

Evaluation of skin bleach creams

K. V. CURRY*

Presented on 12th November 1973 in Nottingham, at the Symposium on 'Evaluation of Product Performance', organized by the Society of Cosmetic Chemists of Great Britain.

Synopsis—The use of test animals to screen DEPIGMENTING CHEMICALS is described and an attempt is made to correlate these results with tests carried out on humans.

A comparison of various techniques used for measuring changes in SKIN LIGHTNESS showed that the method of visual assessment using a simple numerical scoring system was the simplest and gave the most statistically significant results.

Products based on hydroquinone or its derivatives, and designed to lighten the skin colour of the dark skinned races in Africa, Asia and the U.S.A., have been marketed for many years. A considerable amount of data on the depigmenting effects of the chemicals on black guinea pigs is available (1-3) and since the products containing these chemicals have a fairly high volume of sales one must assume that they have some degree of effectiveness on humans. However, although there have been reports in the literature of the treatment of specific problems of skin hyperpigmentation with hydroquinone compounds, no work on the quantitative evaluation of the depigmenting effect of skin lightening creams on coloured races has yet been published.

There are two closely related problems that are encountered in the selection of suitable depigmenting agents for skin lightening creams. The first is that of setting up a suitable test method for the rapid screening of large numbers of compounds in order to narrow them down to one or two with the required effects. Coupled with this is the problem of selecting a method to measure changes in skin colour produced by the test material (4).

* Unilever Research, Isleworth Laboratory, 455 London Road, Isleworth, Middlesex, TW7 5AB.

PRELIMINARY SCREENING TESTS

The classical method of measuring the depigmenting effect of chemicals is to use black-skinned guinea-pigs which do not pose very great problems in rearing, handling and testing. It is also generally agreed that the skin of the guinea-pig has enough similarity to human skin to make comparisons meaningful.

Many of our early screening tests were in fact carried out on black guinea-pigs and many compounds and combinations of compounds were examined for their depigmenting effects on these animals. However, the interpretation of the results of these tests was difficult since the degree of lightening could not be extrapolated to humans.

It was decided, therefore, to evaluate four creams containing various depigmenting agents first of all on black guinea-pigs, then on black pigs, and finally on a panel of humans. This was considered necessary in order to:

- ascertain the usefulness of guinea-pigs as a screening technique by correlating the results with those obtained on humans;
- determine the usefulness of black swine as an alternative test animal;
- establish the best practical method for measuring changes in the lightness of skin.

TESTS ON BLACK GUINEA-PIGS

In this test four products were evaluated for their depigmenting effect on black-skinned guinea-pigs:

Product A	}	Commercially-available skin lightening creams
Product B		
Product C	}	Experimental skin lightening creams
Product D		

(Products C and D were simple water-continuous emulsions containing about 20% oil phase.)

The creams were applied daily for 26 days in a latin square design to the shaved backs of four black-skinned guinea-pigs.

During the course of the experiment a visual estimation of the lightening effects was made at intervals using a set of grey standards. These standards formed a photographic step wedge consisting of shades of grey and each step was numbered arbitrarily with increasing lightness from 1 to 7 (*Fig. 1*).

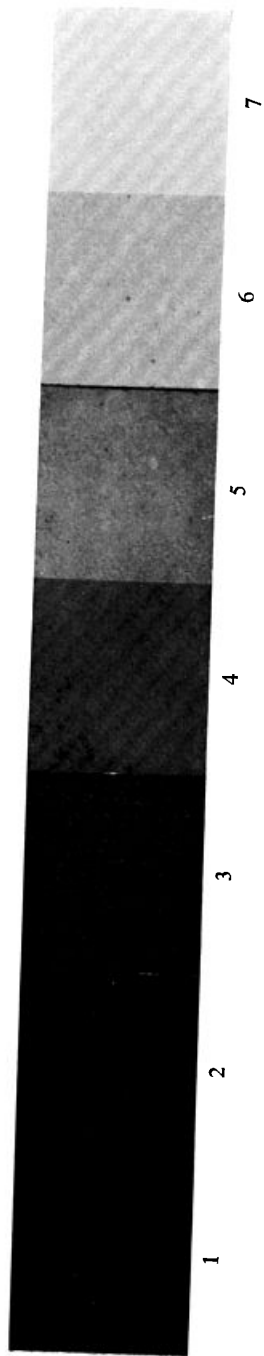


Figure 1. Photographic wedge.

(Facing p. 340)

Measurements of skin lightness were made by holding the wedge alongside the animal and noting the step numbers which most nearly matched each site.

The time course of depigmentation produced by the four products is shown in *Fig. 2* where each point is the mean of four observations. Product A was apparently the most effective of the products tested since it produced a more rapid and a greater degree of depigmentation than the other creams which were all similar in performance. It was noted however that this product (Product A) was also somewhat irritant, producing more scaliness than the other products and so giving a false impression of enhanced whiteness.

One noticeable feature of tests of this sort carried out on black guinea-pigs was that the maximum lightening effect was achieved after about 15 days.

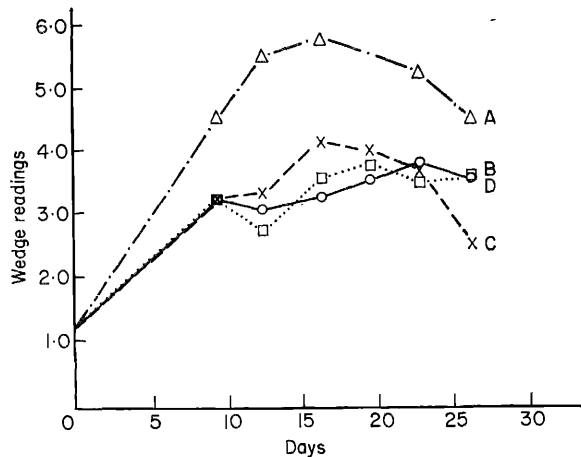


Figure 2. Time course of depigmentation on black-skinned guinea-pigs—wedge technique.

TESTS ON BLACK-SKINNED SWINE

Whilst the guinea-pig is a suitable animal for screening large numbers of compounds, it was considered advisable to obtain comparative data on at least one other black-skinned species other than man. It was not easy to find another suitable species since most animals with black hair have, in fact, white skin. The pig appeared to be a satisfactory choice for a number of reasons.

- (1) The anatomy of the skin is similar to that of man.
- (2) The hairs are more sparse than on the guinea-pig.
- (3) A large area of skin is available for application of materials.

It should be borne in mind, however, that the physiology of pig skin may be completely different from that of human skin in that the barrier properties of the stratum corneum may be different, resulting in differing rates of penetration of active ingredients. The sensitivity of the melanocytes to skin depigmenting agents may also be different.

The animals used in this test were nine pure bred Wessex Saddlebacks (8 hogs and 1 gilt) and at the beginning of the experiment the pigs were shaved and the pigmented area of the back and flanks divided into sites about 5 cm square using a *Magic Marker*. Sites had to be frequently re-marked during the experiment. The skin reflectance was measured using an *EEL* reflectometer which in this case was considered to be more useful than the stepped wedge since the colour of the pig skin was less grey than that of the guinea-pig. The instrument consisted of a reflectance head comprising a photocell and a light source connected to a *Unigalvo 20* galvanometer. Reflection readings of the skin were expressed as a percentage of the reflectance of a standard MgCO_3 block (the reflectance of which was set at 100%). The creams were packed in tubes and applied daily for 38 days, 6 mm of cream being applied to the relevant site on the animal. Measurements of skin reflectance were made at weekly intervals, after washing the skin with water to remove traces of dirt and old cream.

The time course of depigmentation is shown in *Fig. 3*. These results are expressed as the change in reflectance of the skin compared to its initial reflectance before treatment. Positive values indicate lightening whilst darkening is shown by negative values. Each point on the graph is the mean of nine observations (one reading on each of nine pigs).

A statistical analysis of the total results (confined to the results recorded on day 38, that is, the final level of depigmentation reached), showed that there was no significant difference between any of the four products tested.

A comparison of *Fig. 2* and *Fig. 3* does show, however, that the depigmentation occurs more slowly on pigs than on guinea-pigs and it was considered highly likely that depigmentation would be slower still on humans.

The *EEL* proved to be a simple instrument to use although there were occasions when it did not seem to be differentiating markedly between sites that appeared visually to be of different lightness. This is probably because the eye measures the contrast between a specific site and its background and the observer can also ignore differences in surface texture such

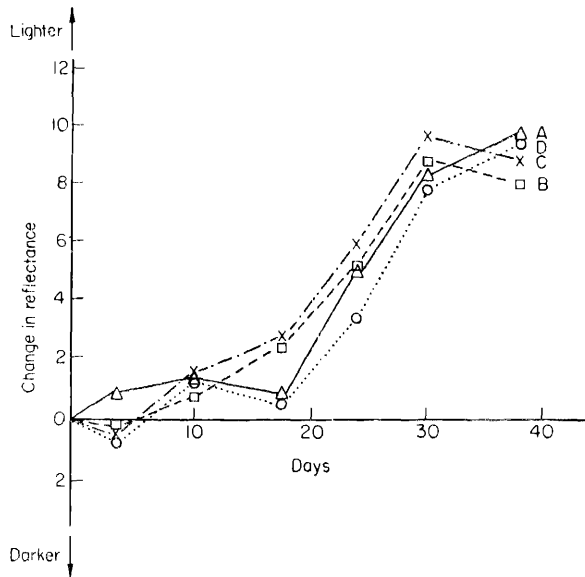


Figure 3. Time course of depigmentation on black-skinned swine—reflection technique.

as scaling due to irritation, and stubble, which may in fact contribute to the surface reflection measured by the instrument.

At this stage, therefore, we had two sets of rather conflicting data, neither of which was satisfactory. Although all creams had a marked depigmenting effect it was still not clear whether Product A was more effective than the others. It was felt at the time that the irritation of A may have affected both the wedge and reflectance values. It was at this stage that it was decided to measure the effectiveness of skin lightening creams on humans and for the time being to disregard the animal tests.

TESTS ON HUMANS

In order to demonstrate the efficacy of skin lightening creams on coloured subjects and if possible to correlate any effects achieved with those shown on the experimental animals, the four products previously described (A, B, C, D) were tested. The opportunity was also taken to compare various techniques of measuring changes in the lightness of skin.

Test method

The test was carried out in Africa on a panel of 40 male volunteers. Most of the subjects were chosen because they had jobs which caused them to spend a considerable amount of their time out of doors, since it was suspected that strong sunlight might be playing some part in the efficiency of the bleach creams. The groups were matched on the basis of the reflectance values of the skin on the outside of the forearm. Each group used only one of the previously mentioned products during the whole test period.

Application of the creams

These creams were applied to four test sites on each person, namely the outside upper forearm and the outside of the wrist on each arm, by means of a cardboard template. In this way the central portion of the outside forearm could act as the control site for each arm. It was also hoped that the subjects would agree to apply the appropriate skin bleach cream product to their faces, but only a few were willing to do so. The cardboard template was located on the outside of the forearm by placing one end on the head of the ulna bone of the wrist. The presence of the template caused some difficulty during the rubbing in of the creams and it was found that the operators applying the cream became so adept in judging the area of application that the template was only used to locate the application area and then removed when the cream was rubbed in. The creams were applied twice daily during the working week (5 days) by skilled operators (the few subjects who agreed to use the cream on their faces were allowed to apply it themselves, but only to their faces). Approximately 0.25 g of cream was applied to each test site on each occasion.

Since the test lasted 8 weeks there was a total of $10(\text{subjects}) \times 40(\text{days}) \times 2(\text{applications}) \times 4(\text{test sites}) = 3200$ applications of each cream during this test.

Examination

The subjects were examined before the test began using a variety of subjective and objective evaluation techniques. In addition the panel members were examined by a consultant dermatologist to ensure that they were

suitable for the test (i.e. had no existing diseases or history of allergy/sensitization), and also during the test in order to detect any signs of primary irritation that might develop.

Evaluation techniques

Both subjective and objective methods for assessing the degree of skin lightening were used in this test. To avoid any bias, conscious or unconscious, in the subjective evaluations, it was decided to carry out the subjective evaluations before the objective measurements were made. In addition the panel members were examined in a completely random order which was dictated by their availability.

The evaluation techniques employed were the following.

Grey scale wedge assessment

The shade intensity of colours of the skin on both the test sites and control sites of the forearms were matched with the same set of numbered photographic standards as that used in the guinea-pig tests (*Fig. 1*).

Visual assessment

The colours of the treatment sites were visually compared with the control site on each arm and the degree of lightening was assessed using the following simple numerical scoring system:

<i>Score</i>	<i>Degree of Lightening</i>
0	No difference
+1	Slightly lighter
+2	Noticeably lighter
+3	Markedly lighter

Photographic assessment

Photographs were taken of the forearms and faces of all subjects initially and of the forearm on each examination. Final photographs were taken of the faces of a few subjects who agreed to apply the creams to their faces.

The photographs of the forearms were taken with a 'Startec' clinical camera fitted with a flash attachment and the forearms were photographed against a standard grey background. The background was used as an additional control when micro-densitometer traces were made of the test and

control sites of the forearms, since it was thought that differences in photographic developing conditions could lead to a wide variation in photographic image density. Photographs of the face were taken with a 35 mm camera. All the photographs were taken in colour.

Reflectance of skin

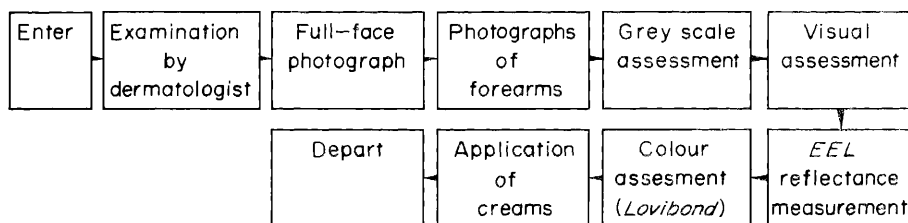
The reflectance of the skin on the test and control sites of the forearms and on the forehead of those subjects applying products to their face was measured using the *EEL* spectrophotometer. 'White Light' was used for the measurements, and the instrument was calibrated to give a reading of 50 units when using the grey tile standard. This instrument had been found to give quite good results in measuring lightening in test animals.

Colour of skin

The colour of the skin was measured with a reflectance *Lovibond* tintometer. The instrument consists essentially of a reflectance head to which are attached two light guides, one delivering the incident light and the other taking away the reflected light. The reflected light is transmitted to a monocular eyepiece where it constitutes half of the field of view while the other half is obtained from a separate source whose colour can be altered by the insertion of various coloured filters. The probe was placed on the skin surface and the colour of the skin matched by means of the red, blue and yellow filters. The colour of the skin was then designated in terms of combinations of these colours required to match the colour of the skin surface.

In all, therefore, there were six different types of assessments made on each subject on eight different sites on each formal examination after 0, 1, 2, 4 and 8 weeks.

A simple flow chart demonstrating the evaluation procedure is shown below.



Various difficulties were encountered with some of the evaluation techniques and these will be dealt with later. However, it was almost immediately apparent that the use of the *Lovibond* tintometer to measure the colour of the skin was a laborious and time-consuming exercise, and since it was found that it took approximately 15 min to carry out the colour measurements on the eight sites on an individual (i.e. 10 h for the whole panel), it was decided to confine this evaluation to only a few randomly-selected individuals.

RESULTS

Grey Wedge Assessment technique

In previous tests carried out on black guinea-pigs this technique had been found to be very useful, and so it was chosen as an evaluation technique in Africa. However, none of the Africans on the panel were really black and ranged from dark red-brown to light brown and considerable difficulty was experienced by all assessors in matching the grey scale with the coloured skin.

The results were analysed statistically and are shown in graphical form in *Fig. 4*.

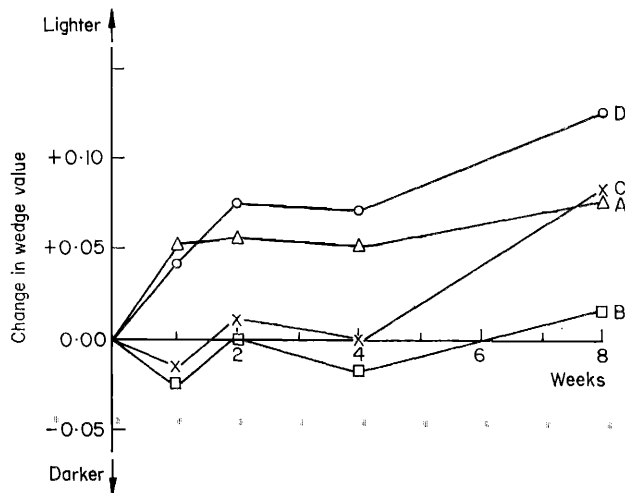


Figure 4. Time course of depigmentation on coloured human subjects—wedge technique.

The main conclusions from this test are as follows.

None of the products gave a statistically significant lightening effect after 4 weeks.

Only Product C and Product D gave a significant lightening effect after 8 weeks.

There was no significant difference between any of the products after 8 weeks.

The lack of statistical significance arises from the very wide scatter of results obtained with this technique, which is indicative of the difficulty of using the grey wedge on brown-coloured skin.

Visual assessment

This was an extremely simple technique to use and the scoring system is one that has been used successfully for many years. The results of this test which were statistically analysed are shown graphically in *Fig. 5*.

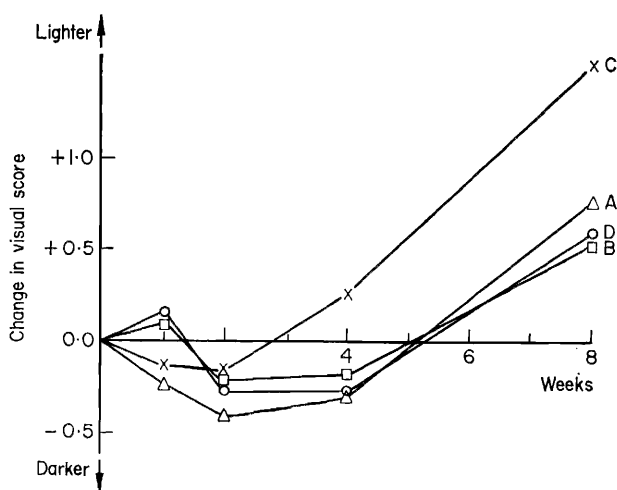


Figure 5. Time course of depigmentation on coloured human subjects—visual technique.

The main conclusions were as follows.

None of the products had a significant lightening effect at the end of 4 weeks.

All products had a significant lightening effect at the end of 8 weeks.

Product C produced a significantly better lightening effect than the other three products, which were equivalent, after 8 weeks. The one interesting feature of *Fig. 5* is that it shows that the effect of Product C on the skin was to produce an effect which was between 'slightly lighter' to 'noticeably lighter' than the control area after eight weeks. The other three products gave effects that were much less than 'slightly lighter'.

Photographic assessment

Forearms

Visual examination of these photographs showed that there would be obvious difficulties in carrying out traces with the microdensitometer. This was due to the presence of very strong highlights on various parts of the arm arising from the use of a flash attachment and the different curvatures on control and test sites. After careful consideration it was decided not to carry out microdensitometer traces on forearms.

Faces

The same sort of difficulty due to highlights was also experienced on the photographs of subjects' faces before and after the test and therefore no microdensitometer traces were carried out on these photographs.

It was disappointing that the photographic technique did not yield any useful information, but this was probably due to the fact that the lightening effects were much smaller than anticipated.

Colour of skin (Lovibond)

As mentioned previously this was a slow and laborious method for measuring the skin colour and only a few subjects were evaluated with this technique. An examination of the results of this assessment did not reveal any pattern of change of colour either with regard to the product used or the length of treatment.

Reflectance of skin (EEL spectrophotometer)

This direct measurement made on the forearms (and foreheads where applicable) was a very simple technique to carry out. The only experimental

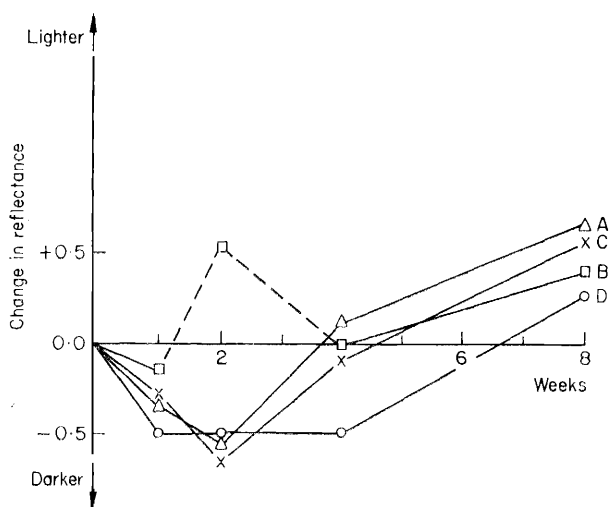


Figure 6. Time course of depigmentation of human subjects—reflectance technique.

difficulty observed was in a slow drift of the galvanometer but this was overcome by frequent recalibration with the standard grey tile. The results are shown in graphical form in *Fig. 6*.

A statistical analysis of the results of this test showed the following.

None of the products produced a significant lightening effect after 4 weeks.

Only Product A had a significant lightening effect after 8 weeks.

There was no significant difference between any of the four products after 8 weeks.

Product B did give a significant lightening effect after 2 weeks but this appears to be an anomalous result.

The lack of statistical significance is perhaps again indicative of the difficulties associated with instrumental measurements on non-homogeneous surfaces. The skin surface on the arms of quite a few subjects showed considerable scarring due to previous wounding of the skin (both deliberate tribal cuts and accidental wounding), and since the area of skin evaluated by the spectrophotometric head was only about 1 cm² this gave rise to considerable variation in reflectance readings owing to slight differences in positioning the reflector head. However, the results are still useful since they also indicate that appreciable lightening effects are not obtained until between 6 and 8 weeks. The high reflectance values for Product A in this test

may be due to the irritation properties of this product causing 'white scaling' on the skin surface which would increase the reflectance measurement. This effect has been noted previously in animal tests.

Because of the large variation in initial skin colour of the subjects it was decided to analyse the test data to determine whether the products had greater or less lightening effect on dark skins than on light coloured skins. The panel was split into three groups for the purpose of statistical analysis.

Group 1 Those giving a skin reflectance of less than 28 units.

Group 2 Those giving a skin reflectance of between 28 and 32 units.

Group 3 Those giving a skin reflectance of more than 32 units.

The statistical analysis was carried out on Group 1 who were arbitrarily classified as 'dark' and also on Group 3 who were classified as 'light'.

The test did in fact reveal that the effects of the various products were more apparent on the subjects with 'dark' skin than on those with 'light' skin. Unlike the results from the total panel, the dark-skinned subjects all showed significant lightening effects after eight weeks whereas none of the light-skinned panel showed any significant lightening effects (*Fig. 7*).

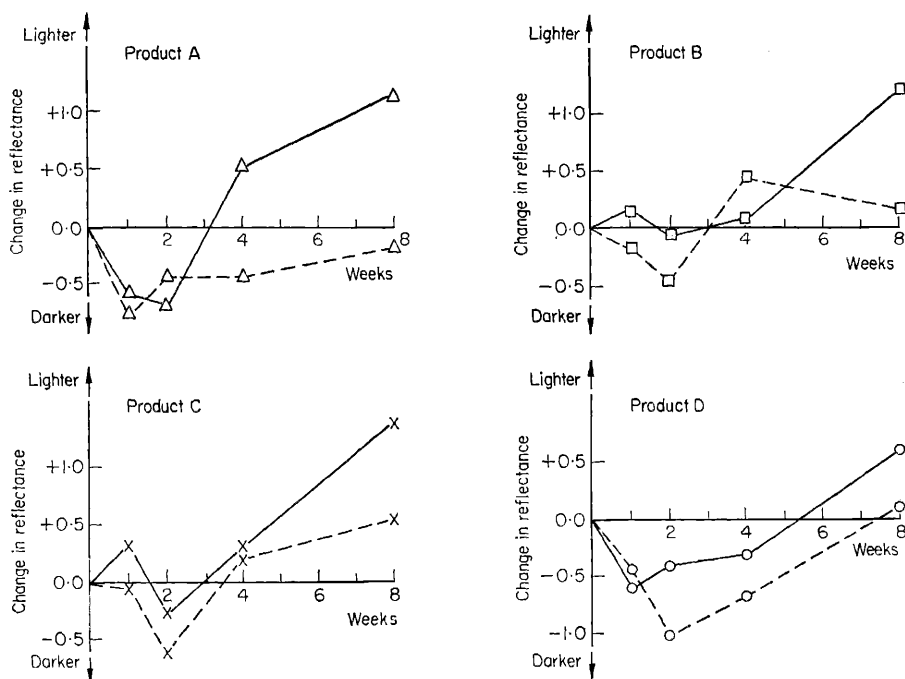


Figure 7. Variation of depigmenting effects of Products A-D on (—) dark-skinned and (- - -) light-skinned human subjects.

Although the three groups were arbitrarily chosen, and the perception of greater or lesser lightening effects on groups may change depending on the group boundaries chosen, it is still nevertheless apparent that clinical tests carried out on light-skinned subjects would need larger panels and longer testing periods to show significant differences.

DISCUSSION

Of the various techniques used to measure lightening of the skin, only three gave results that were capable of being analysed. These techniques were (i) the grey scale; (ii) *EEL* reflectance technique; (iii) comparative visual assessment. A comparative summary of the results obtained with these three techniques is shown below.

	Grey wedge	Visual	Reflectance (<i>EEL</i>)
Lightening effect after 4 weeks	None	None	None
Lightening effect after 8 weeks	Only products C and D were effective	All products effective	Only product A effective
Differences between products	None	C better than ABD	None

It is apparent that the only technique capable of clearly showing the depigmenting effects of all the skin lightening creams is that based on comparative visual assessment. This is probably due to the fact that the eye can compensate for odd blemishes, scaling due to irritation and can also integrate the lightening effect over the total area of skin treated.

The technique is extremely simple to use, and the scatter of results compared to the other techniques is smaller which gives a greater degree of statistical significance. In addition since it is a visual technique, it is giving a measurement of effectiveness of the products which is directly related to the way in which users judge the effectiveness of skin lightening creams.

ACKNOWLEDGMENTS

The author is indebted to many of his colleagues at Unilever Research Laboratory, Isleworth and Unilever Research Laboratory, Colworth House,

for their assistance in carrying out the animal tests and to Messrs. Wilkinson and Allright for their invaluable assistance during the tests carried out in Africa.

REFERENCES

- (1) Beehan, S. S., Pathak, M. A., Hori Y. and Fitzpatrick, T. B. Depigmentation of the skin with 4-isopropylcatechol, mercaptoamines and other compounds. *J. Invest. Dermatol.* **50** 103 (1968).
- (2) Frank, E., Pathak, M. A., Szabo, G. and Fitzpatrick, T. B. Selective action of mercaptoethylamines on melanocytes in mammalian skin. *Arch. Dermatol.* **97** 465 (1968).
- (3) Gellin, G. A., Possick, P. A. and Perone, V. B. Depigmentation from 4-tertiary butyl catechol—an experimental study. *J. Invest. Dermatol.* **55** 190 (1970).
- (4) Gibson, I. M. Measurement of skin colour *in vivo*. *J. Soc. Cosmet. Chem.* **22** 725 (1971).