

Review Article

Validity and ethics of the human 4-h patch test as an alternative method to assess acute skin irritation potential

MICHAEL K. ROBINSON¹, JOHN P. MCFADDEN² AND DAVID A. BASKETTER³

¹The Procter & Gamble Co., Miami Valley Laboratories, Cincinnati, Ohio, USA

²St John's Institute of Dermatology, St Thomas' Hospital, London, UK

³SEAC Toxicology Unit, Unilever Research, Colworth House, Sharnbrook, UK

For more than 50 years, the Draize rabbit skin irritation test has reigned supreme as the regulatory method of choice for the identification of skin irritant chemicals. To date no in vitro alternative test has been validated as an adequate replacement. However, one potential option, to test the endpoint of concern (skin irritation) in the species of concern (man) has been overlooked. The advent of predictive in vitro tools for the identification of substances corrosive to the skin has opened up the practical possibility of carrying out safe and ethical studies on small panels of humans. The human 4-h patch test has been developed to meet the needs of identifying chemical skin irritation potential, providing data which is inherently superior to that given by a surrogate model, such as the rabbit. This paper reviews in detail the present state of the human 4-h patch test, highlighting its advantages and noting its utility as the 'gold standard' on which to build future in vitro models.

Key words: skin irritation; 4-h patch test; alternatives to animal testing; EU legislation; R38.
© Munksgaard, 2001.

Accepted for publication 23 January 2001

The Need for Alternative Test Methods

The manufacture, transport and marketing of chemicals and products requires prior toxicological evaluation of the potential skin corrosivity and skin irritation that might result from intended or accidental exposures. This is done both for the protection of those exposed as well as to comply with regulations (1). Traditionally, the method of choice for evaluating the corrosive or irritant properties of chemicals has been the Draize rabbit skin test (2). This procedure, developed over 50 years ago (3), has become a world standard, although it has been modified to varying extents by regulatory authorities in different countries (4). Variations of the procedure form the basis for classification of skin corrosion and irritation hazard to man (5). All are based on the severity, speed of onset, and persistence and reversibility of skin reactions after

occluded/semi-occluded topical application of test material to the shaved skin of albino rabbits. Labeling or transport restrictions imposed on chemicals or finished products are based on the results obtained from this test (6).

For many years, the Draize method has been criticized for a lack of reliability, particularly when directly comparing skin testing results in the rabbit and in man (7–9). More recently, pressure has come from animal welfare organizations to eliminate its use altogether. In Europe, legislation is pending that would restrict marketing of products, the ingredients of which have been tested on animals (10). This produces a dilemma for the toxicologist faced with the need to protect workers and consumers, comply with regulations, and, at the same time, reduce or eliminate testing of chemicals and products on animals.

What has been created is an extraordinary need

for alternative test methods to replace the traditional animal protocols. This need has driven 2 major initiatives. The first of these initiatives has been the development of in vitro test methods, both cellular and non-cellular. The second initiative has been the development of predictive human skin irritation test methods. To date, the in vitro test methods have been primarily directed toward the evaluation of skin corrosive properties of chemicals. Several methods have been progressed from development (11–13) through increasingly rigorous pre-validation (14) and validation (15) testing programs. In vitro testing procedures for skin irritation have lagged behind (16). Recently, several different cellular systems have been described as having utility for acute and/or cumulative irritation testing (17–21). A pre-validation program is underway (22).

During the past several years, considerable effort has also been invested in the development and evaluation of an ethical and robust human test procedure for assessment of the acute skin irritation potential of chemicals (23–28). As a human test method, this 4-h patch test requires no species or system extrapolation of data for prediction of skin irritation potential. Subsequent to its early development (23–25), this test method has undergone initial inter-laboratory evaluation (26, 28, 29) and has been applied to the study of different aspects of the acute skin irritation response (27, 30–33).

Limitations to the Draize test

The Draize rabbit skin irritation test procedure has been around for over 50 years and is used, in several variations, for evaluation of the corrosive and irritant properties of chemicals. The procedure involves direct application of a test material to the skin of albino rabbits under an occlusive or semi-occlusive dressing and the subsequent assessment of induced skin damage (necrotic changes, erythema, edema). Various modifications of this basic method are currently required by regulatory authorities worldwide and form the basis for classification of skin corrosion or irritation hazard to man (5). The modifications relate to the number of animals tested, the duration of test material exposure, the application on intact versus abraded skin, and the number and frequency of skin evaluations (4). Depending on the nature and severity of the skin reactions, their speed of onset, and their persistence or reversibility, classifications are made regarding corrosion or irritation hazard. Labeling requirements and/or transport restrictions imposed upon chemicals and products are based upon the length of time required to develop a proscribed level of skin response.

For classification as corrosive, skin lesions in the rabbit must penetrate into the dermis, must occur quickly (<4h of exposure) and be irreversible. The severity of the classification is determined by the time of onset of the corrosive response. Generally, a positive response in a single rabbit is sufficient to trigger a corrosive classification. Skin irritation testing involves longer term exposures (generally 24 h) over intact and/or pre-abraded skin. Skin sites are graded at various time points (generally 24, 48, and 72 h) after patch removal for erythema and edema. Both endpoints are evaluated using a 4 point scale of increasing severity. The erythema and edema grades are averaged across all animals, application sites, and time points and then added to derive a primary irritation index (PII) for the test material. Extended grading time points (generally 7 to 14 days) are included to assess recovery of the skin lesions. Skin irritation classifications of mild, moderate, and severe are ascribed to PII values of <2, 2–5, and >5, respectively. However, different regulatory authorities apply various criteria to determining the classification of a test material. For example, whereas the United States Environmental Protection Agency uses the numerical PII ranges listed above, the European Union classifies into a single category (R38: Irritating to skin) on the basis of a PII value ≥ 2 .

The Draize test has come under increasing criticism from 2 main sources; animal welfare concerns and predictive validity for the human response. Animal welfare groups have long targeted the Draize test procedures as inhumane and unnecessary. This has driven the passage (although not yet the implementation) in Europe of legislation that would ban the sale of products, the ingredients of which have been tested on animals (10). This presents a challenge to the toxicologist attempting to protect worker and consumer safety, comply with regulations, and reduce or eliminate testing on animals. In addition, it is important to note that the reproducibility and relevance of the rabbit test results to human experience have been seriously questioned in the literature (7–9). In particular, the Draize test results are a poor predictor of human skin irritation hazard (34) (Table 1), not surprisingly given the use of a very small number of animals to try and predict skin irritation hazard for a diversely reactive human population (31).

Development and Limitations of In Vitro Test Methods

The concerns over the limitations of the Draize test method have clearly served as impetus for the development and subsequent validation efforts for alternative in vitro skin toxicology methods. In the

Table 1. Correspondence between European classifications of 65 chemicals and human skin irritation potential

Existing EC Classification ^a	Human 4-h patch test classification ^b	
	R38	Non-Classified
R34	3^c	1^c
R38	7	23 ^c
non-classified	2 ^d	29

^a Based upon EC, Supplier, and Trade Association Data.

^b Based on results of human 4-h patch test (see Table 2 for details).

^c Bold values indicate over classification (EC) versus human 4-h patch test.

^d Italicized value indicates under classification (EC) versus human 4-h patch test.

area of in vitro skin corrosion testing, several test methods have been described over the past few years (11, 13, 35) and some received initial regulatory exemptions (11, 13). A recently completed validation study sponsored by the European Center for the Validation of Alternative Methods (ECVAM) identified 2 in vitro methods, transcutaneous electrical resistance and cytotoxicity in the skin equivalent construct, EpiskinTM, as valid alternative methods for skin corrosion (15). A second skin construct, EpiDermTM, has only recently completed successful 'catch-up' validation (36, 37), and has been endorsed by ECVAM as an alternative corrosivity test (38). Also, a non-cellular corrosion assay, CorrositexTM, was cited by the US Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM) as equivalent to the Draize test for predicting corrosivity and non-corrosivity for specified chemical classes (acids and bases) (39). In vitro skin irritation test methods are not as far developed in terms of the validation process (16), although a number of individual research reports have been published (17–21, 40, 41). Efforts are underway to identify potential in vitro acute skin irritation test methods and evaluate them through rigorous pre-validation and validation studies (22).

It is the validation process itself that highlights the limitations inherent to in vitro test methods. The methods described above have all been developed upon a sound scientific basis. The validation process then requires that, for each method, a prediction model be developed to allow for categorization of the test results and correlation of the predicted results (across many chemicals) to a previously established toxicity profile of each chemical in the replaced animal test or in man. Through this validation process, one determines whether the in vitro tests are, indeed, predictive. Of course, if one is correlating in vitro test data to animal test data that are inherently limited in their own predictive potential (see above, and Table 1),

it might call into question whether 'validated' in vitro tests are, in fact, valid. On the contrary, the human 4-h patch test provides a useful clinical correlate for in vitro test method development (16).

Ethical Human Testing as a Valid Alternative

A skin irritation response produced in a human being upon exposure to a chemical or other test material is, by definition, a valid response. It is, after all, a response produced in the relevant target organ in the appropriate species. There is no need to extrapolate from an artificial system (e.g., skin equivalent cultures or CorrositexTM) or another species. In using human subjects to develop an alternative test system, there are 2 critical factors. The first is reproducibility of the method. Given the knowledge that humans are a diverse, out-bred population and can respond quite variably to different stimuli (31, 42), it is critical that any human test method allow for inter-individual variation in response and yet demonstrate good inter-laboratory reproducibility. The prediction method embodied in the test method must, therefore, be sufficiently robust to compensate for inter-subject variations in response.

The 2nd critical factor is ethics. Human testing must always be conducted to the highest ethical standards. This requires adherence to strict testing guidelines (described below) relative to protocol review, informed consent, and overall protection of the wellbeing of the test subjects. This is particularly important in human toxicological testing where, typically, the test subjects derive no direct health benefit from their participation.

The Ethics of Human Skin Testing

The fundamental precept in any clinical research activity is the protection of the human test subjects. This is clearly articulated in the Helsinki Declaration of 1964 (43). Much of the guidance regarding the ethics of human studies relates to the testing of drugs and medical devices in which (in many cases) a medical benefit may be obtained from participation. However, safety testing of drugs on healthy volunteers (phase I testing) and skin safety testing of chemicals and products provide no health benefit. The only benefit to the participants is the compensation they receive. In light of this, certain principles must be followed. First, adequate safety of the test material must be assured before any human exposure is allowed. A thorough and independent risk assessment is required to assure no significant toxicities via the intended route of exposure. If necessary, an independent ethical review panel must approve the proto-

col and informed consent. Second, there must be no coercion for subject participation and enrolled subjects must be free to stop participation at any time without penalty, other than financial compensation. Also, the compensation provided should not be so great as to be a possible inducement for participation; rather, it should simply cover out of pocket costs and inconvenience. Third, the participants must be adequately informed of any and all possible adverse reactions (such as skin irritation, and particularly in pigmented skins the possibility of post-inflammatory changes in pigmentation) and must be guaranteed free medical intervention when necessary to treat such reactions. They must read, understand, and sign an informed consent form prior to entry into a study.

Skin safety testing, as typically conducted to evaluate chemicals and products, is inherently low in risk and is ethically acceptable when conducted to high standards (44). Skin irritation tests are designed to produce some degree of reaction that can be evaluated in a direct or comparative way. However, these reactions are almost always mild or moderate in severity, are limited to small sites of chemical or product exposure, and quickly recover. In practice, the 4-h human patch test has been shown to be inherently low in risk, a fact well documented by an extensive published literature.

Review of the 4-h Human Patch Test

The human 4-h patch test was conceived as part of a strategy to replace the use of animals for the identification of skin irritation potential (23, 24). The procedure was, in many ways, designed to be similar to the rabbit Draize test methodology it was seeking to replace. The measured endpoint (degree of visual irritation) was appropriate given that it is the same as that used in the animal test procedure and a very common method in dermatology testing.

Patch test method

The patch test method involves application of 0.2 ml of liquid materials (or 0.2 gm for solid materials) onto a 25 mm Hill Top Chamber (Hill Top Research, Cincinnati, OH) containing a webril pad. The patch is placed against the upper outer arm of approximately 30 human test subjects for periods of exposure of 15 or 30 min through 1, 2, 3, and 4 h. The shorter exposure durations (<1h) can be omitted if the investigator is satisfied that excessive skin reactions will not occur by going directly to a 1-h exposure. The patches are then removed, the skin sites very gently wiped with wet gauze to remove excess test material, and the sites

graded at 24, 48, and 72 h after patch removal. Irritation responses are measured using a simple scale (0, +, ++, +++) of increasing severity (23). Any subject presenting with a '+' or greater response to a particular test substance is considered to have shown a positive irritation reaction and further treatment with that particular test material is suspended. In evaluating the results of the test, what is measured is the number of subjects who had a positive irritant reaction after an exposure of up to 4 h. It is the incidence of responses, not the severity, that is the basis for evaluation. The interpretation of results in terms of EC classification of test materials is by a simple statistical comparison of the data with a concurrently tested positive control chemical. The statistical test is Fisher's exact test (45). The chemical selected as the positive control is 20% sodium dodecyl sulfate (see below for details on selection of the positive control chemical). Any test material showing a statistically lower incidence of positive responders versus the control would be non-classified. Test materials showing statistically greater incidences of response (or no statistically significant difference versus the control) would be classified as 'irritating to skin' (R38).

In its simplest usage, this protocol has been employed to evaluate the skin irritation potential of a range of substances, of varying irritation potentials, for which there was evidence of a current EC classification (R38 or non-classified). A detailed listing has been published (34). Whereas some substances have been tested diluted to demonstrate the value of the method, the primary aim is the testing of undiluted substances. Where the human results differ (up or down) from existing classifications based upon Draize test results ((34) and Table 1), the human result must obviously be the correct one.

Use of 20% SDS as a positive control for chemical irritation classification

The purpose of the 4-h patch test is not to make predictions about individual or group susceptibility to skin irritation. Rather, the test was developed simply to determine the acute irritation potential of a test material, using a test population of sufficient size to demonstrate this irritation potential versus a suitable benchmark. The relationship between the human response to any chemical and appropriate chemical classification is shown in Fig. 1. While it is possible that a very poorly responsive panel of subjects might be off-scale just as much as a highly reactive one, it has been the prevailing experience that a group of approximately 30 test subjects gives a reasonable assessment of

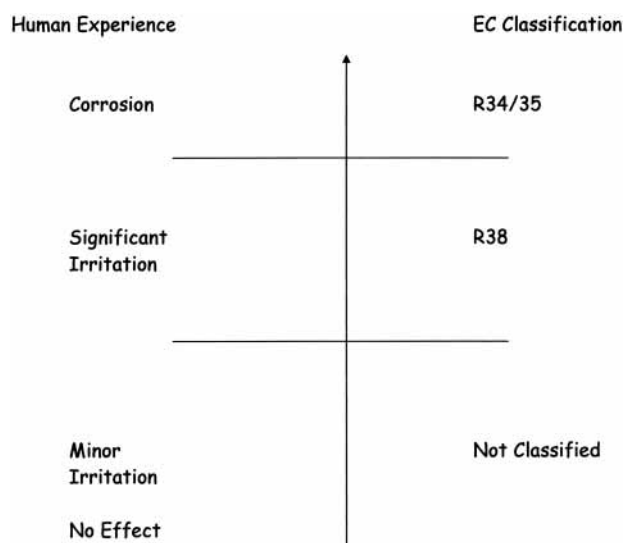


Fig. 1. Classification of skin irritation/corrosion potential.

skin irritation potential in relation to a well-selected positive control. It is the use of the benchmark control as a relational tool that helps provide the consistency in chemical irritancy classification.

Key to the successful development of the 4-h patch test was selection of 20% SDS as the positive control. The human skin response to irritants shows wide inter-individual variation (31). This can be exacerbated by differing environmental conditions. Hence, it was deemed important to 'calibrate' each study population by their collective response to a standard irritant. SDS was selected as the irritant control in part because it is the most widely used skin irritant in clinical investigations (46). Also, it is readily available, inexpensive, stable under normal clinical testing situations (47), and free from other adverse effects.

By including a routine positive control, there was also an opportunity to use it to perform another function, that of a benchmark. Skin irritation is not an absolute phenomenon. Any substance can produce skin irritation, it is strictly a matter of dose and the nature and extent of exposure. For this reason, skin irritation responses in humans are almost always comparative. Thus, SDS was chosen not only to calibrate the degree of response within a test, but to serve as point of comparison for acute irritation classification. The latter point is based upon the fact that SDS is classified as a skin irritant (R38) according to EC criteria (5). The concentration of 20% SDS was selected as the benchmark rather than undiluted material. This concentration was chosen as the threshold or minimum level of SDS (in solution or in a preparation) classified as R38. It was reasonable to conclude that substances which, when

tested undiluted, elicit statistically less skin irritation than this threshold concentration of SDS should not be classified as R38, whereas those substances producing a statistically similar or greater irritation should be classified as R38.

Inter-laboratory evaluation of the 4-h patch test

Before most new test methods can be accepted as suitable for accurately defining a toxic response, they must be 'validated'. Most new methods for skin toxicology are based upon the replacement of an animal testing procedure with an alternative method (e.g., in vitro test) for which there is a sound scientific rationale for the method, but little data to support its predictive value. Hence, validation generally requires development of a prediction model to categorize results and the accurate correlation of these predicted results (across numerous chemicals) to the established toxicity profile of each chemical in the replaced animal test or in man. This helps to establish the relevance of the method (48). A final step is a determination of the consistency of the method across independent laboratories. Only when all these requirements are satisfied, will a method be proposed as a 'valid' alternative (15). Having said this, it is critical to emphasize that the relevance of the human patch

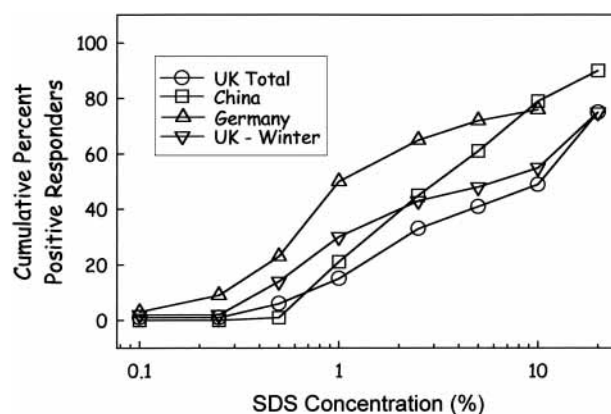


Fig. 2. The human 4-h patch test was conducted under a common protocol at independent testing laboratories in the UK (Unilever Environmental Safety Laboratory, Sharnbrook; St. John's Institute of Dermatology, London), Germany (*proDerm* Institute for Applied Dermatological Research, Hamburg), and China (Shanghai Institute of Dermatology, Shanghai). Subjects, ranging in age from 17–79, were patched with multiple concentrations (1–5% or 1–10%) of SDS under separate patches for 4 h. The patches were removed and the skin sites graded at 24, 48, and 72 h after patch removal. Those subjects not reacting to one or more SDS concentrations were similarly treated with the higher SDS concentration (10% or 20%) at naive skin sites. The graph displays the % positive responders at each SDS concentration for each test population. The German study did not include an exposure to 20% SDS. Figure produced from published data (29)

test method is already inherent in its evaluation of the appropriate target organ, toxic endpoint, and species (see above). The only critical questions relating to its ultimate acceptance are (1) the ability of different laboratories to obtain comparable results and (2) ethical considerations.

To date, there have been 3 major investigations of the inter-laboratory reproducibility of the 4-h patch test. The first of these was an international dose-response study with SDS (29). Independent laboratories in the UK, Germany, and China tested SDS concentrations ranging from 0.1% to 20%. A graphical display of the data is shown in Fig. 2. While there was clearly some variation seen in the dose-response profiles, this was predominantly observed at lower test concentrations. The responses observed with 10%–20% SDS were simi-

lar. In addition, the good intra-laboratory consistency seen across studies with repeat testing of 20% SDS is also relevant to the robustness of the method (27).

The second inter-laboratory comparison study examined multiple chemicals in comparison to 20% SDS to see if the relative classification patterns obtained across laboratories would be consistent. A total of 6 laboratories in the UK, Germany, China, and the US were involved in this comparison testing (26). The results are shown in Table 2.

There was some variation in test procedures employed by the different laboratories and this likely contributed to some variation in results obtained. These are detailed in the original publication (26). However, with minor exceptions, none of the dif-

Table 2. Interlaboratory comparison of chemical irritation study results^a

Test Material	EC Classification ^b	Testing laboratory	Dates of testing	No. pos./total reactions (%)	No. pos./total SDS reactions (%)	HPT classification ^c
acetic acid (10%)	R38	Unilever	Mar-Apr 1994	2/32 (6)	23/33 (70)	NC
		<i>ProDerm</i>	Nov-Dec 1994	4/31 (13)	22/31 (71)	NC
		P&G	Sep-Oct 1995	9/33 (27)	32/34 (94)	NC
benzalkonium chloride (7.5%)	R38	Unilever	Jul-Aug 1994	4/27 (15)	8/27 (28)	R38
		<i>ProDerm</i>	Feb-Mar 1995	15/29 (52)	24/29 (83)	NC
cocotrimethyl ammonium chloride	R38	IRI	Sep-Oct 1994	8/32 (25)	24/32 (75)	NC
		Unilever	Oct 1995	12/57 (21)	26/58 (45)	NC
decanol	R38	<i>ProDerm</i>	Aug-Sep 1995	11/31 (35)	17/32 (53)	R38
		<i>proDerm</i>	Jan 1996	0/29 (0)	20/29 (69)	NC
		Unilever	Jan-Feb 1996	5/28 (18)	23/30 (70)	NC
		P&G	Jan 1996	3/27 (11)	25/28 (89)	NC
hexanol	NC	Unilever	Jan-Feb 1996	8/28 (29)	23/30 (77)	NC
		P&G	Jan 1996	2/27 (7)	25/28 (89)	NC
octanol	NC	Unilever	Jan-Feb 1996	5/28 (18)	23/30 (77)	NC
		P&G	Jan 1996	4/27 (15)	25/28 (89)	NC
sodium dodecyl sulfate (1%)	NC	Shanghai	Feb-Mar 1995	21/100 (21)	90/98 (92)	NC
		St. John's	Mar 94-Sep 95	15/100 (15)	76/100 (76)	NC
		P&G	Feb-Mar 1995	16/31 (52)	30/31 (97)	NC
sodium dodecyl sulfate (10%)	NC	Unilever	Aug 94-Feb 95	42/64 (66)	43/63 (68)	R38
		Shanghai	Feb-Mar 1995	79/100 (79)	90/98 (92)	NC
		St. John's	Mar 94-Sep 95	53/100 (53)	76/100 (76)	NC
		P&G	Aug-Sep 1995	29/31 (84)	30/31 (97)	R38
sodium hydroxide (0.5%)	R38	Unilever	Mar-Apr 1994	20/33 (61) ^d	4/33 (12) ^d	R38
		<i>proDerm</i>	Nov-Dec 1994	13/31 (42) ^d	1/31 (3) ^d	R38
		P&G	Sep-Oct 1995	21/34 (62) ^d	7/34 (21) ^d	R38
tween 80	NC	Unilever	Jul-Aug 1994	1/24 (4)	8/27 (28)	NC
		<i>proDerm</i>	Feb-Mar 1995	1/29 (3)	24/29 (83)	NC

^a Test data presented are from independent tests at 6 laboratories. Results are presented as the cumulative proportion of subjects giving positive skin reactions after 4 h of exposure, with the percentage positive responders in parentheses

^b EC classification is from EC, supplier, and trade association data

^c HPT classification is from the human patch test; The substance is not classified (NC) if the proportion of the positive responders is significantly less than the proportion of the positive responders to the positive control (20% SDS), using the Fisher's exact test. The substance is classified (R38) if the proportion of the positive responders is not significantly different than (or is significantly greater than) the positive control.

^d Results were based on only 1 h of exposure.

Reprinted from (26).

ferences in test method or time (season) of testing was sufficient to alter the irritation classifications for the different chemicals tested. Those few situations resulting in different chemical classifications were attributed to unusually low responses to 20% SDS or to unusual seasonal effects. To address the control SDS response issue, the method now stipulates that positive responses to the 20% SDS control be observed in at least 1/3 of study subjects in order for the test to be valid for irritation classification (24). The seasonal effects question has been resolved through additional studies (27, 33).

Because of the subtle differences in test methods used in the 1997 inter-laboratory comparison study, a recent 2-laboratory study was conducted to compare selected irritation responses and classifications under as controlled a set of conditions as possible (28). A graphical display of the response incidences for the 2 laboratories is shown in Fig. 3 (A, B). This study resulted in the same relative chemical irritation response patterns and classifications for each of the 3 test substances. Recently, a curve-fitting procedure has been applied to the analysis of test results to derive a TR_{50} value, the time of exposure required for 50% of the test subjects to respond to any given test material (27). Though not used for irritation classification purposes, this analysis does add value in comparative studies as shown by the nearly identical TR_{50} values between the 2 laboratories (Fig. 3 C, D). These various intra- and inter-laboratory investigations underscore the robustness of the method.

Use of the 4-h Patch Test to Investigate Response Patterns of Acute Skin Irritation

Since its development 6 years ago, the human 4-h patch test has proven to be quite valuable in dissecting certain elements of the acute skin irritation response in man. Much of this work has been directed toward investigating the influence of extrinsic and intrinsic factors on susceptibility to acute chemical skin irritation.

Geographic variation

The influence of testing common chemicals in different geographical regions (with all the attending influences that may entail) has been touched on above in the section on inter-laboratory comparisons. SDS dose-response and various chemical irritation comparison studies across laboratories in Europe, the US, and China have failed to uncover significant differences in response patterns (26, 28, 29).

Seasonal variation

The dose-response pattern to SDS was compared between studies conducted during winter and summer months in the UK (29). In this study, the winter-time responses were more rapid to develop and more frequent than the summer-time responses, when comparing SDS concentrations of 10%, 20% and 100%. As a result, there was no dose-response seen in the summer time, but a significantly increased response to 100% SDS (v 10% or 20%) in the winter study. These data from the 4-h patch test coincide with earlier studies of surfactant-related skin irritation evaluated by other test methods (49, 50).

Individual variability in skin irritation responses

It is known that humans show wide variation in susceptibility to skin irritation (31). What has also been shown through use of the acute 4-h patch test is the intra-individual variation that can also exist. Within certain studies, comparisons have been made, for individual subjects, of the time required to respond to different chemical irritants. What has been found is that, even for 2 chemicals with similar overall skin irritation potentials (as determined by the response incidence across multiple subjects), individuals may show very different comparative responses. The susceptibility to one chemical (in terms of time to respond) does not always predict the susceptibility to the other (27). Fig. 4 shows reactivity to 2 chemicals, 20% SDS and 0.5% NaOH, plotted for each subject in a study. A large proportion of the subjects responded quite differently to the 2 chemicals, such that the overall correlation coefficient between the 2 chemical responses was low. This same finding was reported in a subsequent study with different chemicals (28). This underscores the need for adequate subject numbers in these studies as well as the potential folly of current animal testing protocols that expect to derive valid results from tests on only one or a very few animals.

Influence of pre-existing atopy

Given what we have learned about individual variations in irritation responses, in developing the 4-h patch test it was also important to know whether there might be differential sensitivities among human subjects at the population level. This would be critical both for the determination of inclusion and exclusion criteria for entry into a study and for the relevance of study results to these different populations. One clinically identified condition that has been associated with differing sensitivity

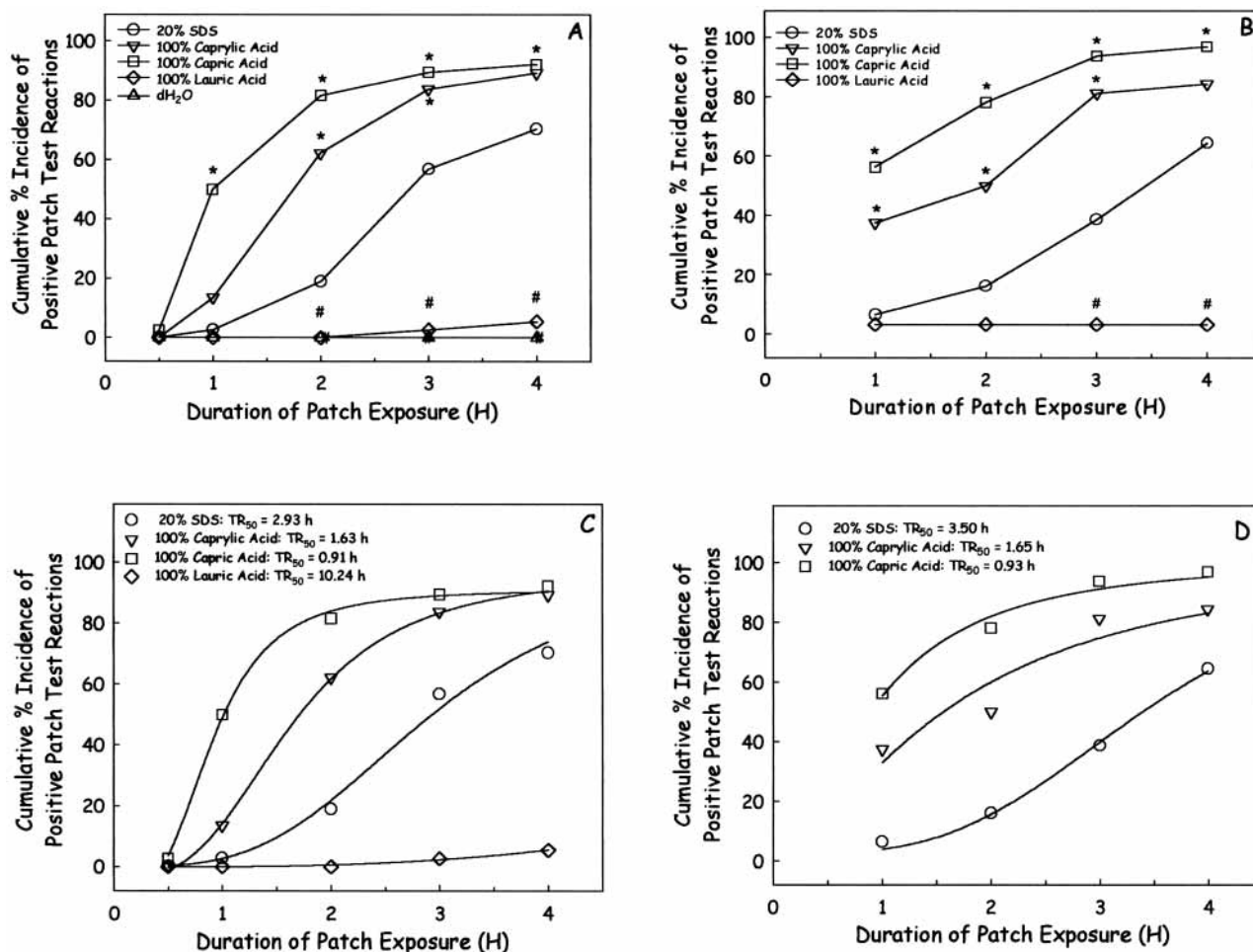


Fig. 3. (A, B) Patches containing the indicated test materials were applied for 0.5 to 4 h, increasing the duration of exposure until a positive response was observed. The asterisk (*) or pound (#) symbols indicate a statistically significant difference (the Fisher exact test) at $p < 0.05$ relative to the response to 20% SDS at the same time point. Graph A represents a study conducted by P&G study and graph B represents a study conducted by Unilever. (C, D) The data shown was graphed using logistic curve fitting procedures and a TR_{50} value (time for 50% of test subjects to respond) was derived for each test material. Graph C is the P&G study and graph D is the Unilever study. Reproduced from (28).

to skin irritants is atopy (51, 52). Thus, a study was conducted to determine if preexisting atopy would have any effect on the sensitivity of the test method (30). Atopic subjects were recruited on the basis of an elevated serum IgE titer to 1 or more common allergens, with or without elevated total serum IgE, and normal subjects were defined as having no detectable allergen specific IgE and total IgE within normal limits. The subjects were tested for acute skin irritation to 20% SDS, 35% cocotrimethyl ammonium chloride (cocoTAC) and 10% hydrochloric acid (HCl). Both test groups responded similarly to all 3 chemical irritants. Though this was a limited study, these results would suggest that the 4-h patch test is not differentially predictive of acute irritation potential among normal versus atopic test subjects and that

atopy not necessarily be used to either include or exclude subjects from test participation.

Influence of skin type and race

Human subjects can be subclassified on the basis of their skin reactivity to sunlight or UVB irradiation. There are 6 gradations of reactivity, from type I ('always burn, never tan') to type VI ('marked constitutive pigmentation'). Black subjects fall into the type VI skin category. There have been reports in the literature regarding possible differences among races in susceptibility to skin irritation (reviewed in (53)). In particular, there have been studies suggesting reduced reactivity among Black subjects (54, 55) perhaps due to a more impenetrable barrier (56). There has been speculation

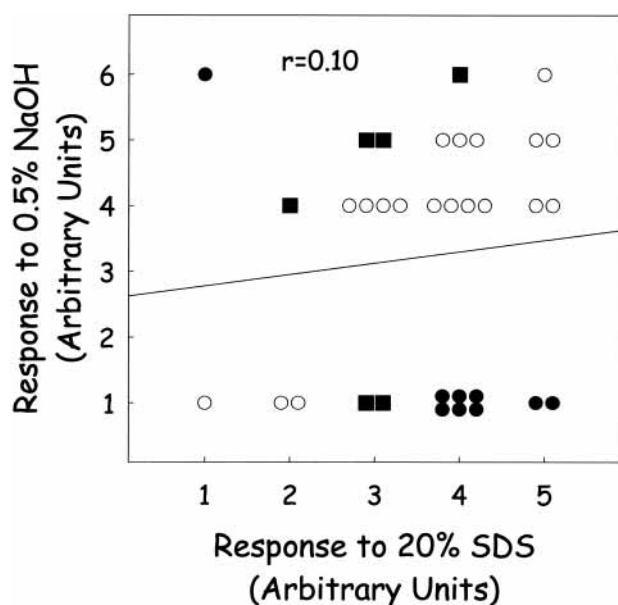


Fig. 4. Within a human 4-h patch test, comparisons were made among test subjects as to their relative response to different irritants. Arbitrary grades were assigned based on how rapidly a subject responded (higher grade=faster response). Comparing 2 strong irritants (20% SDS and 0.5% NaOH), more than $\frac{1}{2}$ the subjects (open circles) showed arbitrary grades within 1 unit for the 2 chemicals. However, a large number of subjects showed arbitrary grades that differed by 2 (filled squares) or greater than 2 (filled circles) units. As a result, the overall correlation in the severity of response (based on time to respond) between the 2 irritants was low. Data from (27).

about differences between other races, particularly a presumed increased reactivity among Asians versus Caucasians (57, 58). However, there is only limited published data supporting this view (59–61). The 4-h human patch test has been a useful procedure to address this controversy.

The study reported earlier on testing in different geographical locations also provided some data on the subject of racial differences. As shown in Fig. 2, there was little difference in SDS dose-responsiveness when populations from different European countries and China were compared. In fact, in that comparison, the Chinese population was among the least reactive. A subsequent study among subjects with different UVB reactive skin types reached the same conclusion (32). No differential SDS dose response was seen in the 4-h patch test (Fig. 5) among subjects with type I to type VI (Black) skin. The lack of difference among the Black subjects contrasts with the older work cited above. However, the earlier studies involved different chemicals, and more chronic and/or repetitive exposures. Under the acute irritation patch test protocol with SDS, Black skin appears to be equally susceptible to irritation relative to other skin types.

Recently, the 4-h patch test has been used to directly investigate the question of differential skin irritation responsiveness between Asian and Caucasian subjects. Populations were tested concurrently at a single laboratory. In an initial set of studies, multiple test chemicals of differing irritation potencies were tested on Caucasian and Chinese subjects (27). There was no significant difference in the response pattern to any of the test substances, even among subjects preselected for heightened sensory skin reactivity. Similar findings were recently reported from studies conducted among Caucasian, Chinese, and Japanese test subjects (33). These results tend to underscore the fact that wide variation in skin irritation among human subjects in general will likely obscure consistently recognizable differences across different human sub-populations. They also highlight the potential pitfalls in ascribing to theories of population differences based only on very limited and often unconfirmed data.

Ethical Considerations in Conduct of the Human 4-h Patch Test

The above discussion has highlighted 3 key features of the human 4-h patch test; it measures acute skin irritation potential in a way that enables appropriate chemical classifications, it shows consistent results in inter-laboratory trials, and it facilitates the investigation of important elements of the acute skin irritation response in the relevant species. What remains to be considered is whether, from an ethical standpoint, this test should ever be conducted. Is it ever ethical to use human beings

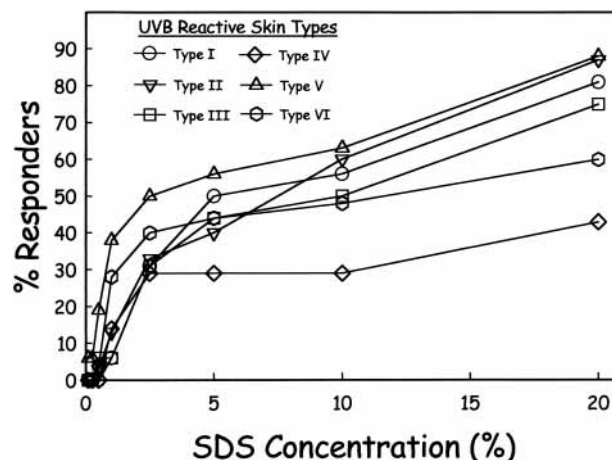


Fig. 5. Subjects ($n=110$) were evaluated for skin pigmentation and ability to tan and placed into one of 6 categories (I: Always burns, never tans; to VI: marked constitutive pigmentation). They were patched for 4h with 0.1% to 20% SDS on the back or upper outer arm. Reproduced from (32).

in a predictive test method where some degree of adverse response is intended to be produced?

Earlier in this paper there is discussion of critical factors for consideration relative to human testing. These included such things as prior knowledge of effects expected, independent review of procedures and risks, and informed consent. A major criticism often directed against tests of this type is that there is an intention knowingly to produce a positive 'adverse' response and then to use that response to make a toxicity prediction or evaluation. How can this practice be considered ethical?

Actually, clinical testing for skin safety is a common practice, particularly in the cosmetics industry. Often the focus is on the 'skin compatibility' testing of finished products (62) or sometimes ingredients (63). Clinical skin sensitization (allergy) testing is generally focused on confirming product safety under exaggerated exposure conditions, not on 'prediction' of clinical sensitization potential (64, 65). However, for prediction of the cumulative or chronic skin irritation potential of relatively mild skin ingredients or products, a number of clinical procedures have been developed over the years, all of which are designed to produce a mild to moderate degree of skin irritation in human skin (66–70). Hence, there is a long precedence for skin irritation testing in human subjects.

It is, first of all, important to consider this question of human skin irritation testing in the overall scheme of the desire to eliminate animal toxicity testing. There is, of course, no suggestion that humans be put at unacceptable risk in order to avoid animal testing. However, given the need to replace the Draize skin irritation test, the inherent validity of the human skin response, and the care taken in this protocol to protect the test subjects, one can make a reasonable argument that this testing is appropriate. In all of the studies conducted to date using this procedure, the large majority of skin reactions recorded (>90%) were mild (grade 1) erythema. Moreover, all induced skin reactions, regardless of severity, were well tolerated by the subjects and resolved quickly. These reactions produced no more significant health risk to the test subjects than any number of trivial skin effects (e.g., insect bites, adhesive bandage irritation, clothing irritation) encountered in day-to-day living. All subjects were well informed of anticipated skin effects and potential, though unexpected, risks and signed well-documented informed consent statements. In situations where subjects were compensated, the amount paid was commensurate with the inconvenience and not sufficient to be considered an inducement to participate.

While the human 4-h patch test is not a drug

study, and provides no direct benefit to the study participant, there is a significant societal benefit. The methodology does ultimately protect the worker and consumer population by providing more relevant and useful data regarding chemical irritation potential and appropriate chemical and product labeling needs. It also serves as the best clinical correlate for future development and validation of in vitro alternative test methods. Considering the high ethical standards to which these studies are conducted, there is much to recommend the method for continued development and use for the assessment of skin irritation potential.

Acknowledgements

The authors are indebted to many colleagues who have supported these studies over several years and to the more than a thousand individuals who have agreed to participate as subjects. In addition, we are very grateful to the positive support given by Dr. Andre Menache, who cares deeply about both the development of non-animal alternatives and the proper and ethical use of human subjects.

References

1. EEC. Annex I to Commission Directive 91/325/EEC of 1st March 1991 adapting to technical progress for the twelfth time Council Directive 67/548/EEC on the approximation of the laws, regulations and administrative provision relating to the classification, packaging and labeling of dangerous substances. *Off J Eur Comm* 1991; L180: 34.
2. OECD guideline for testing of chemicals. 1992: pp. 1–6. 404.
3. Draize J H, Woodard G, Calvery H O. Methods for the study of irritation and toxicity of substances applied topically to the skin and mucous membranes. *J Pharm Exp Therap* 1944; 82: 377–390.
4. Patil S M, Patrick E, Maibach H I. Animal, human, and in vitro test methods for predicting skin irritation. In: Marzulli F N, Maibach H I (eds): *Dermatotoxicology Methods*. Taylor & Francis: Washington, DC, 1998: 89–113.
5. EEC. Annex to Commission Directive 92/69/EEC of 31 July 1992 adapting to technical progress for the seventeenth time Council Directive 67/548/EEC on the approximation of laws, regulations and administrative provisions relating to the classification, packaging and labelling of dangerous substances. *Off J Eur Comm* L84 1992: 36: 1.
6. Department of Transportation. *Method of testing corrosion to the skin*. 1991; Title 49, Appendix A Code of Federal Regulations.
7. Phillips L, Steinberg M, Maibach H I, Akers W A. A comparison of rabbit and human skin response to certain irritants. *Toxicol Appl Pharmacol* 1972; 21: 369–382.
8. Nixon G A, Tyson C A, Wertz W C. Interspecies comparisons of skin irritancy. *Toxicol Appl Pharmacol* 1975; 31: 481–490.
9. Campbell R L, Bruce R D. Comparative dermatotoxicology (I) Direct comparison of rabbit and human primary skin irritation responses to isopropylmyristate. *Toxicol Appl Pharmacol* 1981; 59: 555–563.
10. EEC. Council Directive 93/35/EEC of 14 June 1993 amending for the 6th time Directive 76/768/EEC on the

- approximation of the laws of the Member States relating to cosmetic products. *Off J Eur Comm* 1993; L15: 32.
11. Perkins M A, Osborne R, Johnson G R. Development of an in vitro method for skin corrosion testing. *Fundam Appl Toxicol* 1996; 31: 9–18.
 12. Lewis R W, Basketter D A. Transcutaneous electrical resistance: Application in predicting skin corrosives. In: Elsner P, Maibach H I (eds): *Irritant dermatitis: new clinical and experimental aspects*. Basel: S. Karger AF, 1995: 243–255.
 13. Gordon V C, Harvell J D, Maibach H I. Dermal corrosion, the CORROSITEX[®] system: A DOT accepted method to predict corrosivity potential of test materials. In: Rougier A, Goldberg A M, Maibach H I (eds): *In vitro skin toxicology*. New York: Mary Ann Liebert, 1994: 37–45.
 14. Botham P A, Chamberlain M, Barratt M D, Curren R D, Esdaile D J, Gardner J R, Gordon V C, Hildebrand B, Lewis R W, Liebsch M et al. A prevalidation study on in vitro skin corrosivity testing. The report and recommendations of ECVAM workshop 6. *ATLA-Altern Lab Anim* 1995; 23: 219–255.
 15. Fentem J H, Archer G E B, Balls M, Botham P A, Curren R D, Earl L K, Esdaile D J, Holzhütter H G, Liebsch M. The ECVAM international validation study on in vitro tests for skin corrosivity. 2. Results and evaluation by the management team. *Toxicol In Vitro* 1998; 12: 483–524.
 16. Botham P A, Earl L K, Fentem J H, Roguet R, van de Sandt J J M. Alternative methods for skin irritation testing: the current status. ECVAM Skin Irritation Task Force Report 1. *ATLA Altern Lab Anim* 1998; 26: 195–211.
 17. Boelsma E, Tanjo H, Bodde H E, Ponc M. Assessment of the potential irritancy of oleic acid on human skin: Evaluation in vitro and in vivo. *Toxicol In Vitro* 1996; 10: 729–742.
 18. Augustin C, Collombel C, Damour O. Use of dermal equivalent and skin equivalent models for identifying phototoxic compounds in vitro. *Photodermatol Photoimmunol Ph* 1997; 13: 27–36.
 19. Augustin C, Collombel C, Damour O. Use of dermal equivalent and skin equivalent models for in vitro cutaneous irritation testing of cosmetic products: Comparison with in vivo human data. *J Toxicol Cutan Ocular Toxicol* 1998; 17: 5–17.
 20. Perkins M A, Osborne R, Rana F R, Ghassemi A, Robinson M K. Comparison of in vitro and in vivo human skin responses to consumer products and ingredients with a range of irritancy potential. *Toxicol Sci* 1999; 48: 218–229.
 21. De Brugerolle de Fraissinette A, Picarles V, Chibout S, Kopp M, Medina J, Burtin P, Ebelin M E, Osborne S, Mayer F K, Spake A, Rosdy M, De Wever B, Ettlin R A, Cordier A. Predictivity of an in vitro model for acute and chronic skin irritation (SkinEthic) applied to the testing of topical vehicles. *Cell Biol Toxicol* 1999; 15: 121–135.
 22. Fentem J H, Botham P A, Earl L K, Roguet R, van de Sandt J J M. Prevalidation of in vitro tests for acute skin irritation. In: Clark D G, Lisansky S G, Macmillan R, (eds): *Alternatives to Animal Testing*. (II). *Proceedings of the Second International Scientific Conference Organised by the European Cosmetic Industry*. Newbury, U.K.: CPL Press, 1999: 228–231.
 23. Basketter D A, Whittle E, Griffiths H A, York M. The identification and classification of skin irritation hazard by a human patch test. *Food Chem Toxicol* 1994; 32: 769–775.
 24. York M, Griffiths H A, Whittle E, Basketter D A. Evaluation of a human patch test for the identification and classification of skin irritation potential. *Contact Dermatitis* 1996; 34: 204–212.
 25. Basketter D A, Chamberlain M, Griffiths H A, Rowson M, Whittle E, York M. The classification of skin irritants by human patch test. *Food Chem Toxicol* 1997; 35: 845–852.
 26. Griffiths H A, Wilhelm K P, Robinson M K, Wang X M, McFadden J, York M, Basketter D A. Interlaboratory evaluation of a human patch test for the identification of skin irritation potential/hazard. *Food Chem Toxicol* 1997; 35: 255–260.
 27. Robinson M K, Perkins M A, Basketter D A. Application of a 4-h human patch test method for comparative and investigative assessment of skin irritation. *Contact Dermatitis* 1998; 38: 194–202.
 28. Robinson M K, Whittle E, Basketter D A. A two-center study of the development of acute irritation responses to fatty acids. *Am J Contact Dermatitis* 1999; 10: 136–145.
 29. Basketter D A, Griffiths H A, Wang X M, Wilhelm K P, McFadden J P. Individual, ethnic and seasonal variability in irritant susceptibility of skin: The implications for a predictive human patch test. *Contact Dermatitis* 1996; 35: 208–213.
 30. Basketter D, Blaikie L, Reynolds F. The impact of atopic status on a predictive human test of skin irritation potential. *Contact Dermatitis* 1996; 35: 33–39.
 31. Judge M R, Griffiths H A, Basketter D A, White I R, Rycroft R J G, McFadden J P. Variation in response of human skin to irritant challenge. *Contact Dermatitis* 1996; 34: 115–117.
 32. McFadden J P, Wakelin S H, Basketter D A. Acute irritation thresholds in subjects with Type I-Type VI skin. *Contact Dermatitis* 1998; 38: 147–149.
 33. Robinson M K. Racial differences in acute and cumulative skin irritation responses between Caucasian and Asian populations. *Contact Dermatitis* 2000; 42: 134–143.
 34. Basketter D, Gerberick F, Kimber I, Willis C. *Toxicology of contact dermatitis: allergy, irritancy, and urticaria*. Chichester, UK: John Wiley & Sons.
 35. Basketter D A, Whittle E, Chamberlain M. Identification of irritation and corrosion hazards to skin: An alternative strategy to animal testing. *Food Chem Toxicol* 1994; 32: 539–542.
 36. Liebsch M, Traue D, Barrabas C, Spielmann H, Uphill P, Wilkins S, Wiemann C, Kaufmann T, Remmele M, Holzhütter H G. The ECVAM prevalidation study on the use of EpiDerm for skin corrosivity testing. *ATLA Altern Lab Anim* 2000; 28: 371–401.
 37. Balls M, Fentem J H. The validation and acceptance of alternatives to animal testing. *Toxicol In Vitro* 1999; 13: 837–846.
 38. ECVAM. Statement on the application of the EpidermTM human skin model for skin corrosivity testing. *ATLA-Altern Lab Anim* 2000; 28: 365–366.
 39. Scala R, Fentem J H, Chen J, Derelanko M J, Green S, Harbell J, Kohrman K A, Sauder D N, Stegeman J. Corrositex[®]: An in vitro test method for assessing dermal corrosivity potential of chemicals. 1999 URL: <http://iccvam.niehs.nih.gov/corprep.htm>.
 40. Osborne R, Perkins M A. An approach for development of alternative test methods based on mechanisms of skin irritation. *Food Chem Toxicol* 1994; 32: 133–142.
 41. Bernhofer L P, Barkovic S, Appa Y, Martin K M. IL-1a and IL-1ra secretion from epidermal equivalents and the prediction of the irritation potential of mild soap and surfactant-based consumer products. *Toxicol In Vitro* 1999; 13: 231–239.
 42. Green B G. Regional and individual differences in cutaneous sensitivity to chemical irritants: Capsaicin and menthol. *J Toxicol Cutan Ocular Toxicol* 1996; 15: 277–295.
 43. Chassany O, Duracinsky M. Ethics and clinical trials. *Fundam Clin Pharmacol* 1999; 13: 437–444.

44. Roggeband R, Scailteur V, Basketter D A. Human volunteer studies; a consumer products company view. *Human Exp Toxicol* 1999; 18: 591–593.
45. Mehta C R, Patel N R. A network algorithm for performing Fisher exact test in R X C contingency-tables. *J Am Statist Assoc* 1983; 78: 427–434.
46. Lee C H, Maibach H I. The sodium lauryl sulfate model: An overview. *Contact Dermatitis* 1995; 33: 1–7.
47. Sugar M, Schnetz E, Fartasch M. Does sodium lauryl sulfate concentration vary with time? *Contact Dermatitis* 1999; 40: 146–149.
48. Bruner L H, Carr G J, Chamberlain M, Curren R D. The validation of alternative methods for toxicity testing. In: van Zutphen L F M, Balls M (eds): *Animal alternatives, welfare and ethics*. Elsevier: Amsterdam, 1997: 1091–1098.
49. Frosch P J, Kligman A M. The soap chamber test. *J Am Acad Dermatol* 1979; 1: 35–41.
50. Agner T, Serup J. Seasonal variation of skin resistance to irritants. *Br J Dermatol* 1989; 121: 323–328.
51. Nassif A, Chan S C, Storrs F J, Hanifin J M. Abnormal skin irritancy in atopic dermatitis and in atopy without dermatitis. *Arch Dermatol* 1994; 130: 1402–1407.
52. Klas P A, Corey G, Storrs F J, Chan S C, Hanifin J M. Allergic and irritant patch test reactions and atopic disease. *Contact Dermatitis* 1996; 34: 121–124.
53. Robinson M K. Population differences in skin structure and physiology and the susceptibility to irritant and allergic contact dermatitis: implications for skin safety testing and risk assessment. *Contact Dermatitis* 1999; 41: 65–79.
54. Marshall E K, Lynch V, Smith H W. On dichlorethylsulphide (mustard gas) II. Variations in susceptibility of the skin to dichlorethylsulphide. *J Pharm Exp Therap* 1919; 12: 291–301.
55. Weigand D A, Gaylor J R. Irritant reaction in Negro and Caucasian skin. *South Med J* 1974; 67: 548–551.
56. Wedig J H, Maibach H I. Percutaneous penetration of dipyrithione in man: Effect of skin color (race). *J Am Acad Dermatol* 1981; 5: 433–438.
57. Christensen M, Kligman A M. An improved procedure for conducting lactic acid stinging tests on facial skin. *J Soc Cosmet Chem* 1996; 47: 1–11.
58. Asian skin “more prone” to burning, stinging, redness from H&BA products. *F-D-C Reports: The Rose Sheet* 1998.
59. Rapaport M J. Patch testing in Japanese subjects. *Contact Dermatitis* 1984; 11: 93–97.
60. Ishihara M, Takase Y, Hayakawa R, Nishikawa T, Niimura M, Tokuda Y. Skin problems caused by cosmetics and quasisdrugs: Report by six university hospitals to Ministry of Health and Welfare. *Skin Research (Hifu)* 1986; 28: 80–85.
61. Tadaki T, Watanabe M, Kumasaka K, Tanita Y, Kato T, Tagami H, Horii I, Yokoi T, Nakayama Y, Kligman A M. The effect of topical tretinoin on the photodamaged skin of the Japanese. *Tohoku J Exp Med* 1993; 169: 131–139.
62. Walker A P, Basketter D A, Baverel M, Diembeck W, Matthies W, Mougin D, Paye M, Rothlisberger R, Dupuis J. Test guidelines for assessment of skin compatibility of cosmetic finished products in man. *Food Chem Toxicol* 1996; 34: 651–660.
63. Walker A P, Basketter D A, Baverel M, Diembeck W, Matthies W, Mougin D, Rothlisberger R, Coroama M. Test guidelines for the assessment of skin tolerance of potentially irritant cosmetic ingredients in man. *Food Chem Toxicol* 1997; 35: 1099–1106.
64. Robinson M K, Stotts J, Danneman P J, Nusair T L, Bay P H. A risk assessment process for allergic contact sensitization. *Food Chem Toxicol* 1989; 27: 479–489.
65. Gerberick G F, Robinson M K. A skin sensitization risk assessment approach for evaluation of new ingredients and products. *Am J Contact Dermatitis* 2000; 11: 65–73.
66. Frosch P J, Kligman A M. The chamber-scarification test for irritancy. *Contact Dermatitis* 1976; 2: 314–324.
67. Berger R S, Bowman J P. A reappraisal of the 21-day cumulative irritation test in man. *J Toxicol-Cutan Ocul Toxicol* 1982; 1: 109–115.
68. Lee C H, Maibach H I. Study of cumulative irritant contact dermatitis in man utilizing open application on subclinically irritated skin. *Contact Dermatitis* 1994; 30: 271–275.
69. Andersen K E. Reproducibility of the chamber scarification test. *Contact Dermatitis* 1996; 34: 181–184.
70. Wigger-Alberti W, Hinnen U, Elsner P. Predictive testing of metalworking fluids: A comparison of 2 cumulative human irritation models and correlation with epidemiological data. *Contact Dermatitis* 1997; 36: 14–20.

Address:

M. K. Robinson
 The Procter & Gamble Co.
 Miami Valley Laboratories
 P.O. Box 538707
 Cincinnati, Ohio
 45253-8707
 USA