

Patch tests versus use tests in skin irritation risk assessment

DAVID BASKETTER, GEOFF GILPIN, MICHAEL KUHN¹, DICK LAWRENCE, FIONA REYNOLDS AND
ED WHITTLE

SEAC Toxicology, Colworth, Unilever Research, Sharnbrook, MK44 1LQ, UK

¹Elida Fabergé GmbH, Hamburg, Germany

In the risk assessment of the ability of a substance or preparation to cause skin irritation, a patch test involving human volunteers is often seen as providing definitive information. However, the skin exposure conditions found in an occluded patch may be far removed from those in a particular use situation, not least for such reasons as the high dose per unit area, the duration of exposure and the maceration of the skin. Whilst these factors may enhance sensitivity in the identification of intrinsic hazard, they may do little to ensure accurate safety evaluation for real use. In this paper, we report data from a series of studies with an unmarketed facial skin cosmetic product. Whilst the product was unexpectedly highly irritating in a standard patch test in 30 volunteers, subsequent use tests, including a 6× daily open application to the elbow for 3 weeks and 2× daily application in a half-face test lasting 3 to 4 weeks and involving 52 volunteers failed to show any evidence of skin irritation. It is concluded that the most meaningful results for skin irritation risk assessment are likely to come from studies which involve relevant patterns of exposure.

Key words: skin irritation; risk assessment; patch tests; use tests; cosmetics. © Munksgaard, 1998.

Accepted for publication 16 June 1998

Whilst frank skin irritation to non-surfactant-based facial cosmetics (in this case an aftershave balm) is probably not one of the most important causes of facial eczema, it is nevertheless important that such products are of sufficiently low irritation potential, not least because this is a requirement dictated by the EC Cosmetics Directive (1). One of the common ways in which the skin irritation potential of cosmetic products is examined is by means of a single occluded patch test in groups of human volunteers (2, 3). The occlusion and relatively long duration of contact are deemed to exaggerate exposure, providing a sensitive assessment of skin irritation which may be then extrapolated to the anticipated use conditions. However, the basis for such extrapolation is not well characterized. In the present work, we report on a series of studies which highlight the problem of extrapolation from a patch test, leading in our view to the conclusion that, at least in some cases, accurate risk assessment demands that relevant patterns of exposure are employed in human volunteer studies.

Materials and Methods

Materials

The aftershave balm test material (balm) was a fully formulated, but unmarketed, Unilever product. In part of the work, the balm with either the allyl esters or other fragrance components removed was used. These modified balms were prepared by Elida Fabergé, Buxtehude, Germany. The standard materials used in the test were sodium dodecyl sulfate (SDS, 99%), which was used as a benchmark irritant and was obtained from BDH, Poole, UK, a proprietary skin cream with a well-established positive reactivity under patch test conditions as the positive control, and a proprietary skin cream with a well-established low reactivity under patch test conditions and a long history of safe marketing as the negative control.

Patch test

Proprietary patches, Al-test units (Imeco ac, Södertälje, Sweden), were prepared using adhesive

medical tape (e.g., Leukosilk, Micropore). The test article(s), distilled water (negative control) and the positive control, 0.3% sodium dodecyl sulfate, were applied to individual sites on the patch, according to a randomized sequence. The patch was then applied to the upper outer arm of the panellist and secured with further adhesive tape to ensure proper occlusion. After 23 h, the patch was removed, the sites wiped clean of excess test material, marked and then assessed 1 h later, by a trained assessor. (The quality of training of the assessor is vital to the proper and successful conduct of this type of human volunteer study (5).) The scoring scheme (simplified from 4) is shown in Table 1. At this point, if distinct irritation was evident, the site would not be retreated. The remaining test sites had a replacement patch (identical to the first) applied for 23 h after which time the procedure of wiping, marking and assessing was repeated. A further assessment was performed at 72 h to check that any skin irritation reactions produced were subsiding. The results of the patch tests have been presented as means of the erythema scores taken at 24 h after the application of the 1st patch. There was no difference in the patterns of reactivity at 48/72 h, but the degree of irritation was lessened at the later time point; other endpoints scored, including dryness and oedema, were not changed significantly by the treatments so, for the sake of clarity, the data has not been included.

Elbow test

24 volunteers (age range 20–52 years) were asked to apply 1 material (the balm) to 1 elbow area and the other material (the balm without allyl esters) to the opposite elbow area, 6× daily for 21 days. Treatment sites were the bony protuberances on the inner surfaces of the elbow, effectively being the skin between the elbow crease and point of the elbow. The actual mean number of treatments achieved per day was 5.6 and the mean amount of

sample used per treatment for both materials was approximately 0.1 g (range 0.04 g–0.17 g). The samples were coded to ensure that neither the volunteers nor the trained assessors knew the allocation of materials. Visual assessments of erythema and dryness were made on days 1, 3, 7, 9, 11, 15 and 17 of the treatment period and on days 22 and 24 to monitor any post-treatment effects.

Half-face test

The panel comprised 52 male volunteers, age range 24–58 years. Test materials were coded to avoid bias. 1 material (the balm) was applied to 1 side of the face, the balm without allyl esters to the other side of the face. The treatment/assessment sites comprised the cheeks, jawline (including the area between the lip and nose and the chin) and the neck (including the underside of the chin). Volunteers were asked to apply the materials 2× a day, 1× in the morning and 1× late in the evening, for up to 25 days. Materials were applied after shaving and/or normal face cleansing. Visual assessments were conducted on days 1, 3, 5, 8, 10, 12, 15, 17, 19, 22 and 25. The mean number of applications per day achieved was 1.95, with the dose per treatment (i.e., to 1/2 of the face) being approximately 0.5 g (range 0.1 g–1.0 g). An additional and separate group of 19 volunteers who had no treatment were used to provide background data to help in the evaluation of the results.

Volunteers

For all of the human studies described, ethical support and fully informed written consent from each volunteer was obtained. The age range of the panels varied from 20–58 years and, except in the half-face test, comprised at least 33% of either sex. All volunteers were Caucasian, generally healthy, and free of skin disease at the time of the study.

Results

Patch tests

The initial patch test comparison of the test product against positive and negative control skin creams, the latter with an extensive history of consumer acceptability, is shown in Table 2. The balm was clearly of much greater irritancy under these conditions than the positive as well as the negative control. The SDS data provides a benchmark to help compare data between individual patch test studies.

Since the balm ingredients were generally well known, suspicion fell on the perfume. Table 3 displays the results obtained by testing the balm and

Table 1. Skin irritation scoring scheme

Grade	Description*
0	no visible skin reaction
1	barely perceptible erythema which may not cover the test site
2	weak erythema usually covering the test site
3	moderate erythema covering the test site
4	strong erythema possibly spreading beyond the test site
5	intense erythema spreading beyond the test site

* Whilst other elements of the skin reactions were scored (e.g., dryness) they have been excluded from this table as they were not employed in the analysis of the results, see *Materials and Methods*.

Table 2. Patch test results (I)

	Positive control	Negative control	Balm	0.3% SDS
mean erythema score (<i>n</i> =30)	1.9	0.5	3.1	1.4

Table 3. Patch test results (II)

	Balm	Balm minus perfume	0.3% SDS
mean erythema score (<i>n</i> =61)	2.4	0.3	1.2

Table 4. Patch test results

	Balm	Balm minus allyl esters	Balm minus perfume plus allyl esters	0.3% SDS
mean erythema score (<i>n</i> =30)	3.4	0.8	not done	1.6
mean erythema score (<i>n</i> =32)	2.6	0.5	2.8	1.7

Table 5. Elbow test results

Day of assessment	No. panellists with each reaction grade*					
	balm			balm minus allyl esters		
	0	1	2	0	1	2
1	10	14	0	10	14	0
3	8	16	0	8	16	0
7	7	17	0	7	17	0
9	7	17	0	7	16	1
11	4	18	0	4	17	1
15	5	18	0	6	17	0
17	6	16	0	6	16	0
22	6	15	2	6	15	2
24	9	13	0	10	12	0
total erythema score	148			148		

* No reaction higher than grade 2 erythema was seen. The panel size was 24; where the number of observations does not sum to this total indicates failure of panellists to attend at that scoring day (e.g., on day 11, only 22 subjects came for assessment).

the balmless perfume. In this panel, the response to both the SDS marker and the balm are a little lower than in the 1st panel. However, the response to the balm without the perfume was extremely low, confirming the suspicion about involvement of the perfume in the unexpected irritation.

Having demonstrated the involvement of the perfume, it was then investigated by fractionation. 3 balm preparations were made from each of which components representing approximately 1/3 of the perfume content had been removed and replaced with an solvent (diethyl phthalate). Patch test re-

Table 6. Half-face test results

Day of assessment	Number of panellist sites with each reaction grade*							
	balm				balm minus allyl esters			
	0	1	2	3	0	1	2	3
1	0	68	83	5	0	69	83	4
3	0	65	87	4	0	67	84	5
5	0	55	85	16	0	60	81	15
8	0	66	78	12	0	58	80	18
10	0	60	80	19	0	69	69	18
12	0	72	76	8	0	80	71	5
15	0	44	92	19	0	44	94	18
17	0	45	95	16	0	49	88	19
19	0	58	77	20	0	59	76	21
22	0	45	80	30	0	46	84	26
25	0	63	68	25	0	55	74	27

3 sites per panellist were assessed, for which the data have been summed.

sults indicated that the chemical responsible was in the fraction containing a large number of minor components (data not shown). Within this fraction, 2 allyl esters were noted to be at rather higher than normal levels, with a total final product concentration of 0.11%. The role of these esters in the irritation reaction was evaluated in a further patch test (Table 4). In this study, both the balm and SDS produced results similar to those in the previous tests, but removal of the esters alone almost completely eliminated the skin irritation potential of the fully formulated balm. This result was confirmed in a 2nd panel. As a cross check, the balm with no perfume, but with the allyl esters added, was shown to possess as much irritancy as the formulated product.

Elbow test

Although the balm had shown very clear evidence of skin irritancy under patch test conditions, the known acceptability of this general product type in the market place and our knowledge of the market acceptability of the perfume (in a different product type) prompted investigation of irritancy under more realistic exposure conditions. Using 6× daily repeated open exposure for 3 weeks, the skin response to the balm was compared with that to the balm lacking the allyl esters. The results (Table 5) show that irritation reactions of any significance (either acute or cumulative) did not arise, even to the fully formulated balm with the allyl esters (grades of 0 and 1 effectively represent normal skin). Although an imperfect measure, the simple addition of the observed erythema scores serves further to demonstrate the lack of discernible difference. Statistical evaluation using the sign test also confirmed that there were no significant differ-

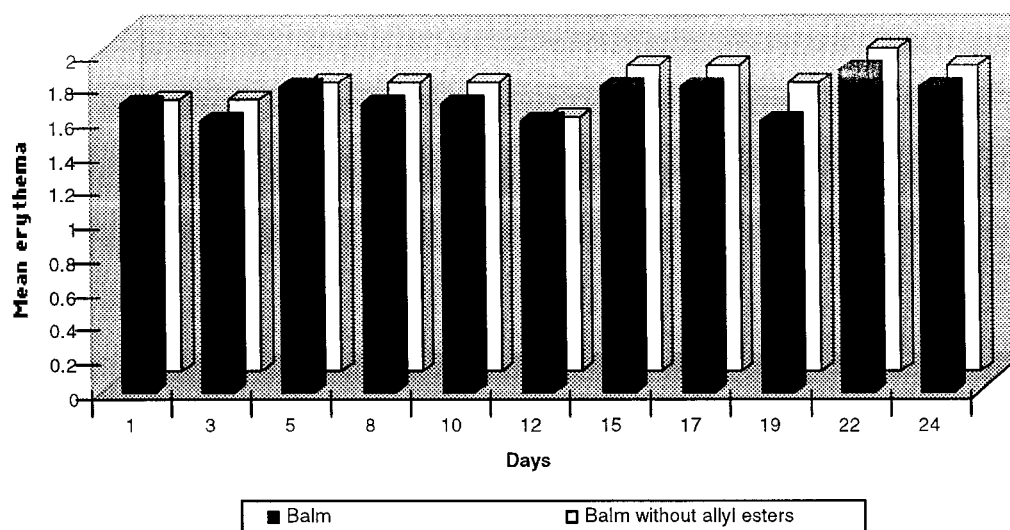


Fig. 1. Results of the half-face test.

ences between the response to balm and, balm minus the perfume, in the panellists.

Half-face test

The lack of reaction in the elbow test provided the necessary reassurance for progression to a half-face test. The results are given in Table 6 and show that 2× daily application of balm or, balm minus the allyl esters, for over 3 weeks failed to cause development of significant or differential skin irritation. This outcome is emphasised by examination of the average erythema scores (Fig. 1), which shows that repeated use of balm or balm without the 2 allyl esters, had essentially no effect on the average degree of erythema on the assessed facial skin sites (cheek, jawline and neck), the values ranging from 1.5 to 1.9 for both groups. A smaller separate group of subjects ($n=19$), who used no treatments at all on their faces, had mean erythema values of 1.5 ± 1 throughout the duration of the study. Of course, such average estimations may mask clear responses in a sensitive/reactive sub-set of the volunteer panel. Examining individual panellists' data reveals that whilst 3 were stopped from continuing to treat with the balm (with allyl esters), only 1 had to stop the treatment with the balm without allyl esters. There was a similar bias in the unsolicited comments from the panellists about sensory effects (data not shown). However, only 1 of these differences reached a level of statistical significance using the sign test (day 10, cheek area only, balm > balm minus perfume). In the context of the study and given that the increased reaction was only on part of the test site and did not progress,

the conclusion was that by itself, this observation did not carry much weight.

Discussion

Product safety evaluation forms a prerequisite for the marketing of cosmetic products. In the EU this is regulated by the EC Cosmetics Directive (1). 1 particular aspect of the evaluation is the assessment of skin irritation, strategies, which are well described (2, 6). A commonly used approach is to apply the product under a patch test, which can give a stringent assessment of acute skin irritation potential. The activity of the product is compared to that of known positive and negative controls, which may well consist of products known either to be well tolerated in the marketplace, or to have some history of marketplace problems.

Whilst this approach offers certain advantages, not least that of being relatively quick and cheap, it does not evaluate the irritancy of the product under realistic use conditions. In this work, we have shown that a product that gave what would have been regarded as totally unacceptable reactions in a patch test, failed in subsequent repeated open application studies to give any significant adverse responses. The chemicals responsible for the irritant effects in the patch test were demonstrated to be allyl esters in the fragrance. These materials were without effect when applied in an alcoholic base in a patch test (data not shown) and are widely used in cosmetic products without giving rise to skin irritation.

In conclusion, it is our view that, when human volunteer studies need to be carried out to establish the safety of products in respect of skin irri-

tation, properly designed studies which reflect anticipated exposure are most likely to give meaningful results.

Acknowledgements

The authors would like to thank Claire Talbot and Lisa Peters for expert technical assistance.

References

1. EC. Directive 93/35/EEC Amending for the 6th time Directive 76/768/EEC on the approximation of the laws of the member states relating to cosmetic products. *Official Journal of the European Communities* 1993; L151: 132.
2. Walker A P, Basketter D A, Baverel M, Diembeck W, Matthies W, Mougin D, Paye M, Rothlisburger R, Dupuis J. Test guideline for assessment of skin compatibility of cosmetic finished products in man. *Food and Chem Toxicol* 1996; 34: 551–560.
3. Patil S, Patrick E, Maibach H I. Predictive skin irritation tests in animals and humans. In: Marzulli F N, Maibach H I (eds): *Dermatotoxicology*, 5th edition. Washington: Hemisphere Publishing Corporation, 1996: 411–437.
4. Basketter D A, Reynolds F S, York M. Predictive testing in contact dermatitis: Irritant dermatitis. In: Goh C L, Koh D (eds): *Clinics in dermatology: contact dermatitis*. Amsterdam: Elsevier, 15: 637–444.
5. Basketter D A, Reynolds F S, Rowson M, Talbot C, Whittle E. Visual assessment of human skin irritation: a sensitive and reproducible tool. *Contact Dermatitis* 1997; 37: 218–220.
6. Jenkins H L, Adams M G. Progressive evaluation of skin irritancy of cosmetics using human volunteers. *International J of Cosmetic Science* 1989; 11: 141–149.

Address:

D. A. Basketter
SEAC Toxicology, Colworth
Unilever Research Sharnbrook
Bedford, MK44 1LQ
UK