only true emollients in this sense are destructive agents that chemically change the usually hard stratum corneum into a softer less protective substance. What we actually mean by the term 'emollient' is a material that 'smooths' the surface to the touch (and makes it look smoother to the eye).

0037–9832/78/0400–0433 $02.00

This is important to remember as it should help in designing techniques for the evaluation of emollients.

Keratolytics are even less appropriately named. This term taken literally signifies a material which 'lyses keratin'. It should not be necessary to remind you that keratin is an intracellular alpha fibrous protein suspended in a matrix of unknown composition and surrounded by the tough plasma membrane of the corneocyte wall. It seems most unlikely that this component is involved in the action of keratolytics. We probably mean by 'keratolytic' an agent that enhances desquamation. It is even more confusing when one realises that by enhancing desquamation one may also 'smooth' so that keratolytics may be also emollients! We probably need different terms and new definitions and I would suggest that we come nearer to the actual action that we seek in choosing them. If we drop the term 'emollient' and substitute 'hydrating agent' we more accurately describe the action of these substances and can define them as agents whose action is to smooth the surface by hydrating it and which is reversible and short lived. 'Descaling agent' is more appropriate than 'keratolytic' and this term may be defined as a substance that acts irreversibly to alter abnormal desquamation such that hyperkeratosis and scaliness are decreased. Thus, water is only a 'hydrating agent' and not a 'descaling agent', even though it may enhance desquamation in the short term, because its action is reversible and short lived. Similarly, salicylic acid is purely a descaling agent as although it may 'smooth' it does not hydrate and its action is irreversible.

HYDRATING AGENTS (HA)

(a) SUBJECTIVE TECHNIQUES

The palpating finger and the eye are remarkably sensitive and the user of the HA or the clinical observer can easily tell whether the skin is more or less smooth than before application. This effect cannot be accurately quantified, however, and is subject to 'bias' and the placebo effect. Middleton and Roberts (2) optimised this method during the evaluation of a 'hand cream' containing PCA, by employing a 'double blind technique' and a complex clinical scoring system. We have also been able to show that there is a difference in the 'clinical score' after the use of an emollient cream for 1 and 2 weeks in twenty patients with dry skin (*Table I*). Although these results are not statistically significant the trends are undeniable.

Table I. Results of use of aqueous cream on skin 'dryness' in subjects with dry skin

Time (weeks)	Subject assessment*	Observer assessment†
0		2·2±0·9
1	1·6±0·8	3·1±0·9
2	1·8±0·9	3·4±0·7

* Mean and S.D. of arbitrary scores: 1 = < 25%, 2 = 25–50%, 3 = > 50% improvement.

† Mean and S.D. of arbitrary scores: 1 = very dry and scaly, 2 = moderately dry and scaly, 3 = slightly dry, 4 = normal.

(b) MORPHOLOGY

The simplest morphological technique is that of macrophotography. Here the surface of

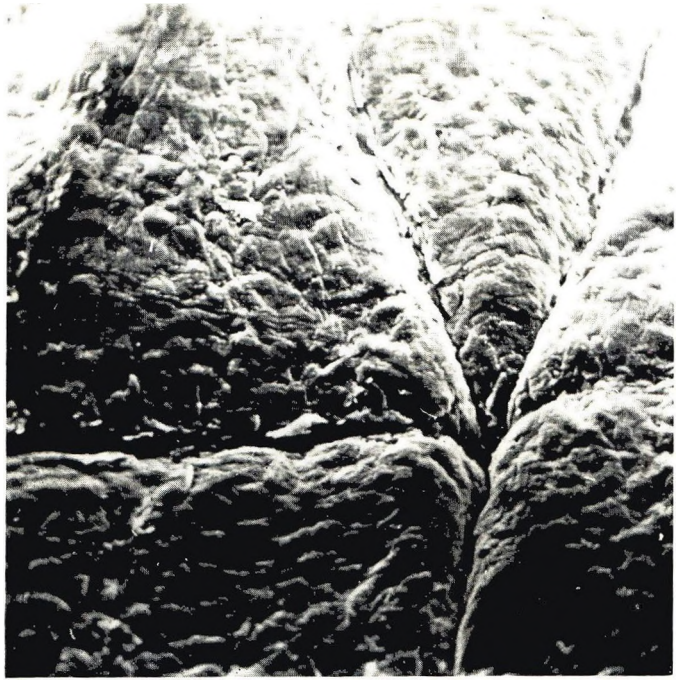


Figure 1. Scanning electron micrograph of a replica taken from skin treated with an HA 2 h previously. Notice shallow skin furrows. (× 16·5).

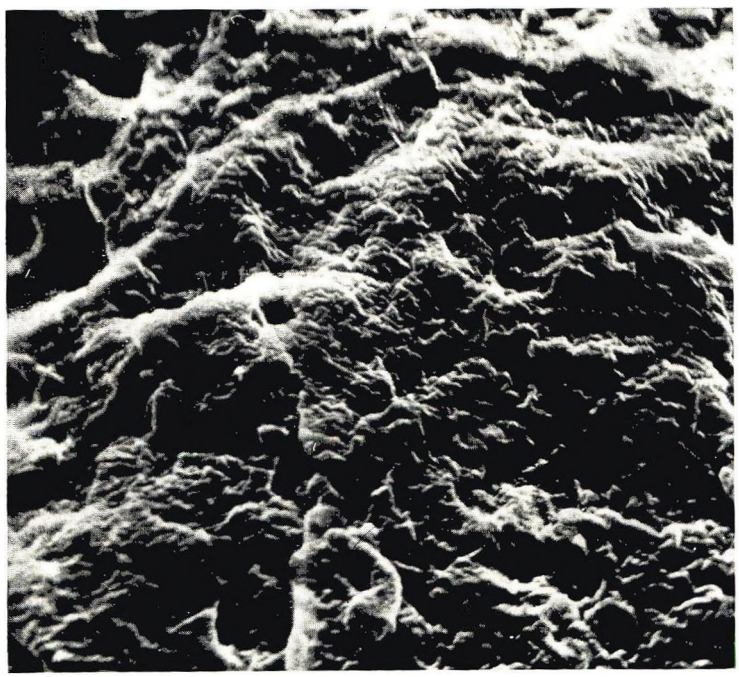


Figure 2. Scanning electron micrograph of replica of skin surface treated with HA 2 h previously. Notice prominent cell borders and thickened individual corneocytes. (× 1·28 K).

Facing page 434

Figure 3. Scanning electron micrograph of replica to show resolution. Notice the detail of the broken hair shaft. (× 165).

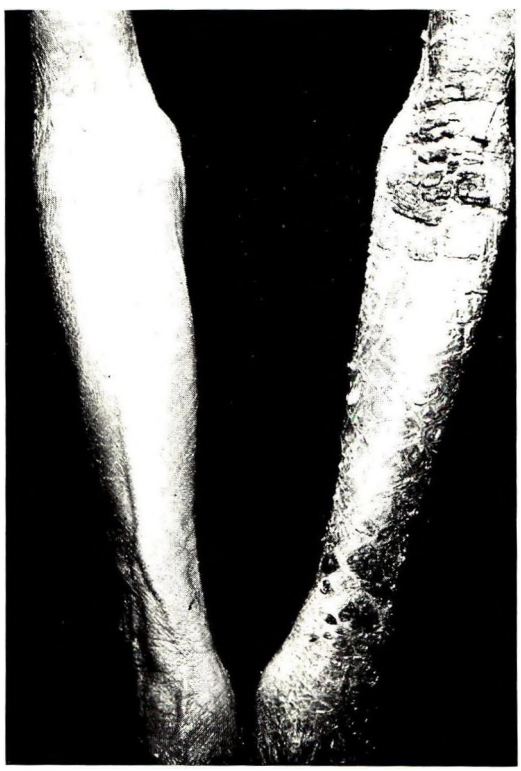


Figure 4. The clinical effects of 6% salicylic acid in white soft paraffin on right arm of man with Lamellar ichthyosis.

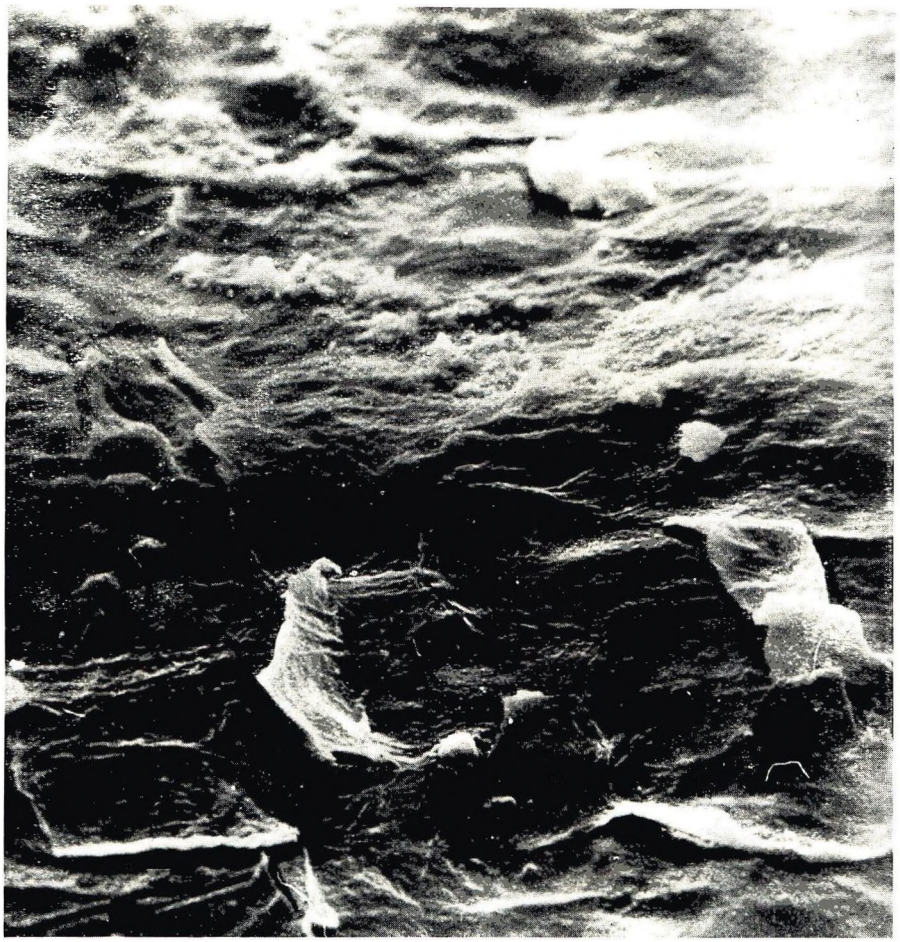


Fig. 5. Scanning electron micrograph of SSB taken from site treated with 6% salicylic acid in white soft paraffin for a 10 day period. (×950).

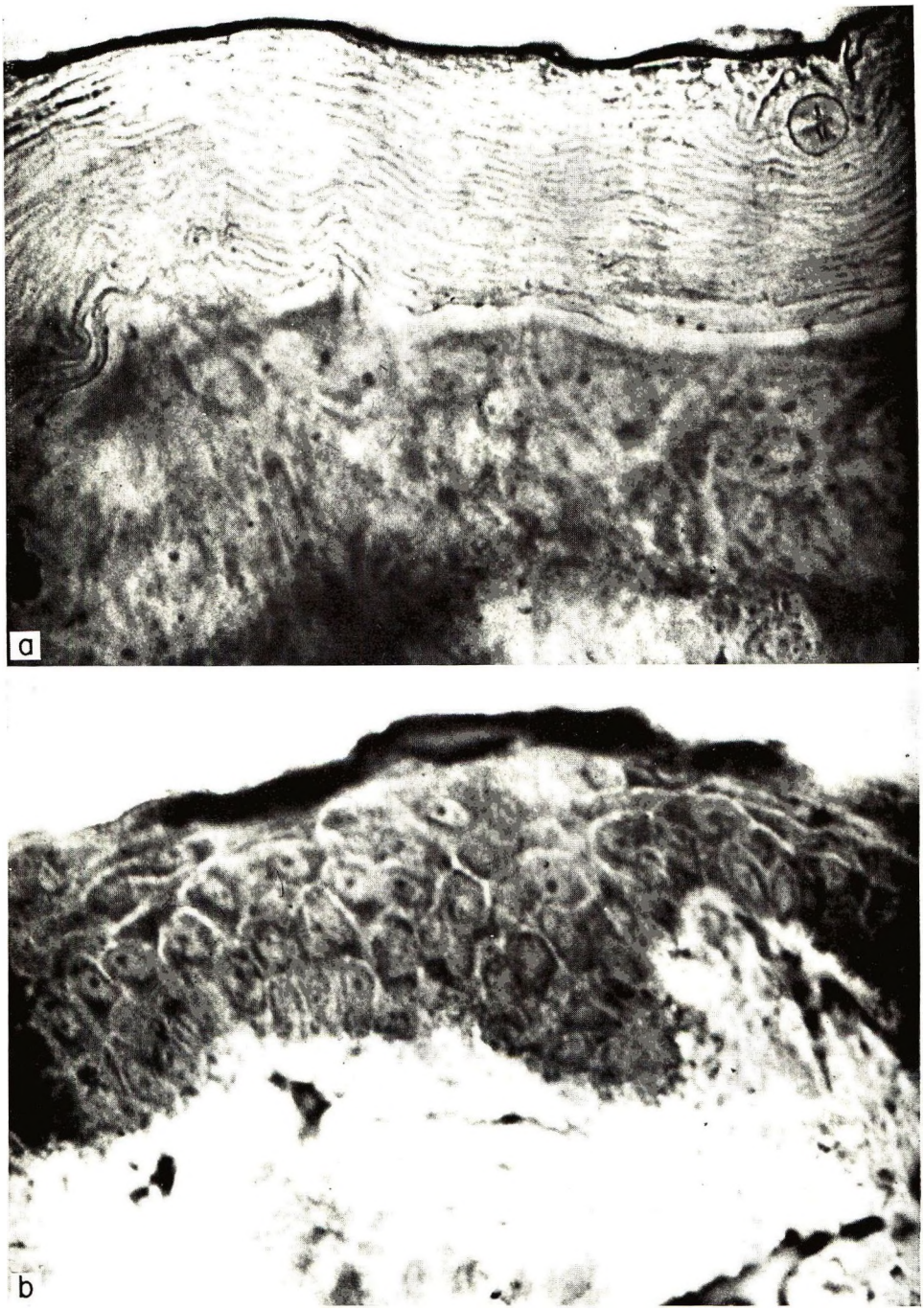


Figure 6. Cryostat sections of skin biopsies. The stratum corneum has been demonstrated by the Mackenzie technique. (a) is from the control site; (b) is from the site treated by salicylic acid.

the test site is photographed so as to give a final magnification of five to eight times. We decided to employ this technique in the evaluation of one emollient. In order to quantify any changes we arranged to measure the width of the skin furrows at five sites chosen in a standardised way by placing the photographs under a template. To remove all bias the photographs were coded and the observer who performed the measurement had no knowledge from which subject or treatment group the photographs came. For the actual measurement a $\times 8$ magnifying lens with built in measuring graticule was used. This experiment was performed on six individuals with normal skin and six with dry skin and was over a 4 h period after application of an emollient (*Table II*).

Table II. Skin furrow width determined from photographs in six normal and six 'dry skin' subjects

	Furrow width (mm ± S.D.)	
Time (min)	Dry skin subjects	Normal subjects
Control	1·41 ± 0·20	1·34 ± 0·19
0†	1·51 ± 0·16	1·54 ± 0·16
30	1·80 ± 0·37*	1·54 ± 0·21
60	1·57 ± 0·11	1·54 ± 0·12
120	1·52 ± 0·16	1·37 ± 0·15
240	1·77 ± 0·40	1·43 ± 0·26

* Difference from control ($P = 0{\cdot}05$)
† Immediately after application

I do not know why furrow width decreases after hydration but there seems to be little doubt that it does and this can serve to give a quantitative index of hydration effect. Scanning electron microscopy (SEM) has been used by several groups including ourselves to assess the effect of hydrating agents (3, 4). In our experience SEM of skin surface replicas or skin surface biopsies (4) can certainly detect alterations after application of HA but they are difficult to quantify. The most prominent change in replicas is the 'filling up' of the normal surface furrows (*Fig.* 1). In addition, individual cell margins are more prominent and the cells appear plumper (*Fig.* 2). Because this is a very cumbersome (and costly) technique and gives no quantitative data, it is, however, best to choose another way of testing HAs.

(c) SURFACE CONTOUR ANALYSIS

Replicas of the skin surface show quite marked changes after application of HA. The replicas that we now employ reproduce the surface contour with an astonishing resolution (*Fig.* 3). We use 'Silflo' dental impression material (a silicone rubber material) for the negative and we then coat the negative with 'DPX' (a styrene plastic slide mounting medium). The 'positive' is dried in a desiccator overnight and then separated from the 'negative'. After separation the positive is mounted on a glass microscope slide and subsequently its surface contour is traced using an instrument known as a surfometer (6). The stylus of this instrument barely indents the surface and the excursions of the stylus give an accurate representation of the surface contour of the specimen. All the HAs tested in this system show a decreased contour profile in the few hours following their use. The decrease in contour profile is consistent and time dependent but not dramatic. In one series of tests six normal volunteers and six volunteers with 'dry' but otherwise normal skin used a popular (oil in water emulsion lotion) on the lower legs and replicas

were taken at intervals for up to 4 h subsequent to application. The results are shown in *Table III*.

Table III. Surfometric areas ($cm^2 \pm$S.D.) for six normal and six dry skin subjects after application of an HA

	Areas ($cm^2 \pm$S.D.) in tracing from replicas	
Time (min)	Normal skin	Dry skin
Before	9·2±3·2	10·5±2·0
Immediately after	7·9±2·6	10·1±3·7
30	7·8±2.8	9·1±2·4
60	7·4±1·1	8·8±3·2
120	6·7±1·7	8·0±2·4
240	6·9±3·1	9·0±1·8

Interestingly the maximum effect has always been, whatever HA we use, between 2 and 4 h and rarely results in more than a 25% decrease in area under the contour traced. A more interesting and fundamentally more important question is whether HAs effect any cumulative change in the stratum corneum after persistent use. In one experiment a HA was applied twice daily for 28 days to the forearm of five normal volunteers and the replicas were taken at a constant time of the day after the last application. The results are given in *Table IV*.

Table IV. Area ($cm^2 \pm$S.D.) under contour tracings from replicas taken from five normal subjects at different times after application of HA. Application stopped at 28 days

Time (days)	Area ($cm^2 \pm$S.D.)
0	10·78±1·56
2	9·68±1·12
7	9·32±0·77
14	8·68±0·91
21	9·66±1·27
28	9·16±0·8
33 (5, after treatment)	9·5 ±2·28

Although the maximum change is seen at the surface, HAs also influence the internal structure of the stratum corneum (SC) and alterations in contours of the ruptured internal face of skin surface biopsies can be detected In *Table V* the results of immersing the lower leg in water containing a bath oil for six volunteers with normal skin are shown. Skin surface biopsies (5) (SSBs) were taken before and at intervals after the period of exposure (20 min).

Table VI is directly comparable to *Table III* as it is taken from the same experiment but documents the effect of SSBs rather than from replicas. Interestingly the trend to reduction in area is more noticeable in the normal individuals but does not reach statistical significance.

Similarly some effect can be determined on SSBs after 14 days application of an aqueous cream. *Table VII* shows results obtained in the same experiment as outlined for *Table I*.

Table V. Results of use of bath oil to lower leg in six individuals on surfometer tracing area

Time after exposure (min)	Area ($cm^2 \pm$S.D.) in tracing
Before	8·5±2·2
5	5·6±1·5*
60	8·2±1·9
240	6·9±2·3

* Significance difference fromcontrol $0{\cdot}05 > P > 0{\cdot}02$.

Table VI. Results of use of HA on normal and dry skin subjects on surfometric tracing of SSB contrast with Table III

	Areas ($cm^2 \pm$S.D.) in tracings from SSBs	
Time (min)	Normal skin	Dry skin
Before	6·2±1·2	6·4±1·8
Immediately after	5·2±1·3	6·6±2·0
30	5·5±1·7	7·9±0·8
60	6·5±1·9	6·6±0·9
120	5·4±1·7	5·8±1·8
240	5·0±1·2	5·7±1·7

Table VII. Results obtained from SSBs after application of aqueous cream for 14 days

Time (days)	*n*	Area ($cm^2 \pm$S.D.) in tracing from SSBs
0	8	12·9±3·7
7	9	12·0±1·8
14	8	11·7±2·0

Thus, it seems that the results of the use of HAs discernible at the surface, are also present within the substance of the SC although from the results presented in *Tables V*, *VI* and *VII* it may be that they are not as pronounced.

(d) OTHER *In Vivo* TECHNIQUES

There are two other techniques that we have used for the evaluation of HAs *in vivo* which have some functional significance. The first concerns the cohesive property of the SC. Our group have devised an instrument which measures intracorneal cohesion (cohesography) (7). The method determines the force required to distract a disc of SC with a vertically orientated piston stuck to the skin surface with a cyanoacrylate glue. On the basis that 'hydration' was likely to alter the strength of the SC we used this instrument in the same study as that from which the results in *Tables I* and *VII* are derived, and these observations are set out in *Table VIII*.

It is clear that although the results do not reach statistical significance there is a definite trend towards decreased intracorneal cohesion. We do not have as much experience with this technique as with contour analysis for 'acute hydration' experiments.

Table VIII. Results obtained after application of aqueous cream for 14 days using cohesography technique

Time (days)	*n*	Mean grams force ± S.D.
0	7	109·1 ± 49·0
7	8	94·4 ± 34·3
14	5	85·2 ± 19·9

Table IX. Cohesography results before and 3 h after application of HA

Time (h)	Grams force (Mean ± S.D.)
0 (control)	222 ± 69·2
3	157·7 ± 64·3

Table IX, however, documents one such experiment using an oil in water emulsion in three normal subjects.

We have also started to use a method that measures 'point penetrability' (8) and which depends on the measurement of the force required for a rapidly moving needle to move a measured distance into the SC. So far we have found that 'acute hydration' with a water soaked gauze pad results in a drop in the force required of 20–30%.

Both the above techniques utilise important physical properties of the SC as indicators of hydration. The next *in vivo* technique I want to briefly mention does not appear to be a measure of an important physical property, but is probably a function of many variables. This is the measurement of the electrical properties of the skin. Clar, Her and Sturelle (9) have made significant contributions to this topic and this group claims that low frequency impedance is located almost entirely in the horny layer. They and others (10) have derived complex relationships between impedance and hydration based on the movement of ions in hydrated SC. Certainly, these groups appear to be able to evaluate HAs using this approach although the constraints necessary to obtain consistent results would appear to render the technique awkward for routine use. Our own experience with electrical measurements is very limited and although we believe that the techniques can be useful we intend to evaluate them alongside the methods with which we are more familiar, before making a commitment to their regular use.

Another approach is to measure the changes in transepidermal water loss (TEWL) after hydration. The water barrier property of the SC is quite dramatically changed after relatively short hydration. For example, Quattrone *et al.* (1) report a 20–40% increase in TEWL after 5 min hydration which reverts to normal within a 30 min period. When an 'emollient cream' was used there was a 20% decrease after 1 h and a 10% decrease after 4–6 h. The decrease after the use of an occlusive HA presumably results in hydration of the SC from the movement of H_2O into the SC from the living tissue beneath. This same author reviews other workers' experiences employing TEWL.

(e) *In Vitro* TECHNIQUES

It is difficult to be enthusiastic concerning *in vitro* techniques for the evaluation of HA. The SC is so utterly dependent *in vitro* on the ambient environmental temperature and relative humidity (11) that controlled environments and long periods of equilibration are necessary. Furthermore, the application of topical agents to the *in vitro* specimen does not accutely emulate the 'in use' situation. The *in vitro* techniques that have been used include measurements that depend on extensibility and elasticity (12, 13) differential

scanning calorimetry (14), gravimetric techniques (15), and photoacoustic spectroscopy (16). The last of these has recently been adapted for *in vivo* use and I am told (Pines, personal communication) that the early results are extremely promising.

Clearly such techniques may be useful to screen compounds before formulation or to attempt to dissect out the mode of action of HAs but I doubt whether they will have much application to the routine evaluation of these materials.

DESCALING AGENTS (DAS)

We are in an even worse state for the evaluation of DAs than for HAs and I could begin and end by saying that there are no useful techniques available. I will attempt, however, to summarise briefly our experience so far and indicate what further work, we plan. It seems that the most efficient method to date depends on clinical evaluation. Van Scott and Yu (17) used patients with ichthyotic disorders to investigate the descaling effect of a number of alpha-hydroxy acids and were certainly able to make recommendations on the basis of their clinical observations. Our own studies confirm the usefulness of this approach. *Figure* 4 shows a patient with a rare and severe form of ichthyosis whose right arm was treated with 6% salicylic acid in white soft paraffin twice daily for 10 days and whose left arm received just the vehicle. Fortunately, such patients are rare and it is just not a practical proposition to use this approach routinely.

Clearly, it would be ideal to investigate DAs on normal human (or less ideally on animal) skin *in vivo* or *in vitro*. The real problem is knowing which parameter to measure. I have implied by my use of the term DA that the best kind of measurement would be on the rate of desquamation. Unfortunately, this is a most difficult measurement to make. We have examined the number of corneocytes liberated after a standardised 'scrub' stimulus to the skin surface using a specially constructed apparatus (18). We have not as yet, however, satisfactorily demonstrated increased numbers of cells liberated from a DA treated site. It may be that this technique is not sufficiently sensitive to pick up differences in normal skin. It might also be mandatory to use abnormally keratinised skin to demonstrate the effect. It is possible, however, to demonstrate the action of DAs on normal SC by either SEM (19) using SSBs, or by actually counting the cell strata within the SC. This last technique is quite promising but unfortunately does necessitate biopsy. The tissue is sectioned on a cryostat and the SC demonstrated by the McKenzie technique (20). *Figure* 5 is an SEM of SC taken after 10 days use of 6% salicylic acid in white soft paraffin and *Fig.* 6 shows cryostat sectioned skin treated by the same material compared to a vehicle treated specimen. Other workers (21) have also demonstrated this SC thinning effect of DAs although the way they accomplish this remains mysterious. *In vitro* testing of DAs has been totally unsuccessful in our hands. We have applied keratolytics to SSBs themselves and demonstrated no change in surface contour. We plan many more studies with DAs both *in vitro* and *in vivo* using measures of the rate of cell loss and the tensile properties of the SC.

CONCLUSION

We are on the threshold of an exciting era in dermatology and cosmetic science. I believe that in the not too distant future there will be accurate and convenient techniques for the evaluation and measurement of many of the skin's properties and functions. This certainly appears to be the case for HAs and I am certain that with perseverance the same will shortly be true for DAs as well.

REFERENCES

1 Quattrone, A. J. and Laden, K. Physical techniques for assessing skin moisturization. *J. Soc. Cosmet. Chem.* **27** 607 (1976).

2 Middleton, J. D. and Roberts, M. E. Efficacy of a skin cream containing pyrrolidone carboxylic acid in reducing the incidence of subclinical dry skin. In: Marks, R. and Dykes, P. J. *The Ichthyoses.* 177 (1978) MTP Press Ltd, Lancaster.

3 Garber, C. A. and Nightingale, C. T. Characterizing cosmetic effects and skin morphology by scanning electron microscopy. *J. Soc. Cosmet. Chem.* **27** 509 (1976).

4 Bernstein, E. O. Scanning electron microscopy of skin topography before and after treatment. Ann. Meet. Soc. Cosmet. Chem. December 1974, New York City.

5 Marks, R and Dawber, R. P. R. Skin surface biopsy: an improved technique for the examination of the horny layer. *Brit. J. Dermatol.* **84** 117 (1971).

6 Marks, R. and Pearse, A. D. Surfometry: a method of evaluating the internal structure of the stratum corneum. *Brit. J. Dermatol.* **92,** 651 (1975).

7 Marks, R., Nicholls, S. and Fitzgeorge, D. Measurement of intracorneal cohesion in man using *in vivo* techniques. *J. Invest. Dermatol.* **69** (3) 299 (1977).

8 Nicholls, S., King, C. S., Guibarra, E. and Marks, R. Measurement of point deformation (PD) of human skin in vivo: contribution of the stratum corneum. Ann. Meet. Eur. Soc. Dermatol. Res. Amsterdam (1978).

9 Clar, E. J., Her, C. P. and Sturelle, C. G. Skin impedance and moisturization. *J. Soc. Cosmet. Chem.* **26** 337 (1975).

10 Edelberg, R. Relation of electrical properties of skin to structure and physiologic state. *J. Invest. Dermatol.* **69** 324 (1977).

11 Agache, P., Boyer, J. P. and Laurent, R. Biomechanical properties and microscopic morphology of human stratum corneum incubated in a wet pad *in vitro. Archiv. für Dermatol. Forschung.* **246** 271 (1973).

12 Middleton, J. D. The effect of temperature on extensibility of isolated stratum corneum and its relation to skin chapping. *Brit. J. Dermatol.* **81,** 717 (1969).

13 Ferguson, J. and Agache, P. Influence of site, storage and trypsin treatment on the mechanical properties of the stratum corneum. *J. Invest. Dermatol.* **68,** 256 (1977).

14 Miller, D. L. and Wildanauer, R. H. Thermomechanical probes for the analysis of physical properties of stratum corneum. *J. Invest Dermatol.* **69** 297 (1977).

15 Laden, K. & Spitzer, R. Identification of a natural moisturizing agent in skin. *J. Soc. Cosmet. Chem.* **18** 351 (1967).

16 Rosencraig, A. and Pines, E. Stratum corneum studies with photoacoustic spectroscopy. *J. Invest. Dermatol.* **69** 296 (1977).

17 Van Scott, E. J. and Yu, R. J. Control of keratinization with alpha-hydroxy acids and related compounds. I. Topical treatment of ichthyotic disorders. *Arch. Dermatol.* **110** 586 (1974).

18 Nicholls, S. and Marks, R. Novel techniques for the estimation of intracorneal cohesion *in vivo. Brit. J. Dermatol.* **96** 595 (1977).

19 Davies, N. and Marks, R. Studies on the effect of salicylic acid on normal skin. *Brit. J. Dermatol.* **95** 187 (1976).

20 Mackenzie, I. C. and Linder, J. E. An examination of cellular organization within the stratum corneum by a silver staining method. *J. Invest. Dermatol.* **61** 245 (1973).

21 Huber, C. and Christophers, E. 'Keratolytic' effect of salicylic acid. *Arch. Dermatol. Res.* **257** 293 (1977).

J. Soc. Cosmet. Chem. **29** 441–446 (1978)

Evaluation of mechanical stresses set up in lipstick during application

R. G. DREW *Beecham Products, Applied Research Laboratories, Randalls Road, Leatherhead, Surrey, KT22 7RX*

Received 19 January 1978

Presented as part of a lecture for the Society of Cosmetic Chemists of Great Britain, London, January, 1977

Synopsis

A device is described which measures the stresses set up when **lipstick** is applied; this device is used to assess the application properties of two ranges of commercially available lipsticks. The values obtained facilitate definition of the strength requirements of a lipstick. Knowledge of the in-use stresses will enable formulators to avoid fracture problems during lipstick application. A trial of ten commercially the available lipsticks indicates that the application force is related to the ease of colouring of the lips be applied lipstick. Application forces were greater when the lipstick was applied by older subjects.

INTRODUCTION

A lipstick consists essentially of a solid stick in a retractable applicator. In recent years lipstick formulations have altered considerably; hard sticks with high levels of staining dye have been almost completely superseded by softer sticks in which soluble dye has been largely replaced by insoluble dye in lipsticks (pigments). With the reduction in level of staining dye in lipsticks much easier transference is required in order to achieve adequate colouration of the lips. Elliott (1) indicated that this has been achieved by introducing oilier bases. The increased liquid and semi-solid content of recent formulations has made the mechanical strength requirements of the stick critical.

A common method of lipstick strength assessment described by Lauffer (2) is to hold the lipstick horizontally in a socket fitting over 1 cm of its base and to apply weights at a measured distance from the edge of the socket. The weight is increased by increments (suitably 0·1 N) every 30 min until the lipstick breaks. At least four readings are needed for each batch of lipstick and broken surfaces must be examined to ensure that no weakening flaws reduce the strength. Breaking point measurements must be carried out at a known temperature (suitably 25°C) on sticks stored at that temperature for at least 30 min. This technique is unrealistic as the method of loading involves a disproportionately large shear component and a small bending moment when measuring the total force to fracture. Lauffer (2) also described a method of roughly measuring the force of application. A strip of smooth paper is drawn between two flat lipstick ends at constant speed. The force required to pull the paper when a given total weight is applied to the upper lipstick is recorded.

0037–9832/78/0700–0441 $02.00

This paper describes an apparatus giving realistic measurements of the force of (ease of) application and force to fracture.

EXPERIMENTAL

MATERIALS

Ten lipsticks manufactured for retail distribution were selected for examination from two lipstick ranges*. They were chosen to cover a range of colours from clear gloss, through shades of pink, red and light brown to dark brown. These lipsticks were then arbitrarily coded. The first range was coded a, b, c, d and e, and the second range was coded v, w, x, y and z. All lipsticks tested were 41 mm long and 11 mm in diameter at their base.

APPARATUS

The apparatus consisted of a brass beam carrying strain gauges, spike and base onto which the lipstick was impaled (*Fig.* 1). A conventional polystyrene holder to match the base diameter of the lipstick was fixed to the beam with epoxy resin.

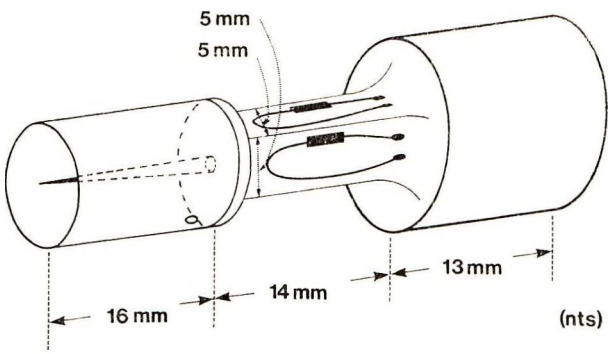


Figure 1. Brass beam, with semiconductor strain gauges, used to hold the lipstick. Sample lipsticks were impaled on the spike set in a cylindrical holder on the left. Applied forces were transmitted to two pairs of strain gauges fitted on opposing sides of the beam.

A semiconductor strain gauge** was fixed to each of the four sides of the square-sectional beam. The strain gauges formed two half Wheatstone resistance bridges. The other half of each Wheatstone bridge circuit was a balancing variable resistor incorporated in the main recording equipment situated away from the device. In the recording equipment the outputs from the two bridge circuits were amplified by a factor of 1000 and fed into a vector sum computation unit.† Hence, by carrying out a vector sum treatment of the two messages the resolved load on the lipstick at any instant could be computed electronically and the output fed to the pen recorder.‡ The bridges were powered

* Margaret Astor Silver Frost and Ultra Soft lipstick from Cofa GmbH, Cosmetik-Fabrikations-Gesellschaft mbH 612 Michelstadt/Odenwald, Bundesrepublik Deutschland.

** The strain gauges used were encapsulated axial-lead type supplied by Kulite Ltd, 20 Wote Street, Basingstoke, Hampshire, U.K. (Code No. DCP–120 090).

† The vector sum circuit was designed by Analog Devices Inc., Norwood, Massachusetts 02062, U.S.A., and based on an integrated circuit chip, AD531 KD.

‡ Phillips pen recorder PM 8221.

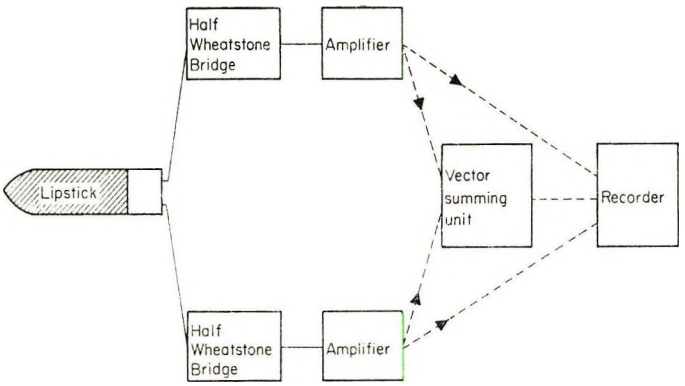


Figure 2. Block diagram of lipstick stress equipment. The applied stresses caused the beam to bend. The resulting strains affected the resistance of opposing strain gauges. The gauges were part of a Wheatstone resistance bridge whose output was then amplified. The two signals were either recorded separately, or combined to give the vector sum (or resolved) force before being recorded.

with a +5 V stabilised power supply. The amplifiers and vector sum device were energised from a stabilised +15, 0, −15 V supply. A block diagram of the equipment is given in *Fig.* 2.

The apparatus was calibrated for pure bending loads by weights being hung from the application point of a lipstick in a horizontally held holder.

LIPSTICK APPLICATION TRIAL

Initially twenty-eight female subjects were asked to describe their method of lipstick application and frequency of use (which varied from non-users, to an hourly application). Each subject selected one lipstick from each range. The first stick was impaled on the strain-gauged beam and the subject was asked to apply the lipstick in front of a mirror in their normal manner. The applied lipstick was removed from the lips with a tissue. This was repeated for a total of six applications. The first stick was removed and replaced by the second stick chosen by the subject from the second range. After repeating the above six applications subjects were asked which stick they preferred. This procedure was repeated on five separate days.

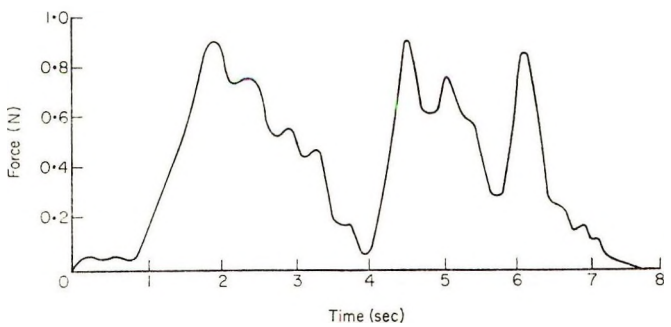


Figure 3. A typical vector sum lipstick application force *v.* time trace. This trace spans 7 sec during which the lipstick was applied in several strokes. The maximum force approached 1 N.

Each application yielded a trace of either the resolved load (via the vector summing device) or the separate components of each resistance bridge on the pen recorder. A typical trace is shown in *Fig.* 3. From each trace the maximum resolved force during an application was recorded.

RESULTS AND DISCUSSION

Maximum application forces obtained from 840 applications were plotted in the form of two histograms (*Fig.* 4).

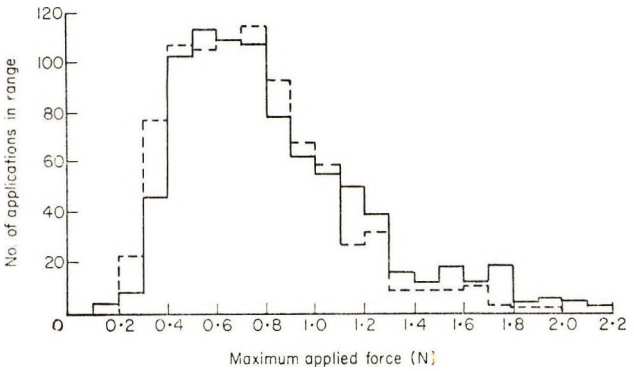


Figure 4. Histogram of the maximum applied forces from each lipstick range during the trial. Most of the subjects applied the lipsticks with a force less than 1 N, but a 2 N force was exceeded occasionally. ——— Ultra Soft; - - - - Silver Frost.

Median forces of application were 0·68 N for the Silver Frost range (10th percentile 0·46 N, 90th percentile 0·92N) and 0·73 N for the Ultra Soft range (10th percentile 0·50 N 90th percentile 1·27 N). The subjects tended to apply the Ultra Soft range using higher application forces than the Silver Frost range. A Mann-Whitney U-test showed that this difference was significant at the $P < 0{\cdot}00001$ level.

Each subject's applications of the same lipstick often covered a wide range of forces (*Table I*). Comparison of the application loads of two lipsticks, however, revealed that within the twenty-eight subjects tested significant differences in maximum application force were found (*Table II*). From these differences a ranking order was compiled indicating which lipsticks would be applied with a greater force: b > c > a > y, e > w > v > x > z > d.

Table I. The maximum application forces observed from one subject using a single lipstick during the trial

Application number	Maximum application force (N)				
	Day 1	Day 2	Day 3	Day 4	Day 5
1	0·604	0·512	0·443	0·660	0·738
2	0·667	0·438	0·708	0·676	0·851
3	0·811	0·678	0·743	0·640	0·808
4	0·397	0·604	0·682	0·996	0·705
5	0·261	0·696	0·732	0·931	0·847
6	0·642	0·774	0·942	0·782	0·829

Table II. Subjects who exhibited significant differences (Mann-Whitney U-test, $P<0{\cdot}02$) in maximum application forces between the two applied lipsticks

Subjects	Ultra Soft *v.* Silver Frost
3, 4	c > w
6	d > w
8, 10, 11	b > z
9	d > z
13	c > x
14	d < x
16, 17	c > v
21	a > y
22, 24	c > y
23	d > y
25	b > y
28	a > y

Four independent volunteers ranked the applied lipsticks in order of colouring power: b > a > c > e, z > v, w > y > x > d. The Kendall Rank Correlation Coefficient, τ, for the ranking of applied load and subjectively assessed covering power was 0·683, indicating that individual subjects would use a lower application force to apply a lipstick with better colouring power.

The subjects were divided into two age groups of approximately equal size (18–29 and 30–60 years old). Subjects over 30 years old used significantly higher application forces (Mann-Whitney U-test, $P<0{\cdot}025$).

A similar analysis of their choice of colours indicated at the $P<0{\cdot}025$ level that the 18–29 years group were biased towards brown colours and the 30–60 years group towards the paler brown, pink or red colours.

Subjects who applied lipstick frequently (daily or more often) used significantly higher application forces than those who applied lipstick less frequently.

Other parameters possibly related to lipstick selection and application were also examined. The subjects' comments upon the applied lipstick was compared with the most common application force range. No overall trend was observed and subjects frequently described the same lipstick using antonyms and the reasons for lipstick selection seemed to have been many and varied. The lipstick choice combinations of a lipstick from each range revealed no bias towards particular colour combinations such as brown and dark brown, or red/pink and pale brown etc. The selection of lipstick was finally compared with subject hair colour but no trend towards any particular choice of lipstick was observed.

DISCUSSION

The main purpose of the work reported here was to assess the capability of the lipstick application equipment; to achieve this an application trial was carried out. The equipment was shown to be capable of meaningful measurements and is a useful tool for assessing the force of application of lipstick.

This trial was designed to assess the maximum application forces of two types of lipsticks. Subjects were expected to apply a lipstick of their own choice in a more typical

manner than a lipstick selected by the investigator and hence only one lipstick of each type was tested by the subject.

Although a difference was detected between the maximum application forces used in each lipstick range, an investigation of the data showed that this may be due to either the colouring ability of the lipstick, or the age of the subject. The colours applied with the greater force were predominantly pale brown or a clear gloss. The mainly dark colours, of better colouring ability, were applied with a smaller force.

The application data implies that lipsticks with relatively low colouring power (usually pale colours) are more likely to fracture during use if the strengths of all lipsticks are equal.

Differences in maximum application force were detected between frequent and occasional lipstick users, but not between users and non-users of lipstick. This may be explained by the selection of lipsticks, because non-users of lipstick more frequently selected glossy and paler lipsticks.

REFERENCES

1 Elliott, T. J. Present and future trends in make-up. *Cosmet. Perfum.* **88** 35 (1973).

2 Lauffer, P. G. I. *Lipsticks.* In: Balsam, M. S. and Sagarin, E. *Cosmetics, Science and Technology* Vol. 1 388 (1972). Wiley-Interscience, New York.

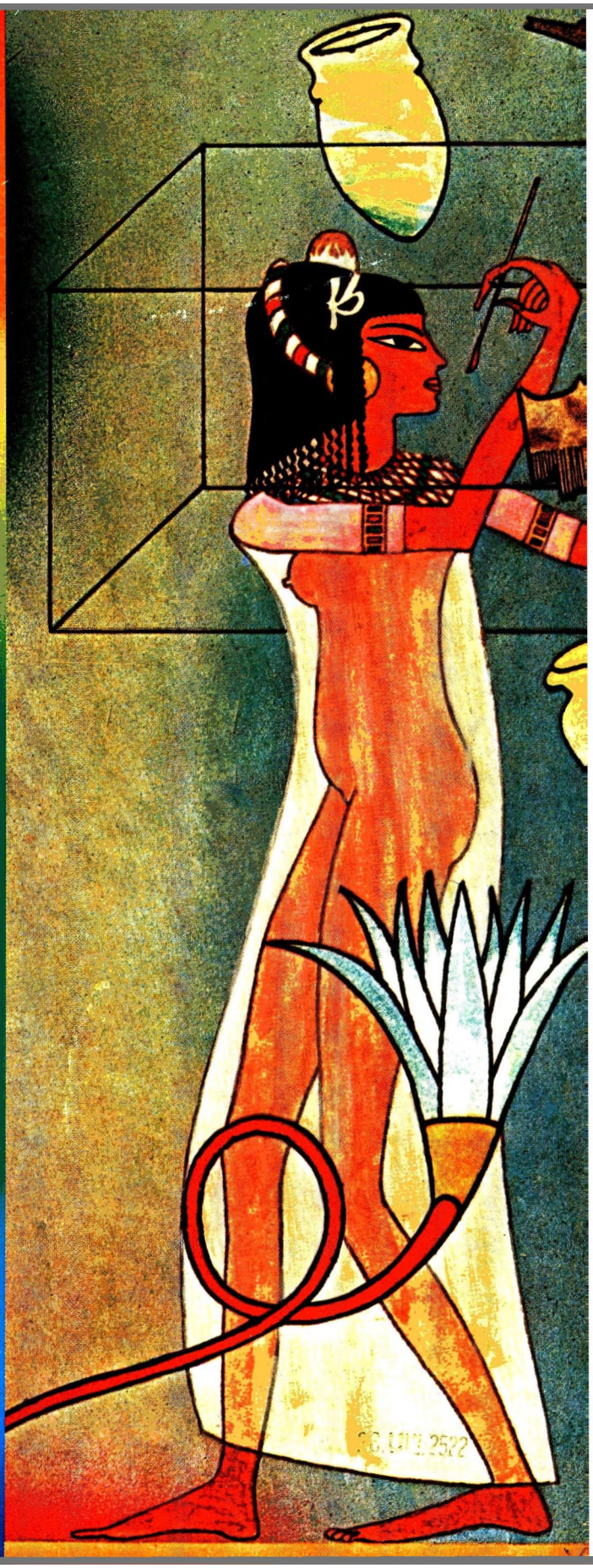
Corrigendum

Changes in sunburn and mechanisms of protection, by B. E. Johnson. *J. Soc. Cosmet. Chem.* **29** 31–44 (1978).

Reference (10) Coblentz, W. Ultraviolet radiation useful for therapeutic purposes: specification of minimum intensity or radiant flux: second communication. *J. Am. Med. Assoc.* **99** 125 (1932) is cited incorrectly for the division of the uv spectrum. Moreover, the wavelength limits for uv-A, uv-B, and uv-C given in the text (p. 32) are not necessarily those universally recognised. It would be better if reference 10 was '*International Lighting Vocabulary*, 3rd edn, 4, Publication CIE No. 17 (E-1.1) (1970), Bureau Central de la CIE, Paris'. The spectrum limits would then be uv-A 315–400 nm, uv-B 280–315 nm, uv-C 100–280 nm.

Published on behalf of the Society of Cosmetic Chemists of Great Britain by Blackwell Scientific Publications Ltd, Osney Mead, Oxford OX2 0EL. Printed in Great Britain by Burgess & Son (Abingdon) Ltd, Station Road, Abingdon, Oxfordshire.

Cosmetic Colorants



Sun distributors in Europe

S. Black (Import & Export) Ltd.
Independent Works
Wennington Road
Rainham, Essex, England

Expandia S.A.
13 Ave. de l'Opera
Paris, France

Nordmann, Rassmann & Co.
Kajen 2
Hamburg, West Germany

S.A. Especialidades Quimicas
Via Augusta 137
Barcelona, Spain

Unitrade Ltd.
8 Constantinoupoleos St.
Nea Smyrni-Athens, Greece

Carbocrom
Via Broni N. 16
20139 Milano, Italy

Sun Chemical Corporation

Pigments Division