

INVESTIGATIVE REPORT

Skin Cleansers: Three Test Protocols for the Assessment of Irritancy Ranking

JULIA SPOO, WALTER WIGGER-ALBERTI, UNDINE BERNDT, TOBIAS FISCHER and PETER ELSNER

Department of Dermatology, Friedrich Schiller University, Jena, Germany

In this article, the results of 3 studies on 2 hand cleansers (products A and B) are analysed. Three different test models (the patch test, the forearm wash test and the use test) are used to obtain information on the skin irritancy of these 2 products. Test reactivity was assessed by clinical scores and bioengineering methods such as corneometry for skin moisture, transepidermal water loss measurements for barrier function and chromametry for erythema. A correlating trend of product A being more aggressive than product B could be confirmed in all 3 studies and was statistically significant in the patch-testing series. Distinction of the results was dependent on the test protocol employed. Models for testing cleansing preparations should be chosen carefully, depending on the hypothesis to be evaluated. Key words: bioengineering methods; cleanser; patch test; wash test; use test.

(Accepted November 26, 2001.)

Acta Derm Venereol 2002; 82: 13–17.

J. Spoo, Department of Dermatology, Friedrich Schiller University, Erfurter Str. 35, D-07743 Jena, Germany. E-mail: spoo@derma.uni-jena.de

Different protocols for assessing the safety of skin cleansers have been proposed (1–5) and standardized for broad applications: it is possible to discriminate the level of irritancy between them by means of different test designs already established and published (6, 7). These include the one-time occlusive test, repeated occlusive test, repeated open test, wash test and use test. All these methods have specific advantages and disadvantages.

In the past, patch testings with soap chambers have been considered problematic (8) because simulation of in-home use is not adequately warranted and the test material itself is aggressive and irritating. However, large numbers of products can be tested within a short period of time (<1 week) and low numbers of subjects are required to achieve statistically significant results to discriminate the products tested.

Exaggerated wash tests are predictive of normal use tests (6) and, because of exaggerated washing, products can be tested in a short time. Low numbers of subjects are necessary for validated results. The methods (several variations of these tests are used) can be considered as

semi-aggressive. Use tests have the substantial disadvantage of requiring large numbers of volunteers using the test products over a considerable period of time for statistically relevant statements. Additionally, interindividual confounders at home influence the outcome of the test. On the other hand, they are considered the best means of emulating in-home-use situations, while being the least aggressive method for testing hand cleansers (6).

For detergents, discrepancies between the one-time patch test and the wash test regarding the ranking of irritancy have been found in the literature (9). Measurements of transepidermal water loss (TEWL) have resulted in a higher concordance than visual scoring among different exposure methods (10).

In this article, we compare testings of 2 hand-cleansing products (products A and B) by 3 different methods – the patch test, the use test and the forearm wash test. Data for the products are extracted from 3 studies conducted on hand-cleansing preparations. To our knowledge, this is the first report of identical substances being assessed by these 3 different study designs. Skin impairment was evaluated by measuring skin physiological parameters, such as hydration, TEWL, erythema, and by visual scoring. A critical interpretation of the results is attempted.

MATERIALS AND METHODS

Test products and their formulations

All tests were performed on healthy volunteers free of skin diseases and after obtaining informed consent in accordance with the Helsinki declarations. Product formulations tested were as follows: *Product A*. Aqua, Sodium Laureth Sulfate, Cocamidopropyl Betaine, PEG-200 Hydrogenated Glycerol Palmitate, Sodium Chloride, Glycol Distearate, Glycerol Laurate, Laureth-4 and Myreth-4, PEG-7 Glycerol Cocoate, Parfum, Citric Acid, Propylene Glycol, C.I. 45100, Methylchloroisothiazolinone, Methylisothiazolinone. *Product B*. Aqua, Sodium Laureth Sulfate, Disodium Cocoamphodiacetate, Sodium Laureth-11-Carboxylate, Cocamidopropyl Hydroxysultaine, Polysorbate 20, Citric Acid, Parfum, Aloe Barbadensis, Methylchloroisothiazolinone, Methylisothiazolinone, C.I. 19140, C.I. 15985, C.I. 28440, C.I. 14700, Calendula officinalis, Anthemis nobilis, Tilia Codata, Cenataurea Cyanus, Matricaria, Chamomilla, Hypericum Perforatum, Propylene Glycol.

Use test (Test I). Twenty-four volunteers, 12 per group, were asked to apply the test products at least 20 times a day during

normal hand-wash procedures for 2 weeks. Daily frequency of hand-washing was recorded in a diary. The participants were asked to abstain from using other cleansers for hand-washing during the 2 weeks of the study period. Additionally, standardized shower gels and hand creams were provided from the study sponsor in order to minimize external confounders.

Forearm wash test (Test II). This study consisted of 3 cleansers including products A and B being tested in 24 test volunteers. Each preparation was tested 16 times, each person receiving 2 of the products, products A and B thus being applied simultaneously in 8 subjects and tested in combination with another cleanser in 8 more subjects. The volar forearms of the panelists were moistened by water spray, and lather generated from a defined amount of soap was applied on a marked area in the middle of the forearm, as demonstrated earlier (8). The skin was consecutively washed for 1 min by gentle forward and backward movements of the investigator's gloved fingers. After 1 min the washing solution was absorbed by a paper tissue, the skin moisturized again and additional lather generated and applied for another minute. After 2 wash cycles the forearm was rinsed with clear water for 30 sec. The same procedure was repeated with the contralateral forearm.

Patch test (Test III). Forty volunteers agreed to having 50 μ l of 4 different cleansers, including products A and B (concentration 10%), applied in large Finn chambers[®] (diameter 12 mm) placed on their backs. The chambers remained in that position for 48 h.

Clinical and instrumental evaluation

All clinical and instrumental evaluations were conducted in a room with standardized environmental conditions (room temperature 22°C, humidity 45–55%) and after acclimatization of the subjects for 10 min in a quiet position.

Clinical evaluation. Visual scores were used for clinical evaluations as follows:

Test I: The skin status of the hands was evaluated by a dermatologist directly prior to the first hand-wash and 3 h after the final wash.

0: very slight erythema/glossy surface/no scaling/no edema/no fissures

1: slight erythema/roughness/scaling/edema/fissures

2: medium erythema/roughness/scaling/edema/fissures

3: severe erythema/scaling/edema/fissures with exudation

Halfgrade steps were allowed.

Test II: The forearms were evaluated concerning erythema by visual scoring of a dermatologist prior to the first wash and 3 h after the final wash on day 5. The score used was 0 = none to 5 = severe erythema.

Test III: Patch test results were evaluated according to the same clinical erythema score prior to Finn chamber application and 3 h after removal of the Finn chambers.

Instrumental evaluation. Instrumental evaluations were performed for assessment of hydration of the stratum corneum, TEWL and erythema. The measurements were conducted at the following points of time:

Test I: day 1 before the first wash and 3 h after the last wash on day 12

Test II: day 1 before the initial wash cycle and 3 h after the last wash cycle on day 5

Test III: day 1 before application of the Finn chambers and 3 h after removal of the Finn chambers on day 3.

Hydration of the stratum corneum was evaluated with the Corneometer CM 825 (Courage and Khazaka, Cologne, Germany). Each value represents the average of 3 individual measurements (11).

Duplicate measurements of the TEWL as indicator for barrier function were performed using the Tewameter TM 210 (Courage and Khazaka, Cologne, Germany) and conducted according to the guidelines described by the Standardization Group of the European Society of Contact Dermatitis (12). Means were calculated for statistical evaluation. Skin colour (degree of erythema) was assessed with the Chromameter CR-200 (Minolta, Osaka, Japan) according to the guidelines of the ESCD (13). This instrument measures the colour of the skin by 3 parameters: the luminance (L^* -value), the red-green scale (a^* -value) and the blue-yellow scale (b^* -value). The red coloration assessed by the a^* -value characterizes the formation of erythema as indicator of irritation. Each value represents the average of 3 individual measurements.

Statistics

Statistical analysis was conducted using SPSS for Windows computer software. Arithmetic means were calculated. Differences between means of all assessed parameters from first to last day of observation were checked for significance using the Wilcoxon test. Since data were extracted from studies in which more than 2 cleansing preparations were tested, statistics were re-evaluated for group-to-group analyses (differences between products A and B) where appropriate. Statistics for the forearm wash test were reconsidered using the two-sided adaptive test following Fisher's combination test. The Mann Whitney U test was applied in the use test, the Wilcoxon test for the patch test. Significance was given with a 5% probability of error ($p \leq 0.05$).

RESULTS

The Δ values of visual scores, skin moisture, erythema and TEWL are graphically displayed in Figs. 1–3. Determination of the visual scores showed a tendency of product A to be more skin impairing than product B as determined by all test methods. The reaction was most severe in the patch testing series, followed by the wash and use tests. Discrimination between the 2 products was most distinct in the patch test. Visual scores of

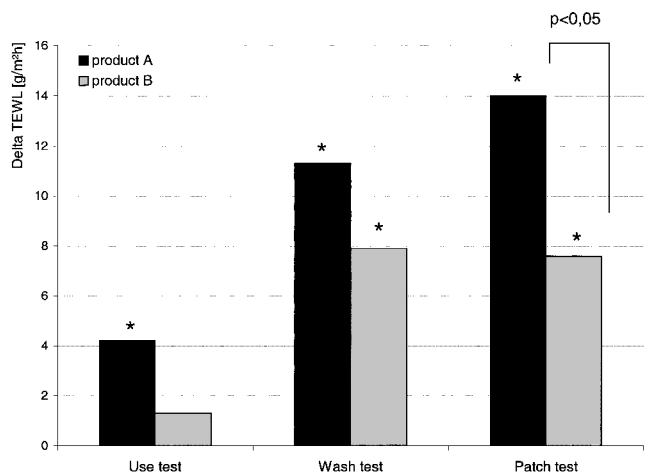


Fig. 1. Transepidermal water loss measurements of the 3 test methods compared (Δ -values (mean difference between measured values on the last and first day of observation): use test $n = 12$, d12-d1; wash test $n = 16$, d5-d1; patch test $n = 40$, d3-d1). *Significant change from first to last day of observation in the original measurement series ($p \leq 0.05$).

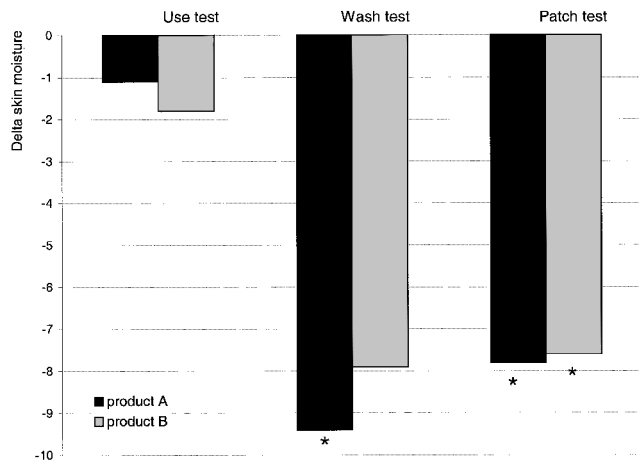


Fig. 2. Stratum corneum hydration in the 3 test methods compared (Δ -values (mean difference between measured values on the last and first day of observation): use test $n=12$, d12-d1; wash test $n=16$, d5-d1; patch test $n=40$, d3-d1). *Significant change from first to last day of observation in the original measurement series ($p \leq 0.05$).

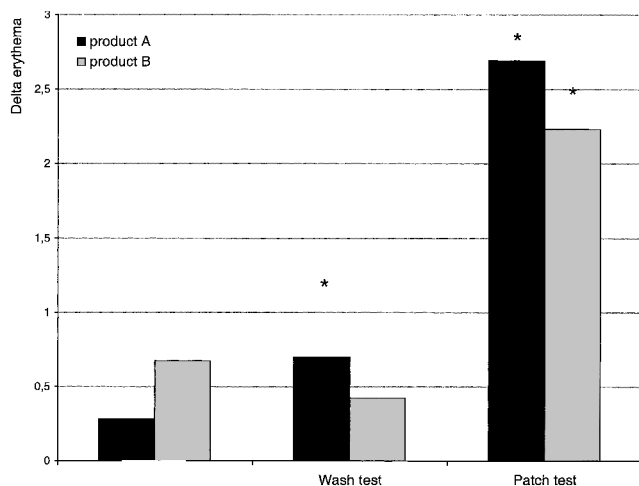


Fig. 3. Erythema measurements in the 3 test methods compared (Δ -values (mean difference between measured values on the last and first day of observation): use test $n=12$, d12-d1; wash test $n=16$, d5-d1; patch test $n=40$, d3-d1). *Significant change from first to last day of observation in the original measurement series ($p \leq 0.05$).

both products in this series changed significantly from day 0 to day 2.

Evaluated by TEWL measurements, all test methods used proved a higher aggressiveness of product A than product B as determined on the last day of measurement. In the use test, absolute changes were least distinct and on a lower level than in the other tests. Impairment of the barrier function induced by product A was significant in all tests performed. The wash and the patch test also showed significant Δ values regarding product B.

Skin hydration as determined by corneometry was most affected by the wash test procedure. Product A showed a tendency to be more aggressive than product B by lowering stratum corneum moisture significantly from day 1 to day 5. In patch testing, skin hydration

was equally affected by the 2 products and was significant. Skin dehydration was only minimal in the use test and changes from day 1 to day 5 were not significant. It was affected slightly more by product B than by product A. Erythema induction was on a comparable and low level in the wash and the use test and significantly increased by the patch test procedure. Irritation by product A compared with product B was more distinct in the patch test and the wash test. In the use test, product B showed a tendency toward greater irritation.

Comparing the cleansers with each other, intergroup analysis proved significant in the patch testing series for TEWL measurement ($p \leq 0.05$) only.

DISCUSSION

Dermatological test methods are important for quality assurance and compatibility assessment of cosmetic products. New guidelines have been proposed for cosmeceutical testings by the European Cosmetic Toiletry and Perfumery Association (COLIPA) for better standardization of research performed in humans in relation to ethical as well as scientific endpoints (7). It was the aim of this study to compare the results of 3 different test regimens testing 2 hand-cleansing products in order to evaluate the degree of corresponding statements.

It is obvious that comparisons, and especially statistical evaluation, of different test protocols should, ideally, be identical with respect to subject recruitment and study design. Therefore, test procedures as such were not subject to statistical evaluation in this study. Nevertheless, the comparison of trends is considered a worthwhile contribution to the discussion of test procedures for cleansing preparations.

With our data, we were able to show that, despite different test methods being used and, therefore, different endpoints being assessed, the trend of product A to be more aggressive than product B could be confirmed by the wash test and the patch test, and, to some degree, the use test. Because of the different clinical scores used, the results for visual scoring should be considered questionable as regards comparison of irritation severity and be regarded as a trend only. However, when evaluated visually, no discrepancy in irritancy rank-order between the 3 test models was observed. Clinical discrimination of both products as well as erythema induction were most obvious in the patch test, which reflects the recommendations for patch testings of cleansing preparations: a good way of rapidly obtaining information on the degree of skin impairment induced by cleansing preparations (6, 7). In our data, intergroup analysis showed significant differences between products A and B concerning barrier function. The evaluation of skin hydration and barrier function, on the other hand, is less relevant in patch testing, since the test procedure as such – application of irritating substances in Finn cham-

bers for 48 h – is aggressive and can by no means represent ordinary use of these substances. Therefore, measurements obtained by evaporimetry and corneometry during patch testings of cleansing preparations might be considered unnecessary.

The use test corresponded with the other tests concerning clinical scoring and TEWL measurement, whereas results for the parameters skin moisture and erythema were not in accordance. For all parameters evaluated, the Δ -values determined in the use test were least distinct and on a lower level than in the other test methods described. The only significant change reported in the use test was in product A for TEWL. This method should be employed for evaluation of long-term-use situations, since even minimal impairment of the barrier function can be assessed, and the method has reliable discriminating abilities from a statistical point of view – adequate subject numbers being guaranteed. Referring to the data analysed, significant statistical results in intergroup analysis between the 2 cleansers were not obtained in either the use or the wash test. The reason for this may be that there is little difference in the products themselves, considering the study designs employed; however, it may also be due to the small numbers of subjects recruited for both the use test ($n = 12$) and the wash test ($n = 16$).

Product A proved more aggressive than product B in the laboratory-controlled environment of the wash test, with skin hydration being significantly reduced by product A. TEWL was also significantly influenced by the aggressive wash procedure, while erythema induction should be considered as a parameter of minor significance to the wash test.

A wide range of test methods for assessing cleansing products is at our disposal. Before entering into studies, test models should be selected strictly with regard to the hypothesis to be evaluated. When testing cleansers containing grits (14), for example, wash test protocols or use tests should be chosen for evaluation of skin barrier impairment, whereas the patch test, due to application under occlusion for 48 h which does not imply any washing procedure, cannot answer any particle-associated question.

The patch test should be chosen for orientation about the degree of irritancy of different products; a number of <30 subjects has been advised for testing and in our study 40 volunteers participated. A significant difference between products A and B was observed employing this test method.

For evaluation of mildness and dryness potential over a wide range of products, exaggerated wash tests should be conducted. Group sizes of 20–40 subjects have been suggested by the COLIPA study (7) and Murahata & Nicoll (6).

Imitation of everyday-use situations is warranted by the use test; however, because of the effects of external

confounders beyond the control of the examiner over a considerable period of time, this test requires large numbers of subjects for statistically significant discriminating results (6). External confounders should be kept to a minimum by distributing standardized home-use products for skin care and cleansing to the volunteers participating in these kinds of studies.

CONCLUSIONS

The wash test, the patch test and to some degree the use test showed good correlation concerning skin impairment of the 2 cleansing products tested, although each test protocol is recommended for the assessment of different endpoints in cosmeceutical testings. A corresponding trend could be confirmed for all methods used, although statistical significance between the tested cleansers could be proved in the patch-testing series only. Conduction of clinical studies for the evaluation of cleansers in accordance with the recommendations of the COLIPA is emphasized for obtaining valid results of statistical and biological significance.

For patch testing, visual scoring remains the “gold standard” in the prediction of irritancy potential. The benefit of TEWL proved high, especially in the situation closest to real life – the use test. However, there is still a lack of knowledge about the concordance of different exposure models with field studies that should be compared for better recommendations as to which test model should be performed.

ACKNOWLEDGEMENT

The study was supported by Kimberly-Clark Europe, Flintshire, UK.

REFERENCES

1. English J, Ratcliffe J, Williams H. Irritancy of industrial hand cleansers tested by repeated open application on human skin. *Contact Dermatitis* 1999; 40: 84–88.
2. Gehring W, Gloor M, Kleesz P. Predictive washing test for evaluation of individual eczema risk. *Contact Dermatitis* 1998; 39: 8–13.
3. Grunewald A, Gloor M, Gehring W, Kleesz P. Damage to the skin by repetitive washing. *Contact Dermatitis* 1995; 32: 225–232.
4. Hannuksela A, Hannuksela M. Irritant effects of a detergent in wash and chamber tests. *Contact Dermatitis* 1995; 32: 163–166.
5. Hannuksela A, Hannuksela M. Irritant effects of a detergent in wash, chamber and repeated open application tests. *Contact Dermatitis* 1996; 34: 134–137.
6. Murahata R, Nicoll G. Mildness testing for personal washing products. In: Aust L, ed. *Cosmetic claims substantiation*. New York: Marcel Dekker, Inc., 1997: 153–169.
7. Basketter D, Baverel M, Diembeck W. Cosmetic product test guidelines for the assessment of human skin compatibility. In: *COLIPA Symposium on Alternatives to Animal Testing*, Brussels, 1995.
8. Nicoll G, Murahata R, Grove G, Barrows J, Sharko P.

- The relative sensitivity of two arm-wash test methods for evaluating the mildness of personal washing products. *J Soc Cosmet Chem* 1995; 46: 129–140.
9. Imokawa G, Mishima Y. Cumulative effect of surfactants on cutaneous horny layers: adsorption onto human keratin layers *in vivo*. *Contact Dermatitis* 1979; 5: 357–366.
 10. Tupker R, Bunte E, Fidler V, Wiechers I, Coenraads P. Irritancy ranking of anionic detergents using one-time occlusive, repeated occlusive and repeated open tests. *Contact Dermatitis* 1999; 40: 316–322.
 11. Courage W. Hardware and measuring principle: Corneometer. In: Elsner P, Berardesca E, Maibach H, eds. *Bioengineering of the skin: water and the stratum corneum*. Boca Raton: CRC Press, 1994: 171–175.
 12. Pinnagoda J, Tupker R, Agner T, Serup J. Guidelines for transepidermal water loss (TEWL) measurement. A report from the Standardization Group of the European Society of Contact Dermatitis. *Contact Dermatitis* 1990; 22: 146–178.
 13. Fullerton A, Fischer T, Lahti A, Wilhelm KP, Takiwaki H, Serup J. Guidelines for measurement of skin colour and erythema. A report from the Standardization Group of the European Society of Contact Dermatitis. *Contact Dermatitis* 1996; 35: 1–10.
 14. Wigger-Alberti W, Fischer T, Greif C, Maddern P, Elsner P. Effects of various grit-containing cleansers on skin barrier function. *Contact Dermatitis* 1999; 41: 136–140.