65

Introduction to the Proof of Claims

Marc Paye
Colgate-Palmolive Research and Development, Inc., Milmort, Belgium

André O. Barel
Free University of Brussels, Brussels, Belgium

With the continuous increase in the variety of cosmetic products proposed to consumers over these last decades, it has become more and more difficult for them to decide what the most appropriate products are for their needs. Aware of such difficulties, cosmetic manufacturers have understood that the success of a product today is not only a question of performance, but also a question of how it is promoted to the potential buyer. Progressively, product promotion took more importance and advertising claims became more aggressive and closer to the limit of what could be scientifically shown and consumer-perceived. In order to monitor the claims made about cosmetic products and protect the consumer against misleading advertisement, several national/federal agencies have issued rules under the form of laws, or directives, to ensure that proper substantiation of claims exists. Furthermore, relying on such rules, competitors always remain ready to challenge unfair or doubtful claims. Last but not least, the consumers themselves have become more critical and, when they feel that their product does not provide the properties that it claims, do not hesitate to stop buying the product as well as the other products of the same brand. It has thus become a priority for the cosmetic chemist to be able to show and substantiate the properties that are claimed for his or her product.

The objectives of this introduction to the proof of claims are as follows:

1. To briefly describe the regional requirements related to the proof of claims.
2. To explain the different existing categories of claims.
3. To review the types of support that can be made.

REGIONAL REQUIREMENTS

Although all over the world the motivation exists to protect the consumer against misleading claims, the current situation is quite different between Europe, the United States, and Japan regarding claim substantiation requirements and the limit of definitions of a cosmetic product. This latter point has been discussed in previous sections (see Part 7 of this book). Specific regulations are summarized hereafter.

Marcel Dekker, Inc.
270 Madison Avenue, New York, New York 10016

The United States

The U.S. federal law does not require premarketing proof of claims but prohibits false advertisement. In the case of a challenge (e.g., by a competitor, a consumer association, a government agency), the manufacturer must be prepared to defend the claims made on the product. However, the challenger has to first provide arguments questioning the validity of the claim. It is quite frequent in the United States for claims to be challenged, and most companies preferably develop scientifically valid claims support strategies and dossiers before marketing their new product.

Several federal authorities controlling cosmetic claims exist. The U.S. Food and Drug Administration (FDA), through the Federal Food, Drug and Cosmetic Act (FDC Act) and the Fair Packaging and Labeling Act (FPL Act), has the main jurisdiction and responsibility on claims made on the labeling of the products. The Federal Trade Commission (FTC) monitors product advertising (e.g., television, radio, magazine). When a claim is related to both advertising and labeling, the two agencies usually collaborate with each other. However, even if both agencies condemn in their respective Acts consumer misleading, neither clearly defines the legal standard for illegality. The manufacturer will thus rely upon a ''reasonable'' basis to support its claims, and most challenges will be treated on a case-by-case basis.

Another significant control of advertising is performed by the National Advertising Division (NAD) of the Council of Better Business Bureaus, which is a self-regulatory, nongovernmental body evaluating the truth and accuracy of challenged advertising. NAD is usually the first body to receive complaints about claims from competing companies or consumer associations. Through several control and communication steps between the two challengers, NAD may decide to involve the appropriate government Agency (FDA or FTC). Several examples of challenged claims have been summarized by Davis and McNamara [1] and Friedel [2], that can help in understanding the U.S. situation.

The European Union

In the European Union, cosmetic claims substantiation is regulated by the 6th Amendment to the Cosmetic Directive, effective since January 1, 1997. In that amendment, it is stated that cosmetics and toiletries manufacturers making claims for their products have to demonstrate the proof of their claims. The dossier containing these proofs has to be readily available if requested by the competent authority. The dossier may be written in English or in the language of the country where it is deposited. More details on the Cosmetic Directive can be obtained in Chapter 60.

As in the United States, no clear definition about the meaning of proof of claims has been given, so manufacturers have to define by themselves what they consider to be a ''reasonable'' and ''acceptable'' proof for their claim. Such a consideration will often depend on the type and originality of the claim, the type of product and the market in which it will compete, the consequences and benefits that the consumer can expect from the claimed effect, and the image, scientific honesty, and competency of the manufacturer.

Although the 6th Amendment to the Directive aimed at uniformizing the differences between countries, big differences still exist regarding how to monitor the proof of claims dossiers, which is basically subject to the interpretation of the Directive within individual state laws. In most countries, such monitoring will essentially be postmarketing in the case of a challenge, but in some countries a premarketing review of the claims can be

Marcel Dekker, Inc.
270 Madison Avenue, New York, New York 10016

requested by a National Review Board (e.g., Greece). Some types of claims are also not uniformly accepted for cosmetic products by all E.U. members; this is, for instance, the case of claims that can be overlooked as ''medically oriented'' such as ''dermatologically tested'' or ''hypoallergenic'' (e.g., not allowed in Denmark for cosmetics). The decision for acceptable claims and reasonable supporting dossier should thus always be reviewed in line with the individual national laws, if any, of the country where marketing is intended.

Japan

In Japan, the situation is different in the sense that claims are reviewed before marketing of the cosmetic product. The Ministry of Health and Welfare (MHW) has to provide a license to the product to allow its marketing. The limit of the definition of purely cosmetic products is also different in Japan than in the European Union and United States, with the existence of ''quasi-drugs'' classified between cosmetics and drugs. This has been reviewed in Chapter 62.

CATEGORIES OF CLAIMS

However they are used (e.g., label, television, or magazine advertising), claims related to cosmetic products can be subdivided into several categories. Table 1 summarizes these categories and provides some examples for each. As previously explained, all claims are not applicable everywhere in the world for cosmetic products and can fall under different regulations in some places.

TABLE 1 Categories of Claims

Categories of Claims	Examples
Claims related to physical and chemical properties	Contains x% of an active
	Neutral pH
	20% more in the bottle
	More concentrated
Claims related to the test procedure or to an endorsement	Dermatologist, dermatologically tested
	Tested under the Good Clinical Practices Principles
	Tested and approved by an institute
Safety-related claims	Mild, gentle on the skin
	For sensitive skin
	Skin-repair properties
	Hypoallergenic
Objective efficacy claims	Moisturizing, hydrating
	Improves elasticity, firmness of skin
	Skin-whitening effect
	Sunscreen effect
	Antiperspirant, deodorant
Subjective claims	Skin will feel softer; more hydrated
	With a pleasant feel, texture
	Smells fresher
Cultural claims	Contains 100% natural ingredients
	Not tested on animals
Juxtaposition claims	Contains an ingredient known for such a property

1. Claims related to physical and chemical properties of the product can be substantiated by measuring directly the claimed characteristic in the product by an analytical method. The measurement methodology has, however, to be well established and validated.

2. Claims related to the test procedure or to an endorsement by an outside authority simply describe the way, person/title, or place where the product has been tested. They are usually perceived by the consumer as proof of a well-tested, quality product. It is essential for such claims to show the property that the consumer can expect from the product, even if it is not directly advertised. For example, the claim ''dermatologist tested'' means that the product has been tested by a dermatologist, but also implies that the results of the test were good and that the product is, e.g., mild on the skin, or has a skin-beneficial property shown by the dermatologist.

3. Safety-related claims make the consumer confident about the innocuousness of the product and the benefits to their body. These claims usually require clinical tests on human volunteers according to protocols published in the scientific literature and performed under high-quality standards. In some cases, in vitro tests can also be accepted if it can be shown that they are able to prove the claimed property for the type of product in test.

4. Objective efficacy claims are probably the most frequent, and those inducing the highest expectation from consumers. This is why they require solid efficacy data dossiers. Many biometric methodologies currently allow getting a direct measurement of the skin properties [3] that are expected to be respected or modified by the cosmetic product. In vivo tests with human volunteers are often recommended or even the only possibility offered to the cosmetic chemist, but other types of tests can also be used in some cases, such as cell-culture tests [4–6].

5. Subjective claims are related to a property or function of the product that is perceived by the consumer. The property does not necessarily have to be objectively substantiated by direct measurement. Only tests on human volunteers can be performed, such as sensory tests (Chap. 71) or well-designed consumer tests.

6. Cultural claims are usually related to the composition of the product and take advantage of the current trends. Their value to the consumer is often dependent on the education, country, or environment. They link the composition of the product to ecological, ethical, or moral considerations (e.g., naturality of ingredients, absence of tests on animals).

7. Juxtaposition claims refer to the presence of an ingredient in a product and to the known property of the ingredient, without claiming that the complete product has the property. This type of claim can be supported by proving the presence of the ingredient in the product (analytical methods) and relating the claimed property to that ingredient through literature data or any type of appropriate test on the pure ingredient.

Several of these categories can be further subdivided in terms of absolute or comparative claims. The following four subcategories can be described as follows:

(a) Noncomparative claims: they simply refer to a property of the product without any direct comparison to another product. However, it is obvious that even if not classified as such, all claims contain a comparative connotation. For instance, claiming that a product is mild means that this is not the case for all other products. Similarly, claiming that a product is hydrating for the skin refers to the hypothesis that some other products are not. Examples include claims that a product is mild, hydrating, protects the skin, and softens the skin.

Marcel Dekker, Inc.
270 Madison Avenue, New York, New York 10016

(b) Claims comparing a new product to the one it replaces in the market place: in the proof of claim dossier, a direct comparison between the two products will be required. The kind of test depends on the claim. Examples include x% more efficacy, milder than ever, and even milder than before.

(c) Purely comparative claims comparing the new product to competitive ones for the claimed property: this kind of claim is likely to be challenged by competitors and requires a solid supporting dossier where direct comparison between the products is made. The test methodology has to be well justified and validated for the objective. Such comparative claims are quite usual in the United States; in Europe they are allowed only under restricted and severe conditions. Examples include milder than product x and y, and more hydrating than product z.

(d) Absolute claims: the comparison is not limited to a few mentioned products as previously discussed, but the product claims to be the best in the market for a given property or to completely fulfill a specific function. Such claims require very solid dossier and can be invalidated if even one competitive product can be shown to be superior on this property. Examples include the mildest, nothing more hydrating, total protection, and complete diet for the skin.

TYPE OF SUPPORT

Whenever the nature of the effect or the product justifies it, the claims on cosmetic products must be shown. However, the type of support has never been clearly and officially defined, so that any kind of support could be acceptable at the condition it can be scientifically and reasonably justified. Different ways to support cosmetic claims [7] are reviewed hereafter; some have already been briefly discussed earlier in this chapter.

Comparison to a Similar Formula

If the product is derived from another formula by a minor modification, it is not always necessary to repeat the claim-supporting test for the new product. In such a case, it has however to be clearly justified that the change is not to affect the claimed property. Depending on the claims, certain modifications can be considered minor or not. Similarly, for a line of products with minor differences between individual products, some claims can often be shown on only a few products of the line and then extended to the other products.

Literature Search

For some types of claims, literature data can be considered as effective claim-support dossier. This is, for instance, the case of claims on ingredients entering into the composition of the cosmetic product; often, the proof of the ingredient property can be found into the scientific literature. It should be noticed, however, that peer-reviewed literature usually has more credit than supplier literature in the case of a challenge, although the latter can also be used if supported by well-controlled tests.

In Vitro Tests

In vitro tests never have the same value as in vivo data obtained from clinical tests run on human volunteers. This is why they are mostly used in combination with other types

Marcel Dekker, Inc.
270 Madison Avenue, New York, New York 10016

of data. However, in some cases (depending on the claim or the availability of alternative tests), in vitro tests can be used on their own to support claims, provided that the test is proven to be scientifically valid for the intended objective. From the most promising and usual in vitro tests, 3-D cell-culture methodologies probably receive the most credit for investigating many cosmetic product properties, from skin mildness to more specific properties like sun protection [4–6].

For special dermatocosmetic claims, such as fat reduction and anticellulites effects, in vitro data are mostly presented as direct support. However, in such cases, scientifically valid in vivo testing about the efficacy of these treatments is not always available, and often cumbersome, difficult, and of long duration. Extrapolation of in vitro data or even of supplier literature on the efficacy of the actives is often used, without really proving the claims.

In Vivo Tests on Human Volunteers: Clinical Studies

The most direct proof of a claim is to show the product effect directly on the human volunteers using the product. Many test protocols may be used depending on the objective. Most protocols have been published in scientific literature and are well-established tests. They go from very exaggerated application conditions [8–11] to normal usage of the product by the subjects in the laboratory or at home [12,13]. It is obvious that the more realistic the application condition, the more powerful the demonstration of the effect. Besides the application procedure, these protocols can also differ by the assessment technique of the claimed effect: scoring of the effect by an expert evaluator, objective measurement of the property by a biometric technique, or self-assessment of the subjective effect by the user.

Assessment by an Expert Evaluator

This type of evaluation concerns a cosmetic effect that can be determined by visual, tactile, or olfactive assessment. Examples of scoring scales for the assessment of dry skin have been given by Serup [14]. The evaluator is trained to make such an assessment, reliable and fully independent of the product manufacturer. In some cases, the evaluator will be a dermatologist, an ophthalmologist, or a dentist, but this is not mandatory provided that the evaluator can justify his/her qualification. When the test protocol is appropriate, expert evaluations are frequently combined with other assessment methods. Examples of claims easily supported by expert evaluation are: skin whitening, antiwrinkle, hair shine, and deodorancy. Safety claims are also appropriate for such an assessment to check the absence of erythema or dryness after product application.

Measurement by Means of Biometric Methodologies

A huge amount of biometric methods have developed over the two last decades, which now allow objective and quantitative measurement of most skin properties, such as elasticity, firmness, color, barrier properties, moisture content, relief, and blood flow [3,15,16]. This kind of evaluation is highly valuable thanks to its objectivity and sensitivity, and can identify small differences between products to support comparative claims. Those biometric measurements must however take into account several key rules: 1) many external factors can affect the measurements that have to be made according to specific guidelines [17–20]; 2) the interpretation of the data has to be done by an expert in the field able to relate the collected data to physiological parameters; and 3) the instruments must be highly

Marcel Dekker, Inc.
270 Madison Avenue, New York, New York 10016

reliable. Under such conditions, instrumental measurements are highly valuable and have revolutionized the way of supporting cosmetic claims. However, instrumental methodologies have nowadays become so sensitive that questions can sometimes be raised about the relevancy of such small differences between products for consumers who are not always able to detect them.

Self-Assessment of the Effect by the Volunteers

When used in clinical tests, this type of evaluation is usually combined with other assessments. It is, however, not applicable to all test procedures, and requires that the product has been placed in contact with a sufficiently large area of the body to provide an effect that can be self-perceived. When confirming objective measurements of the property, this self-evaluation is extremely powerful because it expresses that the measured effect is really meaningful to the consumer.

Sensory Tests with Human Volunteers

The self-perception of the product effect by the volunteers can be obtained independently of a clinical test; in such a case a specific test procedure has to be designed. Sensory tests are limited to the so-called sensory claims, which clearly state that the product modifies the perception of a property of the skin or hair (e.g., you can feel your skin ''softer'' or ''more hydrated''). When the sensory effect of the product is obvious and can be easily perceived by a large majority of people, the test can be performed on a panel of regular (or ''naïve'') volunteers, without any specific selection criteria regarding their capability to feel differences. However, the self-perception of stimuli or of a skin feel is very variable between subjects and, often, differences between products are not so obvious for a ''naïve'' user; it is then necessary to run the sensory test either on a very large panel (sometimes several hundreds of volunteers) or to use a panel of volunteers specifically selected and trained to perceive small differences for the kind of product in test. (For more details on sensory tests, the reader is referred to Chapter 71.)

Consumer Tests

These tests are performed at the end of the development phase of the product, and consist in providing consumers with the product to use at home for a certain period of time, according to their usual habits and practice. Expected sensory/efficacy properties of the product can be checked from these tests by means of a questionnaire filled out by the users.

The information collected from consumer tests are very helpful in supporting cosmetic claims, as it will reassure the manufacturer that its product is not misleading the user about the claimed property. Consumer tests, like sensory tests, are mostly used to support claims such as those related to odor perception, skin sensation, tactile or visible properties of skin or hairs, and taste of oral-care products. However, because of the subjectivity of the data, such tests, to be valuable, must be performed very carefully. The questionnaire has to be prepared by a specialist on this kind of test, and cannot be oriented toward the answers of the users. For more details, the reader is referred to specific guidelines for consumer testing [21].

Multiapproach for Claim Support

As previously shown, all the approaches described for supporting claims on cosmetic products have own advantages as well as weaknesses. In order to combine the strengths

Marcel Dekker, Inc.
270 Madison Avenue, New York, New York 10016

of several, a multitest or multievaluation approach, combining expert assessments, instrumental measurements, and subjective data from the user, is often considered to be an ideal support for a claim. However, depending on the type of claim, for cost and time reasons, it is not always necessary to go so far in the dossier if one test obviously and undoubtly provides the proof of the claimed effect.

CONCLUSION

Claims on cosmetic products are extremely varied and often depend on the product, the market, and the current trends. However, several claims have been used on different product types for many years. Testing strategy for some is described in the following chapters. They cover some safety-related claims (e.g., mildness, sensitive skin–designed products, noncomedogenicity claims), some efficacy claims (e.g., skin-hydration effect, smoothing and antiwrinkle effect) and sensory claims. The proposed tests especially aim at guiding the skin scientists to design their own protocols based on reasonable scientific considerations, and do not intend to impose strict testing procedures.

REFERENCES

1. Davis JB, McNamara SH. Regulatory aspects of cosmetic claims substantiation. In: Aust LB, ed. Cosmetic Claims Substantiation. Vol. 18. New York: Marcel Dekker, 1998:1–20.
2. Friedel SL. Technical support for advertising claims. J Toxicol-Cutan Ocular Toxicol 1992; 11:199–204.
3. Willoughby M, Maibach HI. Cutaneous biometrics and claims support. In: Aust LB, ed. Cosmetic Claims Substantiation. Vol. 18. New York: Marcel Dekker, 1998:69–86.
4. Jackson EM. Supporting advertising claims. Reviewing a three-dimensional in vitro human cell test. Cosmet Toilet 1993; 108:41–42.
5. Majmudar G, Smith M. In vitro screening techniques in dermatology. A review of the tests, models and markers. Cosmet Toilet 1998; 113:69–76.
6. Roguet R. Intérêt des modèles de peaux reconstruites en cosmétologie. Cosmétologie 1997; 13:38–43.
7. DGCCRF-D'UMA Commission 30. Evaluation de l'efficacité des produits cosmétiques. Les recommandations de la DGCCRF. Cosmétologie 1997; 15:44–46.
8. Frosch PJ, Kligman AM. The soap chamber test: a new method for assessing the irritancy potential of soaps. J Am Acad Dermatol 1979; 1:35–41.
9. Sharko PT, Murahata RI, Leyden JJ, Grove GL. Arm wash with instrumental evaluation—a sensitive technique for differentiating the irritation potential of personal washing products. J Derm Clin Eval Soc 1991; 2:19–27.
10. Clarys P, Manou I, Barel AO. Influence of temperature on irritation of the hand/forearm immersion test. Contact Dermatitis 1997; 36:240–243.
11. Charbonnier V, Morrison BM Jr, Paye M, Maibach HI. Open application assay in investigation of subclinical irritant dermatitis induced by sodium lauryl sulfate (SLS) in man: advantage of squamometry. Skin Res Technol 1998; 4:244–250.
12. Jackson EM, Robillard NF. The controlled use test in a cosmetic product safety substantiation program. J Toxicol-Cutan Ocular Toxicol 1982; 1:117–132.
13. Paye M, Gomes G, Zerweg Ch, Piérard GE, Grove GG. A hand immersion test in laboratory-controlled usage conditions: a need for sensitive and controlled assessment methods. Contact Dermatitis 1999; 40:133–138.
14. Serup J. EEMCO guidance for the assessment of dry skin (xerosis) and ichtyosis: clinical scoring systems. Skin Res Technol 1995; 1:109–114.

Marcel Dekker, Inc.
270 Madison Avenue, New York, New York 10016

15. Wiechers JW, Barlow T. Skin bioengineering techniques for substantiating cosmetics claims. Cosmet Toilet 1998; 113:81–83.
16. Kajs TM, Gartstein V. Review of the instrumental assessment of skin: effects of cleansing products. J Soc Cosmet Chem 1991; 42:249–271.
17. Rogiers V, Derde MP, Verleye G, Roseeuw D. Standardized conditions needed for skin surface hydration measurements. Cosmet Toilet 1990; 105:73–82.
18. Berardesca E. EEMCO guidance for the assessment of stratum corneum hydration: electrical methods. Skin Res Technol 1997; 3:126–132.
19. Piérard GE. EEMCO guidance for the assessment of skin colour. J Am Acad Derm Venereol 1998; 10:1–11.
20. Morrison BM Jr. ServoMed evaporimeter: precautions when evaluating the effect of skin care products on barrier function. J Soc Cosmet Chem 1992; 43:161–167.
21. The Advertising Research Foundation. Guidelines for the public use of market and opinion research. New York, NY, 1981.

Marcel Dekker, Inc.
270 Madison Avenue, New York, New York 10016

Marcel Dekker, Inc.
270 Madison Avenue, New York, New York 10016

66

Tests for Sensitive Skin

Alessandra Pelosi, Sabrina Lazzerini, and Enzo Berardesca
University of Pavia, Pavia, Italy

Howard I. Maibach
University of California at San Francisco School of Medicine, San Francisco, California

INTRODUCTION

Sensitive skin is a condition of subjective cutaneous hyperreactivity to environmental factors. Subjects experiencing this condition report exaggerated reactions when their skin is in contact with cosmetics, soaps, and sunscreens, and they often report worsening after exposure to dry and cold climate.

Although no sign of irritation is commonly detected, itching, burning, stinging, and a tight sensation are constantly present. Generally, substances that are not commonly considered irritants are involved in this abnormal response. They include many cosmetic ingredients such as dimethyl sulfoxide, benzoyl peroxide preparations, salycilic acid, propylene glycol, amyldimethylaminobenzoic acid, and 2-ethoxyethyl methoxycinnamate [1].

Sensitive skin and subjective irritation are widespread but still far from being completely defined and understood. Burckhardt [2] hypothesized a correlation between sensitive skin and constitutional anomalies and/or other triggering factors such as occupational skin diseases or chronic exposure to irritants. On the other hand, Bjornberg [3] proposed that no constitutional factors play a role in the pathogenesis of sensitive skin, although the presence of dermatitis shows a general increase in skin reactivity to primary irritants lasting months.

EPIDEMIOLOGICAL STUDIES

Recent findings suggest that higher sensitivity can be attributable to different mechanisms. Hyperreactors may have a thinner stratum corneum with a reduced corneocyte area, causing a higher transcutaneous penetration of water-soluble chemicals [4]. In 1977, Frosch and Kligman [5], by testing different irritants, showed a 14% incidence of sensitive skin in the normal population, likely correlated to a thin permeable stratum corneum, which make these subjects more susceptible to chemical irritation.

Many epidemiological studies have been carried out to assess whether or not a correlation with sex, skin type, or age could be found. Contradictory findings have been re-

Marcel Dekker, Inc.
270 Madison Avenue, New York, New York 10016

ported. Some investigators [6,7] documented a higher reactivity to irritants mostly in females, but other experimental studies did not confirm this observation. Bjornberg [8], using six different irritants by patch-test application, found no sex-related differences. Moreover Lammintausta [9], studying the response to open and patch-test application of sodium lauryl sulphate (SLS), found mild interindividual variations in transepidermal water loss (TEWL) and dielectric water content (DWC) values, but no sex-related differences in the reaction pattern.

In 1982, Frosch [10], using dimethylsulfoxide, show a correlation between the minimal erythema dose (MED) and the response to irritants; the higher the inflammation, the lower the MED. Subsequently, a correlation between skin reactivity and skin type was reported; higher reactions were detected in subjects with skin type I [11]. Moreover, in eczema skin reactivity is enhanced [12]. Studies performed on animal models showed that strong irritant reactions in guinea pigs significantly reduced the threshold of skin irritation [13]. On the other hand, hyporeactive states may be induced by skin treatment. Subclinical dermatitis, after repeated cutaneous irritation by open application, may induce skin hyporeactivity [14]. This can also be one of the mechanisms of false-negative patch test.

Skin reactivity seems also to change depending on age. The literature is contradictory. For example, Nilzen and Voss Lagerlung [15] reported higher reactivity patch-test reactions to soaps and detergents in the elderly, whereas Bettley and Donoghue [16] reported a lower reactivity in the same group.

Coenraads et al. [17] showed a higher skin reactivity to croton oil in the older patient group, but no differences by testing thimochinone or croton aldehyde. Recently, Grove [18], by testing croton oil, cationic and anionic surfactants, weak acids, and solvents, reported a lower susceptibility in older subjects in terms of less-severe skin reactions. Aged skin seems to have a reduced inflammatory response either to irritants or to irritation induced by UV light [19]. On the other hand, after irritating the skin, increased TEWL values were recorded in the older subjects compared with the young. This finding could be related to a deficient ''early warning detection system'' in the elderly. The lack of any visible response can lead to continued exposure to external irritants and higher risk of damage to skin-barrier function.

CLINICAL PARAMETERS

Because of the lack of clinical signs, the phenomenon of sensitive skin is difficult to document. Attempts to identify clinical parameters in subjects with subjective irritation indicate that these individuals tend to have a less hydrated, less supple, more erythematous, and more teleangiectatic skin compared with the normal population. In particular, significant differences were found for erythema and hydration/dryness [20].

TESTS FOR SENSITIVE SKIN

Recently, because no visible clinical signs of irritation are detected in sensitive skin, new methods of sensory testing have been increasingly used to provide definite information.

QUANTITATION OF CUTANEOUS THERMAL SENSATION

The superficial skin layer includes sensory nerve fibers connected to specialized receptors such as corpuscles or naked nerve endings. A Beta fibers, myelinated (conduction velocity

of 2–30m/sec), mediate the touch, vibration, and pressure sensation. A Delta fibers, smaller and myelinated (conduction velocity of >30m/sec), mediate cold and pain sensation. C fibers, small and nonmyelinated, mediate warm and itching sensation. Quantitative sensory testing (QST) methods have been used mainly to study the impairment of somatosensory function in neurological diseases; particularly in dermatology, thermal sensation testing analysis is becoming the most used QST technique [21]. It assessed function in free nerve endings and their associated small myelinated and nonmyelinated fibres. This method is able to measure the threshold of warm and cold sensation as well as hot and cold pain.

All modern automated thermal testing instruments include a thermode (Poltrier device) with semiconductor junctions made of different metals. Depending on the polarity of the electric current, the skin is heated or cooled; different thermic sensations are reproduced on the different sides of the junctions.

In the center of the thermode, a thermocouple records the temperature. TSA 2001 (Medoc Company, Ramat Yshai, Israel) is considered one of the most advanced portable thermal sensory-testing devices. Basically, its measures the hot or cold threshold and the suprathreshold pain magnitude. It operates between 0°C and 54°C.

The thermode in contact with the skin produces a stimulus whose intensity increases or decreases until the subject feels the sensation. As the sensation is felt, the subject is asked to press a button. The test is then repeated two more times in order to get a mean value. Using this method, artefacts can occur because of the lag time the stimulus needs to reach the brain. This inconvenience can be avoided by using relatively slow rates of increasing stimuli. The stimulus can also be increased stepwise and the subject is told to say whether or not the sensation is felt. When a positive answer is given, the stimulus is decreased by one-half the initial step and so on, until no sensation is felt. The subject's response determines the intensity of the next stimulus. The limitation of this second method is that a longer performance time is required.

STINGING TEST

Stinging seems to be a variant of pain that develops rapidly and fades quickly any time the appropriate sensory nerve is stimulated. Although this method lacks objective criteria, it is widely accepted as a marker of sensitivity and has often been used in skin-irritation studies [5]. It provides information to establish those subjects experiencing invisible cutaneous irritation.

It is performed by applying to the skin hydrosoluble substances such as lactic acid or capsaicin. The test is usually carried out on the nasolabial fold, a site richly innervated with sensory fibers. Subjects first undergo a facial sauna for 5 to 10 minutes, then an aqueous lactic acid solution (5–10% according to different methods] is rubbed with a cotton swab on the test site. In order to have a more reliable response, it is recommended to apply an inert control substance, such as saline solution, to the contralateral test site. After application, within a few minutes a moderate to severe stinging sensation occurs for the ''stingers group.'' These subjects are then asked to describe the intensity of the sensation using a point scale. Hyperreactors, particularly those with a positive dermatological history, have higher scores. An alternative test involves the application of 2 mL of 90% aqueous dimethylsulfoxide (DMSO) in a small glass cup on the cheek for 5 minutes. This procedure causes intense burning in stingers and, after application, tender wheal and persistent erythema often occur. By contrast, lactic acid produces no visible changes. Us-

ing this screening procedure, 20% of the subjects exposed to 5% lactic acid in a hot, humid environment were found to develop a stinging response [5]. Lammintausta et al. confirmed these observations [22]. In this study, 18% of subjects were identified as stingers. In addition, stingers were found to develop stronger reactions to materials causing nonimmunological contact urticaria, to have increased values of TEWL and increased blood-flow velocimetry values after application of an irritant under patch test.

EVALUATION OF ITCHING RESPONSE

Recent studies show that a new class of C fibers with an exceptionally lower conduction velocity and insensitivity to mechanical stimuli can likely be considered as afferent units that mediate the itchy sensation [23]. Indeed, this subjective feeling has been extensively investigated but no explanation of the individual susceptibility to the itching sensation, without any sign of coexisting dermatitis, has been found. Laboratory investigation of the itch response has also been limited.

An itch response can be experimentally induced by topical or intradermal injections of various substances such as proteolytic enzymes, mast cell degranulators, and vasoactive agents. Histamine injection is one of the more common procedures: histamine dihydrochloride (100 μg in 1 mL of normal saline) is injected intradermally in one forearm. Then, after different time intervals, the subject is asked to indicate the intensity of the sensation using a predetermined scale and the duration of itch is recorded. Information is always gained by the subject's self-assessment.

A correlation between whealing and itching response, produced by applying a topical 4% histamine base in a group of healthy young females, has been investigated by Grove. The itching response was graded by the subjects using the following scale: none, slight, moderate, and intense. The data showed that, despite the fact that 90% of the wheals were greater than 8 mm in diameter, only 50% of the subjects experienced pruritus; patients with large wheals often had no complaints of itching, suggesting that the dimensions of the wheals do not correlate well with pruritus. In addition, itch and sting perception seem to be poorly correlated. Grove [18] compared the cumulative lactic acid sting scores with the histamine itch scores in 32 young subjects; all the subjects who were stingers were also moderate to intense itchers, while 50% of the moderate itchers showed little or no stinging response.

Yosipovitch [24], studying the effects of drugs on C fibers during experimentally induced itch, showed that topically applied aspirin rapidly decreases histamine-induced itch. This result can be attributed to the role that prostaglandines play in pain and itch sensation [25]. Localized itching, burning, and stinging can also be a feature of nonimmunological contact urticaria. This condition, still not completely defined, is characterized by a local wheal and flare after exposure of the skin to certain agents. Different combinations of mediators such as non–antibody-mediated release of histamine, prostaglandins, leukotriens, substance P, and other inflammatory mediators may likely be involved in the pathogenesis of this disorder [26]. The fact that prostaglandins and leukotriens may play a role in the inflammatory response is supported by the inhibition of the common urticants by both oral acetylsalicylic acid and indomethacin and by topical diclofenac and naproxen gel [1]. Several substances, such as benzoic acid, cinnamic acid, cinnamic aldehyde, and nicotinic acid esters, are capable of producing contact nonimmunological urticaria, eliciting local edema and erythematous reactions in half of the individuals. Provocative tests

Marcel Dekker, Inc.
270 Madison Avenue, New York, New York 10016

are usually used to identify subjects experiencing this condition: benzoic acid, sorbic acid, or sodium benzoate in open application well reproduce the typical symptoms in subjects suspected of contact nonimmunological urticaria.

WASHING AND EXAGGERATED IMMERSION TESTS

The aim of these tests is to identify a subpopulation with an increased tendency to produce a skin response. In the washing test [27], subjects are asked to wash their face with a specific soap or detergent. After washing, individual sensation for tightness, burning, itching, and stinging is evaluated using a point scale previously determined. The exaggerated immersion test is based on soaking the hands and forearms of the subjects in a solution of anionic surfactants (such as 0.35% paraffine sulfonate, 0.05% sodium laureth sulfate-2EO) at 40°C for 20 minutes. After soaking, hands and forearms are rinsed under tap water and patted dry with a paper towel. This procedure is repeated two more times, with a 2-hour period between each soaking, for 2 consecutive days. Before the procedure, baseline skin parameters are evaluated. The other evaluations are taken 2 hours after the third and sixth soakings and 18 hours after the last soaking (recovery assessment). All of the skin parameters are performed after the subjects have rested at least 30 minutes at 21 $\pm$ 1°C.

CORNEOSURFAMETRY

This method, recently described [28], investigates the interaction of surfactants with the human stratum corneum. It is performed as follows: cyanoacrylate skin-surface stripping (CSSS) is taken from the volar aspect of the forearm and sprayed with the surfactant to be tested. After 2 hours, the sample is rinsed with tap water and stained with basic fuchsin and toluidine blue dyes for 3 minutes. After rinsing and drying, the sample is placed on a white reference plate and measured by reflectance colorimetry (Chroma Meter® CR200; Minolta, Osaka, Japan).

The index of mildness (CIM = luminacy L*-chroma C*) is taken as a parameter of the irritation caused by the surfactant. This index has a value of 68 $\pm$ 4 when water alone is sprayed on the sample and decreases when surfactant is tested, with stronger surfactants lowering the values. Piérard et al. [29], testing different shampoo formulations in volunteers with sensitive skin, showed that corneosurfametry correlates well with in vivo testing. A significant negative correlation ($p < 0.001$) was found between values of colorimetric index of mildness (CIM) and the skin compatibility parameters (SCP) that include a global evaluation of the colorimetric erythemal index (CEI) and the TEWL differential, both expressed in the same order of magnitude.

In the same study, corneosurfametry showed less interindividual variability than in vivo testing, allowing a better discrimination among mild products. An interesting finding showed that sensitive skin is not a single condition. Goffin et al. [30] hypothesized that the response of the stratum corneum to an environmental threat might be impaired in different groups of subjects experiencing sensitive skin. Data of the corneosurfametry, performed after testing eight different house-cleaning products, showed that the overall stratum corneum reactivity, as calculated by the average values of the corneosurfametry index (CSMI) and the CIM, is significantly different ($p < 0.01$) between detergent-sensitive skin and both nonsensitive and climate/fabric sensitive skin as well.

Marcel Dekker, Inc.
270 Madison Avenue, New York, New York 10016

CONCLUSIONS

Sensitive skin represents a widespread condition of susceptibility to exogenous factors. The reason why some subjects react with subjective symptoms like itching, burning, stinging, prickling, or tingling is unclear. However, a correlation of increased reactivity in subjects with a history of dermatitis and the association of increased reactivity with skin type I has been reported. Noninvasive evaluation of sensitive skin may successfully predict individual susceptibility to cosmetic-related adverse reactions. All of the efforts in this direction appear undoubtedly important to improve tolerance to the majority of cosmetic products.

REFERENCES

1. Amin S, Engasser PG, Maibach HI. Side-effects and social aspect of cosmetology. In: Baran R, Maibach HI, eds. Textbook of Cosmetic Dermatology. 2d ed. 1998; 60:709–746.
2. Burckhardt W. Praktische und theoretische bedeutung der alkalineutralisation und alkaliresistenzproben. Arch Klin Exp Derm 1964; 219:600–603.
3. Bjornberg A. Skin Reactions to Primary Irritants in Patients with Hand Eczema. Goteborg: Isaccsons, 1968.
4. Berardesca E, Cespa M, Farinelli N, Rabbiosi G, Maibach HI. In vivo transcutaneous penetration of nicotinates and sensitive skin. Contact Dermatitis 1991; 25:35–38.
5. Frosch PJ, Kligman AM. A method for appraising the stinging capacity of topically applied substances. J Soc Cosmet Chemist 1977; 28:197–209.
6. Agrup G. Hand eczema and other hand dermatoses in South Sweden. Academic dissertation. Acta Dermato Venereol 1969; 49(suppl. 161).
7. Fregert S. Occupational dermatitis in 10 years material. Contact Dermatitis 1975; 1:96–107.
8. Bjornberg A. Skin reactions to primary irritants in men and women. Acta Dermatol Venereol 1975; 55:191–194.
9. Lammintausta K, Maibach HI, Wilson D. Irritant reactivity in males and females. Contact Dermatitis 1987; 17:276–280.
10. Frosch P, Wissing C. Cutaneous sensivity to ultraviolet light and chemical irritants. Arch Derm Res 1982; 272:269–278.
11. Lammintausta K, Maibach HI, Wilson D. Susceptibility to cumulative and acute irritant dermatitis. An experimental approach in human volunteers. Contact Dermatitis 1988; 19:84–90.
12. Bettley FR. Non specific irritant reactions in eczematous subjects. Br J Dermatol 1964; 76: 116–121.
13. Roper SS, Jones EH. An animal model for altering the irritability threshold of normal skin. Contact Dermatitis 1985; 13:91–97.
14. Lammintausta K, Maibach HI, Wilson D. Human cutaneous irritation: induced hyporeactivity. Contact Dermatitis 1987; 17:193–198.
15. Nilzen A, Voss Lagerlund K. Epicutaneous tests with detergents and a number of other common allergens. Dermatologica 1962; 124:42–52.
16. Bettley FR, Donoghue E. The irritant effect of soap upon the normal skin. Br J Dermatol 1960; 72:67–76.
17. Coenraads PJ, Bleumink E, Nofer JP. Susceptibility to primary irritants. Age dependence. Contact Dermatitis 1975; 1:377–381.
18. Grove GL. Age-associated changes in integumental reactivity. In: Léveque JL, Agache PG, eds. Aging Skin. Properties and Functional Changes. New York: XX, 1993; 16:227–237.
19. Gilchrest BA, Stoff JS, Soter NA. Chronologic aging alters the response to ultraviolet-induced inflammation in human skin. J Invest Dermatol 1982; 79:11–15.

Marcel Dekker, Inc.
270 Madison Avenue, New York, New York 10016

20. Seidenari S, Francomano M, Mantovani L. Baseline biophysical parameters in subjects with sensitive skin. Contact Dermatitis 1998; 38:311–315.
21. Yosipovitch G, Yarnitsky D. Quantitative sensory testing. In: Maibach HI, Marzulli FN, eds. Dermotoxicology Methods: The Laboratory Worker's Vade Mecum. New York: Taylor & Francis, 1997; pp 315–320.
22. Lammintausta K, Maibach HI, Wilson D. Mechanisms of subjective (sensory) irritation: propensity of nonimmunologic contact urticaria and objective irritation in stingers. Dermatosen in Beruf und Umwelt 1988; 36(2):45–49.
23. Schmelz M, Schmidt R, Bichel A, Handwerker HO, Torebjörk HE. Specific C-receptors for itch in human skin. J Neurosci 1997; 17(20):8003–8008.
24. Yosipovitch G, Ademola J, Ping Lui, Amin S, Maibach HI. Topically applied aspirin rapidly decreases histamine-induced itch. Acta Demato Venereol (Stockh) 1977; 77:46–48.
25. Lovell CR, Burton PA, Duncan EH, Burton JL. Prostaglandins and pruritus. Br J Dermatol 1976; 94:273–275.
26. Lahti A, Maibach HI. Species specificity of nonimmmunologic contact urticaria: guinea pig, rat and mouse. J Am Acad 1985; 13:66–69.
27. Hannuksela A, Hannuksela M. Irritant effects of a detergent in wash and chamber tests. Contact Dermatitis 1995; 32:163–166.
28. Piérard GE, Goffin V, Piérard Franchimont C. Corneosurfametry: a predictive assessment of the interaction of personal care cleansing products with human stratum corneum. Dermatology 1994; 189:152–156.
29. Piérard GE, Goffin V, Hermanns-Le T, Arrese JE, Piérard Franchimont C. Surfactant-induced dermatitis: comparison of corneosulfametry with predictive testing on human and reconstructed skin. J Am Acad Dermatol 1995; 33:462–469.
30. Goffin V, Piérard Franchimont C, Piérard GE. Sensitive skin and stratum corneum reactivity to household cleaning products. Contact Dermatitis 1996; 34:81–85.

Marcel Dekker, Inc.
270 Madison Avenue, New York, New York 10016

Marcel Dekker, Inc.
270 Madison Avenue, New York, New York 10016

67

Tests for Skin Hydration

Bernard Gabard
Spirig Pharma Ltd., Egerkingen, Switzerland

INTRODUCTION

Writing about skin hydration means simultaneously writing about dry skin and its treatment by moisturizers. Dry skin has never really been defined in a repeatable way. In fact, this expression prejudices into believing that the skin does have a reduced water content, although this was never confirmed or denied. Generally speaking, dry skin signifies that the skin surface looks as though it is lacking in water, this being reinforced by the pharmacological effect of hydrating the skin surface by appropriate treatments.

Experimental models used for measuring skin hydration are basically clinical models incorporating or not noninvasive bioengineering measurements. To ensure meaningful results, the outlines of the intended studies should be of modern design incorporating blinding, randomization, and a suitable statistical control (particularly if different products are to be compared). This last point means including a predetermined adequate number of subjects in the study. The general ethical and legal frames of such clinical studies required for claim support are well defined in corresponding monographs or publications covering extensively the general procedures to be followed and the prerequisite information needed about the products to be tested [1–3].

Regardless of the method used, a further important point concerns standardization of the experimental conditions. To obtain acceptable and reproducible results, measurements should be performed with relaxed patients and/or volunteers already acclimatized for at least 20 minutes to controlled ambient temperature and relative humidity conditions. Both factors mainly affect sweat-gland activity, but other parameters should equally be considered with attention to, e.g., anatomical skin site, test products remaining or not on the skin, and correct handling of the measuring equipment if any. All these possible influences on measurement outcome have been discussed in detail in recent guidelines and in pertinent reviews [4–6].

A CLINICAL EVALUATION: THE REGRESSION METHOD

The dermatologist is perfectly able to clinically grade a given state of skin dryness (e.g., surface roughness, squames, and fissures). Clinical evaluation and grading of skin hydration is based on visual and tactile evaluation of clinical signs. There are numerous possibil-

Marcel Dekker, Inc.
270 Madison Avenue, New York, New York 10016

ities of testing, but basically they rely on the regression method, published in 1978 by Kligman [7], which is still used as an industry standard. Briefly, female subjects with moderate to severe xerosis of the legs are selected following strict criteria. The test products are applied under controlled conditions by trained employees twice daily 5 days a week for 3 weeks. Three days after treatment ends, the follow-up period begins. Scoring is also completed 3 and 7 days later. Treatment period may be shortened to 2 weeks if necessary. A recent guideline ensures that clinical scoring of the hydration state of the skin surface will be conducted based on the same definitions [4]. Caution is given upon scoring by the subjects themselves, as their perception of their skin condition may not be the same as the dermatologist's [4,8].

INCORPORATING BIOENGINEERING METHODS

A large number of bioengineering methods are now available to evaluate hydration (or dryness) of the skin directly or indirectly. Inclusion of these methods in the study protocol opens many possibilities for getting meaningful results such as design variations, optimization of the claim support, and also, most importantly, improvement of cost effectiveness by shortening the duration of experiment, using a lower number of subjects, and strengthening the statistical evaluation.

Concerning the numerous techniques available for the evaluation of skin hydration, the reader is referred to very recent monographs describing these methods in a detailed fashion [8–13]. They mainly include measurements of electrical properties, spectroscopic methods such as infrared absorption spectroscopy and emission, evaluation of the barrier function of the stratum corneum (SC), measurement of mechanical properties, nuclear magnetic resonance imaging, skin-surface topography, and scaling evaluation. However, in this short review, examples of possible designs will be given that use bioengineering techniques based only on the electrical properties of the SC or on measurement of transepidermal water loss (TEWL) (for a review of modern suitable measuring equipment see Refs. 8, 12, 13).

Static Measurements

Short-Term Tests/Single Application

The tests are conducted on the forearm of healthy subjects and allow a randomized side-to-side comparison of test products with a placebo or vehicle, a known active product, and untreated control skin. Four to six products may be simultaneously tested. The products are applied at the rate of 2 mg/cm^2. Two different experimental designs may be used.

1. The test products are left in place for 1 hour (or another suitable duration, e.g., 3 h [14]). Measurements are conducted at different times thereafter. Removal of excess or nonpenetrated product is preferable before measuring, especially if the preparation contains a high proportion of lipids. Most moisturizers show a rapid increase of measured hydration values (Fig. 1).
2. The test products may be applied on similar areas at the same rate but under occlusion with a standard occluding patch overnight for 16 hours. The next morning, measurements are conducted in the same way as in part 1 beginning 1 hour after removal of the occlusion patch (Fig. 2). This last procedure better picks up the activity of a humectant contained in the test preparation, whereas

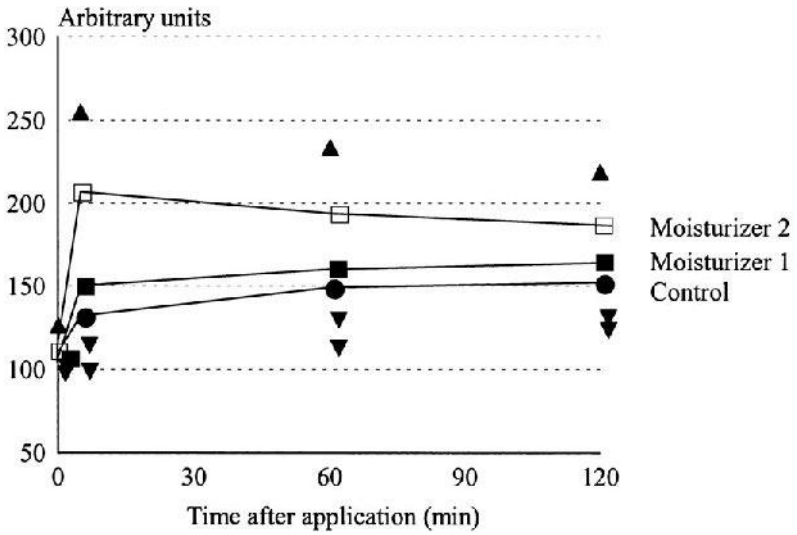


FIGURE 1 Example of hydration changes after 1 h application of two different o/w moisturizers containing both 2% urea as humectant. Hydration evaluation: NOVA DPM 9003; Means + or − half SD: ▲ ▼; Moisturizer 1: ■; Moisturizer 2: □; Control (untreated skin): ○. Start values (Time 0) measured before application of the products.

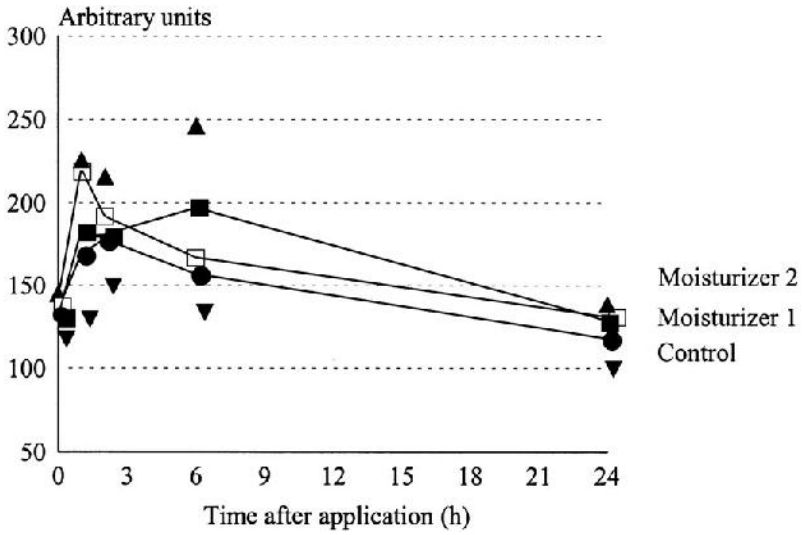


FIGURE 2 Example of hydration changes after 16 h occlusive application of two different o/w moisturizers containing both 2% urea as humectant (same products as in Figure 1). Start values (Time 0) measured before the occlusive application of the products. (For further details, see legend of Figure 1.)

the vehicle effect is strongly attenuated by the uniform conditions encountered under the occlusion patch.

Long-Term Tests/Multiple Applications

The design of these tests and selection of subjects is similar to the regression method previously described but with a modified and shortened regression protocol [15]. The treatment period extends over 1 week only, and the regression phase takes place over the following week. Bioengineering measurements are conducted 12 to 16 hours after the treatment or moisturizer application, and for the last time on the Monday following the regression week. Inclusion of these noninvasive measurements allowed rapid and reliable product-performance evaluation.

Dynamic Measurements

These tests, in addition to the classic evaluation of skin hydration, provide information on some dynamic properties of the SC [16–18]. These properties are likely to be modified by the humectants (e.g., glycerol, urea, and alpha-hydroxy acids) incorporated in the moisturizers used for treatment. Generally speaking, dynamic function tests are characterized by the assessment of the skin response to a given external stimulus that can be of physical (e.g., water, occlusion, stretch, heat) or chemical (e.g., drugs, irritants) nature. These dynamic tests may be used either during short-term or long-term product testing, and will usually be performed before and at different time points after treatment.

The Sorption-Desorption Test (SDT)

This test gives information about the water-binding capacity of the uppermost layers of the SC [16,18]. It is best conducted using measurement devices that are able to measure hydration on a wet surface and that give instantaneous readings on contact with the skin. This first value represents the hydration state of the SC. Then 50 µl of distilled water are pipetted onto the skin, left in place for exactly 10 seconds, wiped with a soft paper towel, and then hydration is immediately measured. This value represents the hygroscopicity of the superficial SC. Further measurements are taken at 0.5, 1, 1.5, and 2 minutes. The area under the curve from 0.5 minutes onwards represents the water-holding capacity of the superficial SC (Fig. 3).

The Moisture-Accumulation Test (MAT)

This test gives information about the quantity of moisture the SC may accumulate during a given time [17,18]. This test is conducted with a device able to measure continuously after bringing the probe in contact with the skin surface. The probe then remains on the skin for 3 minutes, thereby creating occlusive conditions. The MAT measures the accumulation of water under the probe every 0.5 minutes. Water accumulation is evaluated by calculating the area under the time curve until 3 minutes (Fig. 4).

The Plastic Occlusion Stress Test (POST)

The POST may also be considered a dynamic test and gives information about SC hydration, integrity of the barrier function, and SC water-holding capacity [19,20]. It consists of occluding the skin with a plastic chamber (e.g., Hilltop chamber or a similar occlusive device) for 24 hours. Then the occlusion is removed and the evaporation of the accumulated water is measured each minute for 30 minutes as TEWL. This technique has been

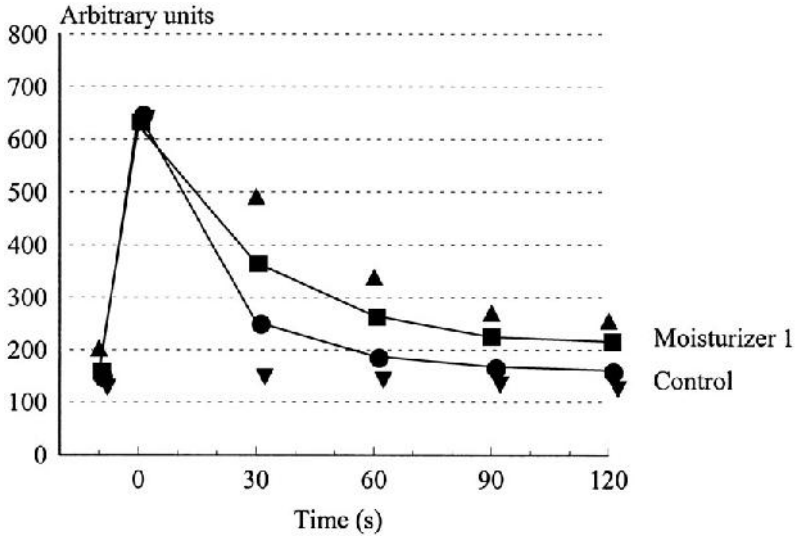


FIGURE 3 Time course of hydration changes during a sorption-desorption test (SDT) performed 60 min after a single 1 h short-term application of moisturizer 1. (For further details, see legend of Figure 1.)

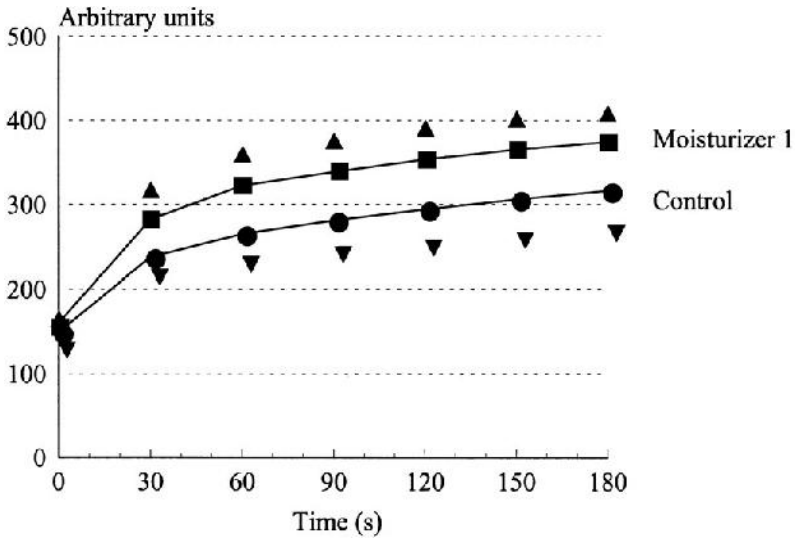


FIGURE 4 Time course of hydration changes during a moisture accumulation test (MAT) performed 60 min after a single 1 h short-term application of moisturizer 1. (For further details, see legend of Figure 1.)

MARCEL DEKKER, INC.
270 Madison Avenue, New York, New York 10016

thoroughly described in recent guidelines [21,22]. The measurement is called skin surface water loss (SSWL) and not TEWL, because it does not represent the true TEWL but the sum of the TEWL and the evaporation of water trapped within and over the SC under the occlusive equipment, at least at the beginning of the measurement period. During these first minutes of evaporation, the SSWL is proportional to SC hydration. At the end of the dehydration time, SSWL is greatly reduced and mainly TEWL is measured. Therefore, changes induced in the last part of the curve reflects the barrier function of SC.

Other Suitable Tests

Some well-defined properties of the skin are more or less dependent on SC hydration and may be evaluated with the following bioengineering methods:

- Mechanical or viscoelastic properties (elasticity, extensibility) [23]
- Skin-surface roughness [24]
- Skin-surface scaling [25]

Some other techniques are also indicated for evaluating SC hydration, but they are not available for routine experimentation at the present moment. They have been critically reviewed and evaluated in a recent publication to which the reader is referred [8].

CONCLUSION

During the evaluation of SC hydration in vivo, it must be kept in mind that no absolute determination of a water content or concentration is possible. This holds for clinical evaluation and for bioengineering measurements as well. For this reason, several measurement techniques should be used simultaneously during a study. Not only is the information gained from these different experimental approaches complementary, and of great benefit if they are integrated in a clinical evaluation, but one should remember that moisturizers may influence skin hydration in different ways. Thus, different aspects of hydration changes need to be investigated, such as water binding, water retention, or emolliency, which is also a further part of a moisturizer's action. Last, it should be remembered that, in order to obtain meaningful results, proper design of the study, inclusion of a suitable number of subjects, strict standardization of measurement conditions, and all other relevant factors need to be tightly controlled. Only by assuring the best quality level will results be obtained that will help to design and use optimal moisturizers.

REFERENCES

1. Seidenschnur EK. FDA and EEC regulations related to skin: documentation and measuring devices. In: Serup J, Jemec GBE, eds. Handbook of Non-Invasive Methods and the Skin. Boca Raton: CRC Press, 1995:653–665.
2. COLIPA (The European Cosmetic, Toiletry and Perfumery Association). Guidelines for the evaluation of the efficacy of cosmetic products, 1997.
3. Davis JB, McNamara SH. Regulatory aspects of cosmetic claims substantiation. In: Aust LB, ed. Cosmetic Claims Substantiation. New York: Marcel Dekker, 1998:1–20.
4. Serup J. EEMCO guidance for the assessment of dry skin (xerosis) and ichthyosis: clinical scoring systems. Skin Res Technol 1995; 1:109–114.
5. Berardesca E. EEMCO guidance for the assessment of stratum corneum hydration: electrical methods. Skin Res Technol 1997; 3:126–132.

Marcel Dekker, Inc.
270 Madison Avenue, New York, New York 10016

6. Wilhelm KP. Possible pitfalls in hydration measurements. In: Elsner P, Barel AO, Berardesca E, Gabard B, Serup J, eds. Skin Bioengineering: Techniques and Applications in Dermatology and Cosmetology. Current Problems in Dermatology, Vol. 26. Basel: Karger, 1998:223–234.
7. Kligman AM. Regression method for assessing the efficacy of moisturizers. Cosmet & Toilet 1978; 93:27–35.
8. Barel AO, Clarys P, Gabard B. In vivo evaluation of the hydration state of the skin: measurements and methods for claim support. In: Elsner P, Merk HF, Maibach HI, eds. In: Cosmetics: Controlled Efficacy Studies and Regulations. Berlin: Springer, 1999:57–80.
9. Tagami H, Ohi M, Iwatsuki K, Kanamaru Y, Yamada M, Ichijo B. Evaluation of the skin surface hydration in vivo by electrical measurements. J Invest Dermatol 1980; 75:500–507.
10. Levêque JL, De Rigal J. Impedance methods for studying skin moisturization. J Soc Cosmet Chem 1983; 34:419–428.
11. Levêque JL. Cutaneous Investigation in Health and Disease: Noninvasive Methods and Instrumentation. New York: Marcel Dekker, 1989.
12. Elsner P, Berardesca E, Maibach HI. Bioengineering of the Skin: Water and the Stratum Corneum. Boca Raton: CRC Press, 1994.
13. Serup J, Jemec GBE. (1995) Handbook of Non-Invasive Methods and the Skin. Boca Raton: CRC Press, 1995:159–170.
14. Serup J. A three-hour test for rapid comparison of effects of moisturizers and active constituents (urea). Acta Derm Venereol (Stockh) 1992; (suppl. 177):29–33.
15. Grove G. Skin surface hydration changes during a mini-regression test as measured in vivo by electrical conductivity. Curr Therap Res 1992; 52:1–6.
16. Tagami H, Kanamaru Y, Inoue K, Suehisa S, Inoue F, Iwatsuki K, Yoshikuni K, Yamada M. Water sorption-desorption test of the skin in vivo for functional assessment of the stratum corneum. J Invest Dermatol 1982; 78:425–428.
17. Van Neste D. In vivo evaluation of unbound water accumulation in stratum corneum. Dermatologica 1990; 181:197–201.
18. Treffel P, Gabard B. Stratum corneum dynamic function measurements after moisturizer or irritant application. Arch Dermatol Res 1995; 287:474–479.
19. Berardesca E, Maibach HI. Effect of nonvisible damage on the water-holding capacity of the stratum corneum, utilizing the plastic occlusion stress test (POST). In: Frosch PJ, Doom-Goossens A, Lachapelle JM, Rycroft RJG, Scheper RJ, eds. Current Topics in Contact Dermatitis. Berlin: Springer, 1989:554–559.
20. Berardesca E, Elsner P. Dynamic measurements: the plastic occlusion stress test (POST) and the moisture accumulation test (MAT). In: Elsner P, Berardesca E, Maibach HI, eds. Bioengineering of the Skin: Water and the Stratum Corneum. Boca Raton: CRC Press, 1994:97–102.
21. Pinnagoda J. Hardware and measuring principles: evaporimeter. In: Elsner P, Berardesca E, Maibach HI, eds. Bioengineering of the Skin: Water and the Stratum Corneum. Boca Raton: CRC Press, 1994:51–58.
22. Pinnagoda J, Tupker RA, Agner T, Serup J. Guidelines for transepidermal water loss measurement: a report from the standardization group of the European Society of Contact Dermatitis. Contact Dermatitis 1990; 22:164–168.
23. Barel AO, Lambrecht R, Clarys P. Mechanical function of the skin: state of the art. In: Elsner P, Barel AO, Berardesca E, Gabard B, Serup J, eds. Skin Bioengineering: Techniques and Applications in Dermatology and Cosmetology. Current Problems in Dermatology, Vol. 26. Basel: Karger, 1998:69–83.
24. Marks R. How to measure the effects of emollients. J Dermatol Treat 1997; 8:S15–S18.
25. Schatz H, Altmeyer PJ. Dry skin and scaling evaluated by D-Squames and image analysis. In: Serup J, Jemec GBE, eds. Handbook of Non-Invasive Methods and the Skin. Boca Raton: CRC Press, 1995:153–157.

Marcel Dekker, Inc.
270 Madison Avenue, New York, New York 10016

Marcel Dekker, Inc.
270 Madison Avenue, New York, New York 10016

68

Tests for Skin Protection: Barrier Effect

Hongbo Zhai and Howard I. Maibach
University of California at San Francisco School of Medicine, San Francisco, California

Barrier creams (BC) may play an important role in the prevention of contact dermatitis [1–6], and various in vitro and in vivo methods have been developed to evaluate their efficacy. In practice, their use remains the subject of lively debate; some reports suggest that inappropriate BC application may exacerbate, rather than prevent, irritation [1–3, 6–9]. The accuracy of measurements depends on the use of proper methodology. This chapter reviews the investigative details of pertinent scientific literature, and summarizes the methodology and efficacy of BC.

IN VITRO METHODS

In 1946, Sadler and Marriott [10] introduced some facile tests to evaluate the efficiency of BC. One method used the fluorescence of a dyestuff and eosin as an indicator to measure penetration and the rates of penetration of water through BC; this is rapid and simple, but provides only a qualitative estimate.

Suskind [11] used a simple method to measure the relative efficacy or repellency of several formulations with film-immersion tests in a specific exposure. Results showed two formulations (containing 52.5% silicone in bentonite and 30% silicone in petrolatum) were both effective against a range of aqueous irritants or sensitizers.

Langford [12] conducted in vitro studies to determine the efficacy of the formulated fluorochemical (FC)–resin complex included against solvent penetration through treated filter paper, solvent repellency on treated pigskin, and penetration of radio-tagged sodium lauryl sulfate through treated hairless mouse skin. He also conducted an in vivo study on 75 persons who had all previously experienced irritation on their hands because of continued contact with solvents. Eighty-three percent of the panelists stated the cream was effective in protecting their hands.

Reiner et al. [13] examined the protective effect of ointments both on guinea pig skin in vitro and on guinea pigs in vivo. The permeation values of a toxic agent through unprotected and protected skin within 10 h as a function of time was determined radiologically and enzymatically. Permeation of the toxic agent was markedly reduced by polyethylene-glycol ointment base and ointments containing active substances. In in vivo experi-

Marcel Dekker, Inc.
270 Madison Avenue, New York, New York 10016

ments on guinea pigs, mortality was greater after applying the toxic agent to unprotected skin. All formulations with nucleophilic substances markedly reduced the mortality rate.

Loden [14] evaluated the effect of BC on the absorption of (^{3}H)-water, (^{14}C)-benzene, and (^{14}C)-formaldehyde into excised human skin. The control and BC-treated skins were exposed to the test substance for 0.5 hours, whereupon absorption was determined. The experimental cream ''water barrier'' reduced the absorption of water and benzene but not formaldehyde. One cream slightly reduced benzene and formaldehyde absorption. The other two creams did not affect the absorption of any of the substances studied.

Treffel et al. [15] measured in vitro on human skin the effectiveness of BC against three dyes (eosin, methylviolet, and oil red O) with varying n-octanol/water partition coefficients (0.19, 29.8, and 165, respectively). BC efficacy was assayed by measurements of the dyes in the epidermis of protected skin samples after 30 minutes of application. The efficacy of BC against the three dyes showed in several cases data contrary to manufacturer's information. There was no correlation between the galenic parameters of the assayed products and the protection level, indicating that neither water content nor consistency of the formulations influenced the protection effectiveness.

Fullerton and Menne [16] tested that the protective effect of various ethylenediaminetetraacetate (EDTA) barrier gels against nickel contact allergy using in vitro and in vivo methods. In an in vitro study, about 30 mg of barrier gel were applied on the epidermal side of the skin and a nickel disc was applied above the gel. After 24-hours application, the nickel disc was removed and the epidermis separated from the dermis. Nickel content in epidermis and dermis was quantified by absorption differential pulse voltammetry (ADPV). The amount of nickel in the epidermal skin layer on barrier gel–treated skin was significantly reduced compared with the untreated control. In vivo patch testing of nickel-sensitive patients was performed using nickel discs with and without barrier gels. Test preparations and nickel discs were removed 1 day after application, and the test sites were evaluated. Reduction in positive test reactions was highly significant on barrier gel–treated sites.

Zhai et al. [17] used an in vitro diffusion system to measure the protective efficacy of Quaternium-18 bentonite (Q18B) gels to prevent 1% concentration of [^{35}S] sodium lauryl sulfate (SLS) penetration on human cadaver skin. The accumulated amount of [^{35}S]-SLS in receptor-cell fluid were counted to evaluate the efficacy of the Q-18B gels over a 24-hour period. These test gels significantly decreased SLS absorption when compared with unprotected-skin control samples. The percent protection effect of three test gels against SLS percutaneous absorption was 88%, 81%, and 65%, respectively.

IN VIVO METHODS

In 1940, Schwartz et al. [18] introduced an in vivo method to evaluate the efficacy of a vanishing cream against poison ivy extract using visual erythema on human skin. The test cream was an effective prophylaxis against poison ivy dermatitis when compared with unprotected skin.

Lupulescu and Birmingham [19] observed the ultrastructural and relief changes of human epidermis after exposure to a protective gel, acetone, and kerosene on humans. Unprotected skin produced cell damage and a disorganized pattern in the upper layers of epidermis. Application of a protective agent before to solvent exposure substantially reduced the ultrastructural and relief changes of epidermis cells.

Lachapelle and co-workers [3, 20–23] used a guinea pig model to evaluate the pro-

tective value of BC and/or gels by laser Doppler flowmetry and histological assessment. The histopathological damage after 10 minutes of contact to toluene was mostly confined to the epidermis, whereas the dermis was almost normal. The dermal blood-flow changes were relatively high on the control site compared with the gel-pretreated sites.

Frosch et al. [1, 8, 9, 24, 25] developed the repetitive irritation test (RIT) in the guinea pig and in humans to evaluate the efficacy of BC by using a series of bioengineering techniques. The cream-pretreated and -untreated test skin (guinea pig or humans) was exposed daily to the irritants for 2 weeks. The resulting irritation was scored on a clinical scale and assessed by biophysical techniques' parameters. Some test creams suppressed irritation with all test parameters, some failed to show such an effect, and some even exacerbated the irritation [9].

Zhai [2] used an in vivo human model to measure the effectiveness of BC against dye-indicator solutions: methylene blue in water and oil red O in ethanol, which are representative of model hydrophilic and lipophilic compounds. Solutions of 5% methylene blue and 5% oil red O were applied to untreated and BC-pretreated skin with the aid of aluminum occlusive chambers for 0 and 4 hours. At the end of the application time, the materials were removed, and consecutive skin-surface biopsies (SSB) obtained. The amount of dye penetrating into each strip was determined by colorimetry. Two creams exhibited effectiveness, but one cream enhanced the cumulative amount of dye.

Zhai et al. [5] introduced a facile approach to screening protectants in vivo in human subjects. Two acute irritants and 1 allergen were selected: 1) sodium lauryl sulfate (SLS), representative of irritant household and occupational contact dermatitis, 2) the combination of ammonium hydroxide (NH_4OH) and urea to simulate diaper dermatitis, and 3) Rhus to evaluate the effect of model protective materials. Test materials were spread onto test area, massaged, allowed to dry for 30 minutes, and reapplied with another 30-minute drying period. The model irritants and allergen were applied with an occlusive patch for 24 hours. Inflammation was scored with an expanded 10-point scale at 72 hours after application. Most test materials statistically suppressed the SLS irritation and Rhus allergic reaction rather than NH_4OH and urea-induced irritation.

Wigger-Alberti et al. [26] determined which areas of the hands were likely to be skipped on self-application of BC by fluorescence technique at the workplace. Results showed the application of BC was incomplete, especially on the dorsal aspects of the hands. Brief data of recent experiments of BC are summarized in Table 1.

CONCLUSIONS

Some BC reduce CD under experimental conditions. But, inappropriate BC application may enhance irritation rather than benefit. To achieve the optimal protective effects, BC should be used with careful consideration based on specific exposure conditions; also, the proper use of BC should be instructed.

In vitro methods are simple, rapid, safe, and are recommended in the screening procedure for BC candidates. With radiolabeled methods, we may determine the accurate protective and penetration results even in the lower levels of chemicals because of the sensitive radiolabeled counting when BCs are to be evaluated. Animal experiments may be used to generate kinetic data because of a closer similarity between humans and some animals (e.g., pigs and monkeys) in percutaneous absorption and penetration for some compounds. But no one animal, with its complex anatomy and biology, will simulate the penetration in humans for all compounds. Therefore, the best estimate of human percutane-

Marcel Dekker, Inc.
270 Madison Avenue, New York, New York 10016

TABLE 1 Brief Data from Recent Experiments of BC

Models						
In vitro	In vivo animals or humans	Irritants or allergens	Barrier creams	Evaluations by	Efficacy	Authors and Refs.
Human skin		Dyes (eosin, methylviolet, oil red O)	16 Barrier creams	Amount of dyes in the epidermis	Various % protection effects	Treffel et al. [15]
Human skin	Nickel-sensitive patients	Nickel disc	Ethylenediaminetetraacetate (EDTA) gels	Nickel content	Significantly reduced the amount of nickel in the epidermis in vitro, and significantly reduced positive reactions in vivo	Fullerton and Menne [16]
Human skin		[^{35}S]-SLS	3 Quaternium-18 bentonite (Q-18B) gels	Amount of [^{35}S]-SLS	% protection effect was 88%, 81%, and 65%, respectively	Zhai et al. [17]
	Guinea pigs	n-Hexane, trichlorethylene, toluene	3 water-miscible crams	Morphological assessment	Limited protective effects	Lachapelle et al. [23]
	Guinea pigs and humans	SLS, sodium hydroxide, toluene, lactic acid	Several barrier creams	Various bioengineering techniques	Some of them suppressed irritation, some failed	Frosch et al. [1, 8, 24, 25]
	Humans	Dyes (methylene blue and oil red O)	Three barrier creams	Amount of dye penetrating into strips	Two of them exhibited effectiveness, one enhanced cumulative amount of dye	Zhai and Maibach [2]
	Humans	SLS, ammonium hydroxide (NH^4OH) and urea, Rhus	Several protectants	Clinical scores	Most suppressed the SLS irritation and Rhus allergic reaction, failed to NH^4OH and urea irritation	Zhai et al. [5]
	Humans	Self-application of BC	An oil-in-water emulsion	Fluorescence technique	Self-application of BC was incomplete	Wigger-Alberti et al. [26]

MARCEL DEKKER, INC.
270 Madison Avenue, New York, New York 10016

ous absorption is determined by in vivo studies in humans. The histological assessments may define what layers of skin are damaged or protected, and may provide the insight mechanism of BC. Noninvasive bioengineering techniques may provide accurate, highly reproducible, and objective observations in quantifying the inflammation response to various irritants and allergens when BC are to be evaluated that could assess subtle differences to supplement traditional clinical studies.

To validate these models, well-controlled field trials are required to define the relationship of the model to the occupational setting. Finally, the clinical efficacy of BC should be assessed in the workplace rather than in experimental circumstances.

REFERENCES

1. Frosch PJ, Schulze-Dirks A, Hoffmann M, Axthelm I, Kurte A. Efficacy of skin barrier creams. (I). The repetitive irritation test (RIT) in the guinea pig. *Contact Dermatitis 28*:94, 193.
2. Zhai H, Maibach, HI. Effect of barrier creams: human skin in vivo. *Contact Dermatitis 35*: 92, 1996.
3. Lachapelle JM. Efficacy of protective creams and/or gels. In: Elsner P, Lachapelle JM, Wahlberg JE, Maibach HI, eds. *Prevention of Contact Dermatitis, Current Problems in Dermatology*. Basel: Karger, 1996; 182.
4. Zhai H, Maibach HI. Percutaneous penetration (Dermatopharmacokinetics) in evaluating barrier creams. In: Elsner P, Lachapelle JM, Wahlberg JE, Maibach HI. eds. *Prevention of Contact Dermatitis, Current Problems in Dermatology*. Basel: Karger, 1996; 193.
5. Zhai H, Willard P, Maibach HI. Evaluating skin-protective materials against contact irritants and allergens. An in vivo screening human model. *Contact Dermatitis 38*:155, 1998.
6. Wigger-Alberti W, Elsner P. Do barrier creams and gloves prevent or provoke contact dermatitis? *Am J Contact Dermatitis 9*:100, 1998.
7. Goh CL. Cutting oil dermatitis on guinea pig skin. (I). Cutting oil dermatitis and barrier cream. *Contact Dermatitis 24*:16, 1991.
8. Frosch PJ, Schulze-Dirks A, Hoffmann M, Axthelm I. Efficacy of skin barrier creams. (II). Ineffectiveness of a popular ''skin protector'' against various irritants in the repetitive irritation test in the guinea pig. *Contact Dermatitis 29*:74, 1993.
9. Frosch PJ, Kurte A, Pilz B. Biophysical techniques for the evaluation of skin protective creams. In: Frosch PJ, Kligman AM, eds. *Noninvasive Methods for the Quantification of Skin Functions*. Berlin: Springer-Verlag, 1993; 214.
10. Sadler CGA, Marriott RH. The evaluation of barrier creams. *Br Med J 23*:769, 1946.
11. Suskind RR. The present status of silicone protective creams. *Ind Med Surg 24*:413, 1955.
12. Langford NP. Fluorochemical resin complexes for use in solvent repellent hand creams. *Am Ind Hyg Assoc J 39*:33, 1978.
13. Reiner R, Roßmann K, Hooidonk CV, Ceulen BI, Bock J. Ointments for the protection against organophosphate poisoning. *Arzneim-Forsch/Drug Res 32*:630, 1982.
14. Loden M. The effect of 4 barrier creams on the absorption of water, benzene, and formaldehyde into excised human skin. *Contact Dermatitis 14*:292, 1986.
15. Treffel P, Gabard B, Juch R. Evaluation of barrier creams: an in vitro technique on human skin. *Acta Derm Venereol 74*:7, 1994.
16. Fullerton A, Menne T. In vitro and in vivo evaluation of the effect of barrier gels in nickel contact allergy. *Contact Dermatitis 32*:100, 1995.
17. Zhai H, Buddrus DJ, Schulz AA, Wester RC, Hartway T, Serranzana S, Maibach HI. In vitro percutaneous absorption of sodium lauryl sulfate (SLS) in human skin decreased by Quaternium-18 bentonite gels. In vitro & Molecular Toxicol 12:11, 1999.
18. Schwartz L, Warren LH, Goldman FH. Protective ointment for the prevention of poison ivy dermatitis. *Public Health Rep 55*:1327, 1940.

Marcel Dekker, Inc.
270 Madison Avenue, New York, New York 10016

19. Lupulescu AP, Birmingham DJ. Effect of protective agent against lipid-solvent–induced damages. Ultrastructural and scanning electron microscopical study of human epidermis. *Arch Environ Health 31*:29, 1976.
20. Mahmoud G, Lachapelle JM, Neste DV. Histological assessment of skin damage by irritants: its possible use in the evaluation of a 'barrier cream'. *Contact Dermatitis 11*:179, 1984.
21. Mahmoud G, Lachapelle JM. Evaluation of the protective value of an antisolvent gel by laser Doppler flowmetry and histology. *Contact Dermatitis 13*:14, 1985.
22. Mahmoud G, Lachapelle J. Uses of a guinea pig model to evaluate the protective value of barrier creams and/or gels. In: Maibach HI, Lowe NJ, eds. *Models of Dermatology*. Basel: Karger, 1987; 112.
23. Lachapelle JM, Nouaigui H, Marot L. Experimental study of the effects of a new protective cream against skin irritation provoked by the organic solvents n-hexane, trichlorethylene and toluene. *Dermatosen 38*:19, 1990.
24. Frosch PJ, Kurte A, Pilz B. Efficacy of skin barrier creams. (III). The repetitive irritation test (RIT) in humans. *Contact Dermatitis 29*:113, 1993.
25. Frosch PJ, Kurte A. Efficacy of skin barrier creams. (IV). The repetitive irritation test (RIT) with a set of 4 standard irritants. *Contact Dermatitis 31*:161, 1994.
26. Wigger-Alberti W, Maraffio B, Wernli M, Elsner P. Self-application of a protective cream. Pitfalls of occupational skin protection. *Arch Dermatol 133*:861, 1997.

Marcel Dekker, Inc.
270 Madison Avenue, New York, New York 10016

69

Objective Methods for Assessment of Human Facial Wrinkles

Gary Grove and Mary Jo Grove
KGL's Skin Study Center, Broomall, and cyberDERM, inc., Media, Pennsylvania

INTRODUCTION

The skin, especially that of the face, undergoes very characteristic changes with advancing age. Although there are other overt morphological changes that can be considered as markers of cutaneous aging, the degree of wrinkling in the ''crow's feet'' area seems to have the greatest impact. Thus, it is not surprising that considerable effort has been expanded to develop skincare products and cosmetic surgical procedures that can effectively restore a more youthful appearance.

Wrinkles can be easily visualized and many clinical studies have involved the use of ranking scales that rely on subjective assessments by expert graders. To improve the validity and reproducibility of this approach, more complex ordinal scales with semiquantive word descriptors and reference photographs have been devised by several investigators [1–3]. For example, Daniell [1] devised a set of reference photographs that illustrates his six-point grading scheme for evaluating crow's-feet wrinkles in the lateral periorbital area. Such reference photographs can be used to train inexperienced graders and periodically review the competency of all evaluators. Nevertheless, the major drawback to this type of approach is that it provides no permanent records that fully describe the skin-surface features or allow retrospective analysis. Instead, we must rely on the subjective judgements of trained graders and their ability to recall from memory the full range of changes in skin-surface features that might occur in each situation.

This problem can be overcome by taking standardized photographs before treatment and at various intervals during the treatment period. This provides a series of photographs that not only documents the study but can also be used to quantify the therapeutic response. This can be done by a panel of blinded, independent readers who are remote from the study environment as was done for photodamaged skin treated with isotretinoin [4] or alpha-hydroxy acids [5]. Although clinical assessments and photography are useful methods for assessing such changes in photodamaged skin, we are concerned that they might not be appropriate for studying wrinkles. This is especially true for photography in which changes in lighting or facial expression can greatly influence the appearance of lines and wrinkles. We are also concerned that concurrent improvements in other facial features, such as a decrease in mottled pigmentation or increased dermal blood flow, might partially

Marcel Dekker, Inc.
270 Madison Avenue, New York, New York 10016

unblind or unduly influence the investigator while judging the drug's impact on wrinkles. Thus, there is clearly a need for a more objective method to evaluate the effects of various treatments on facial lines and wrinkles.

CHAPTER OBJECTIVE

In this chapter, we will describe how optical profilometry [6] can provide an objective measure of wrinkling that can complement clinical assessment of various agents and procedures that might be useful in the therapy of photodamaged skin. Our approach is a variant of the skin-surface replica approach first described by Corcuff and colleagues [7–10]. Silicon rubber impression materials can be used to make a mold of the skin surface that faithfully captures its fine facial lines and wrinkles [11]. Such samples provide a permanent topographic record and can easily be taken serially from the same site with extended periods between each sample. By using image-analysis techniques similar to those used by NASA to map the lunar landscape during the Ranger missions [12], we can extract numeric information that describes the microtopographic features of the skin in the same way. Instead of using the sun to sidelight the moon's craters and crevices, we use a fiberoptic illuminator set at an appropriate angle to bring out the skin-surface details of interest.

BASIC METHODOLOGY

Skin-Surface Impressions

The site to be sampled should be delineated by affixing adhesive paper rings with orientation tabs such as those manufactured by CuDerm (Dallas, TX). Because the crow's feet furrows taper and become less pronounced as you move away from the periorbital area, it is extremely important that the site be located precisely. To facilitate relocating this site for subsequent serial samples, close-up photographs can be taken of the region with the adhesive rings properly placed for each panelist, as shown in Figure 1 for the crow's foot region.

Of the dental-impression materials that have been used, Silflo from Flexico Develop-

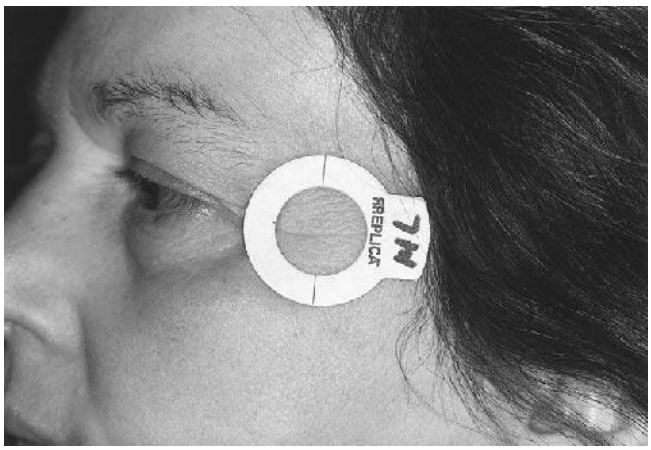


FIGURE 1 Placement of CuDerm Replica Locator Ring for obtaining silicone rubber impression from the crow's feet region.

MARCEL DEKKER, INC.
270 Madison Avenue, New York, New York 10016

ments Ltd. (Potters Bar, UK) is the best choice for skin-surface replicas. It not only offers a very high degree of fidelity, but the white opaque surface is ideal for viewing under reflected light used in image analysis. Moreover, these types of samples could be stored for at least 2 years without any fear of alteration in microtopography.

To make a impression, a thin layer of freshly prepared Silflo is gently spread over the bounded area of the ring and allowed to flow into the various furrows, creases, and fine lines that mark the surface. Within a few minutes, the material will polymerize and the replica is removed by gently lifting away from the skin using the orientation tab of the paper ring. It is important that the panelist remains calm with eyes closed and face relaxed during the polymerization phase. Because of the hydrophobic properties of silicon rubber, holes will form if the panelist is sweating from being too warm or emotionally stressed. Other artifacts such as bubbles can be created if the mixture is stirred too vigorously, causing it to froth. Alternatively, if the resin is not adequately mixed with the hardener or the mixture is allowed to partially polymerize before application, the specimens will lack detail.

Image Analysis

The general principles of image analysis for measuring the microtopography of the skin surface as captured in replica specimens have been previously described [6,7]. Briefly, these instruments consist of a high-resolution, black-and-white digital camera that is interfaced into a computer that contains specially designed image-processing hardware and software. The resulting image consists of a 640×480 pixel matrix with 256 gray levels. By selecting proper thresholds based on gray-level values, the image can be segmented into features of interest, such as wrinkles, and subsequently analyzed. One of the advantages of using replicas over photographs is that only topographic features are captured in the white replicas, which can be studied in all three dimensions by taking lighting angles into consideration. In striking contrast, not only are the photographs limited to two dimensions, but color variations attributable to mottled pigmentation greatly complicate the analysis.

In this application, the replica specimen is sidelighted using a fiberoptic illuminator set at a precisely defined angle to bring out the surface details of interest. In general, the lower the light source the greater the detail will be. In the case of a child, an angle of 15° to 20° will enable the observer to see a large number of fine lines, whereas for deeper furrows and creases such as crow's feet in an adult, an angle of 38° to 45° is optimal. In both cases, lines and wrinkles will cast shadows that are contrasted against the white background of the replica. Figure 2 illustrates that differences in the degree of wrinkling in the crow's foot region can be readily appreciated in this manner.

Because of the extreme anisotrophy of the skin, it is extremely important to take note of the position of the light source relative to the orientation of the specimen. Indeed, the major furrows and lines that are recognized as crow's feet are highly directional with 180° symmetry. For technical reasons, it is far simpler to rotate the sample than to have the lighting system revolve to simulate the movement of the sun. This is accomplished by a using a lazy Susan as a revolving sample holder, and great care is taken to ensure that the replica is held perfectly flat and centered with regard to both the light source and video camera during this movement. In the automated system of Corcuff and Leveque [10], the specimen is rotated at 9° steps through 360°, giving a series of 40 values. When plotted as polar coordinates according to the angle of rotation, one obtains a ''rose of direction.'' A min-max at 180° is readily apparent, and taking measurements in both orien-

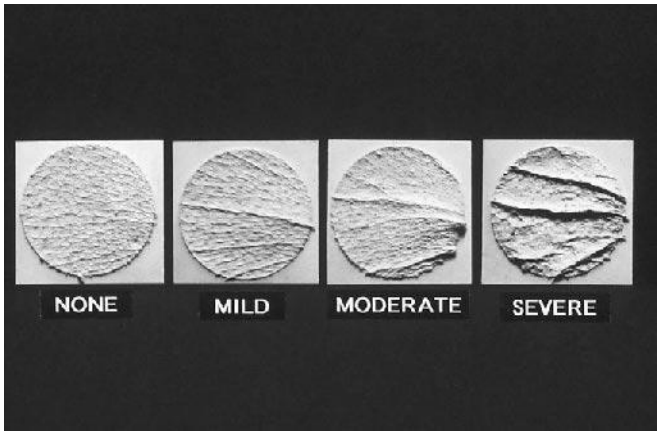


FIGURE 2 Representative silicone rubber impressions cast from the crow's foot region showing different degrees of wrinkling.

tations is sufficient for most applications. In our convention, the north-south axis is when the orientation of the lighting is perpendicular to the major furrows, whereas with the east-west it is parallel.

In addition to the angle of lighting, the uniformity of the incident light is critical. Interfering lights and changes in ambient lighting will influence the reproducibility of the measurements; it is best to work in total darkness. Fluctuations in the electronic circuits of the digitizing camera must also be minimized. As a rule, all these types of errors can be controlled by routinely using a series replicas as reference standards to ensure the analysis is being properly conducted.

The digitized image can be mathematically represented as a three-dimensional matrix of numbers. The x and y values are polar coordinates that provide the location of the pixel whereas the z value represents the brightness of the pixel in terms of its gray level. In one analytical approach, the digitized image is segmented into a binary image consisting of shadows and background by choosing an appropriate gray-level threshold. The percentage of the surface area occupied by shadows in a standard field of view is directly related to its topography. Obviously, if the surface is rather smooth and flat, there will be few shadows and this value will be small. On the other hand, if the skin is wrinkled and rough, the shadowed areas will be correspondingly larger. Moreover, because the angle of illumination is known, the horizontal projection of these shadows can be used to estimate their mean depth. Indeed, more sophisticated analyses such as the coefficient of developed skin surface (CDSS), which is a mathematical expression of true-versus-apparent surface area as pioneered by Corcuff's group [7–10], are possible.

In optical profilometry, a profile that represents the surface features at that specific location is created by plotting the gray-level values across a horizontal segment of this digitized image. This graphic display (Fig. 3) is similar to that achieved through mechanical profilometry with stylus devices, and we can extract numeric information that describes the microtopographic attributes in much the same way. Of the many parameters available for assessing skin-surface topography, both Rz and Ra have proved to be the most useful. To compute Rz, the profile is first divided into five equal segments along the x-axis. The minimum–maximum differences within each of the five segments are then determined,

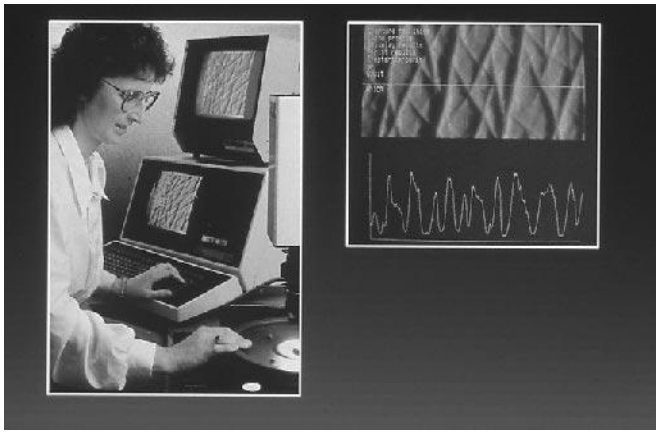


FIGURE 3 Basic set-up and profile of skin surface topography generated by image analysis of side-illuminated silicone rubber impressions.

and Rz is calculated as the average of these five local values. To compute Ra, an average line is generated to run through the center of the profile, and the area that the profile describes above and below this reference line is determined.

REPRESENTATIVE RESULTS

Photodamaged Skin with Topical Tretinoin

Computerized image analysis of silicone rubber impressions of the skin surface has been used to document the effects of topically applied tretinoin cream on photodamaged facial skin in several multicenter clinical trials [6, 12–15]. Although coarse wrinkles have been diminished, it is clear that the primary effect is on superficial, fine lines, as shown in Figure 4.

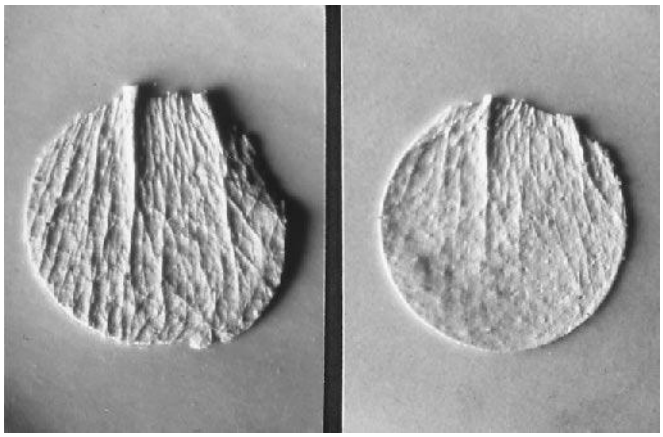


FIGURE 4 Representative silicone rubber impressions from the crow's foot region of a patient before and after 6 months of treatment with 0.05% tretinoin emollient cream.

Cutaneous Resurfacing Using High-Energy, Pulsed CO_2 Lasers

Although cutaneous resurfacing with CO_2 lasers is not a new technique, the older systems were not well suited for the delicate areas around the eyes and mouth. The newest generation of high-energy pulsed (''ultrapulse'') CO_2 lasers produces high-energy bursts that allow maximal lesional ablation with minimal heat conduction to uninvolved skin which greatly reduces the risks of scarring. Alster [16] has shown that although both the surgipulse and ultrapulse high-energy CO_2 lasers are effective in reducing the appearance of periorbital rhytides, computer analysis of skin-surface impressions shows a more substantial improvement after ultrapulse laser treatment. Indeed, the skin-surface texture was found to be comparable to normal. Alster [17] has also used optical profilometry to document that laser resurfacing can also effectively improve or even eliminate atrophic facial acne scars.

CONCLUDING REMARKS

The use of silicon rubber impressions allows one to capture the microtopographic features of the skin surface that can be subsequently measured using well-established image-analysis techniques. These samples are durable, easy to store, and can be readily transported from the clinical center to a remote site for objective measurements in a truly blinded manner. Moreover, these samples allow one to obtain serial samples of the same skin surface over prolonged periods of time. As pointed out by Leveque [18], numerous factors can alter the appearance or actual dimensions of wrinkles. It is extremely important to understand that extreme care must be taken to ensure that the impression be artifact free and truly the skin surface being studied. The before- and after-treatment comparisons must be based on identical areas and use the same conditions of analysis.

REFERENCES

1. Daniell HW. Smoker's wrinkles: a study in the epidemiology of ''crow's feet.'' Ann Intern Med 1971; 75:873–880.
2. Griffiths CEM, Wang TS, Hamilton TA, Voorhees JJ, Ellis CN. A photonumeric scale for the assessment of cutaneous photodamage. Arch Dermatol 1992; 128:347–351.
3. Larnier C, Ortonne JP, Venot A, Faivre B, Beani JC, Thomas P, Brown T, Sendagorta E. Evaluation of cutaneous photodamage using a photographic scale. Br J Dermatol 1994; 130: 167–173.
4. Armstrong RB, Lesiewicz G, Harvey G, Lee LF, Spoehr KT, Zultak M. Clinical panel assessment of photodamaged skin treated with isotretinoin using photographs. Arch Dermatol 1992; 128:352–356.
5. Stiller MJ, Bartolone J, Stern R, Smith S, Kollias N, Gillies R, Drake LA. Topical 8% glycolic acid and 8% l-lactic acid creams for treatment of photodamaged skin. Arch Dermatol 1996; 132:631–636.
6. Grove GL, Grove MJ, Leyden JJ. Optical profilometry: an objective method for quantification of facial wrinkles. J Am Acad Derm 1989; 21:631–637.
7. Corcuff P, Chatenay F, Leveque JL. A fully automated system to study skin surface patterns. Int J Cosmet Sci 1984; 6:167–176.
8. Corcuff P, deRigal J, Leveque JL, Makki S, Agache P. Skin relief and aging. J Soc Cosmet Chem 1983; 34:177–190.
9. Corcuff P, Leveque JL, Grove GL, Kligman AM. The impact of aging on the microrelief of perio-orbital and leg skin. J Soc Cosmet Chem 1987; 82:145–152.

Marcel Dekker, Inc.
270 Madison Avenue, New York, New York 10016

10. Corcuff P, Leveque JL. Skin surface replica image analysis of furrows and wrinkles. In: Scrup J and Jemec GBE, eds. *Handbook of Non-invasive Methods and the Skin.* Boca Raton: CRC, 1995:89–96.
11. Grove GL, Grove MJ. Objective methods for assessing skin surface topography noninvasively. In: Leveque JL, ed. Cutaneous Investigations in Health and Disease. New York: Marcel Dekker, 1989: 1–32.
12. Grove GL, Grove MJ, Leyden JJ, Lufrano L, Schwab B, Perry BH, Thorne EG. Skin replica analysis of photodamaged skin after therapy with tretinoin emollient cream. J Am Acad Dermatol 1991; 25:231–237.
13. Olsen EA, Katz HI, Levine N, et al. Tretinoin emollient cream: a new therapy for photodamaged skin. J Am Acad Dermatol 1992; 26:215–224.
14. Weinstein GD, Migra TP, Poch PE, et al. Topical tretinoin for treatment of photodamaged skin. A multicenter study. Arch Dermatol 1991; 127(5):659–665.
15. Gilchrest BA. Treatment of photodamage with topical tretinoin: an overview. J Am Acad Dermatol 1997; 36(?)S27- 36.
16. Alster TS. Comparison of two high-energy, pulsed carbon dioxide lasers in the treatment of periorbital rhytides. Dermatol Surg 1996; 22:541–545.
17. Alster TS, West TB. Resurfacing of atrophic facial acne scars with a high-energy, pulsed carbon dioxide laser. Dermatol Surg 1996; 22:151–155.
18. Leveque JL, EEMCO guidance for the assessment of skin topography. J Eur Acad Dermatol Venereol 1999; 12(2):103–114.

Marcel Dekker, Inc.
270 Madison Avenue, New York, New York 10016

Marcel Dekker, Inc.
270 Madison Avenue, New York, New York 10016

70

Acnegenicity and Comedogenicity Testing for Cosmetics

F. Anthony Simion
The Andrew Jergens Company, Cincinnati, Ohio

INTRODUCTION

Many people experience facial acne, especially in their teens and early 20s. It typically causes distress, and in some individuals contributes to a lowered self-image [1]. As consumers age the prevalence of acne decreases, although it can be triggered by factors such as stress, medications, and the use of cosmetics [2–4]. Indeed, there are many reports of cosmetics causing comedones, or acneform eruptions. These adverse reactions are of great concern to consumers, many of whom look for products that will not cause such problems. Hence, cosmetics manufacturers strive to develop products that do not cause comedones or acne. Products are frequently labeled noncomedogenic and/or nonacnegenic. Consumers use these terms interchangeably, although comedone formation and acnegenicity are not the same. To consumers, the key issue is that the cosmetics they use do not cause breakouts.

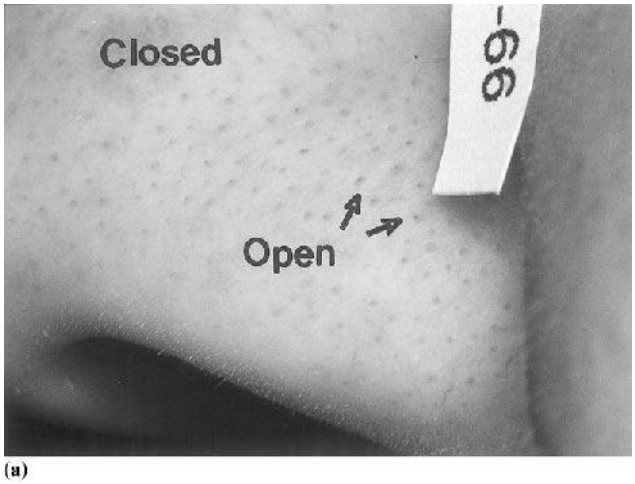
COMEDOGENICITY

Comedone formation occurs when the pattern of keratinization inside the sebaceous follicle changes. Within the keratinocytes, these changes include the production of different keratins and a reduction in the number of lamellar granules [5]. There is also an increase in mitotic activity [6]. As a result the keratinocytes do not desquamate properly and the follicular duct is blocked. It is not known what causes these changes, but the result is a microcomedone. As further keratinized material accumulates, the follicle becomes visible from the surface as a closed comedone, or whitehead (Fig. 1). As more material accumulates, the follicle distends and open comedones, also know as blackheads are formed. The black color is attributable to the oxidation of lipids as they reach the skin's surface. The test methods to assess comedogenesis are designed to quantify the hyperkeratotic plugs.

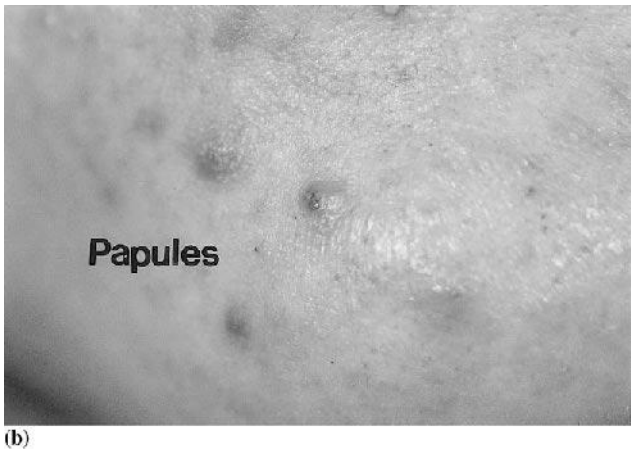
ACNEGENICITY

The hyperkeratotic plug results in sebum accumulating in the follicilar duct and the sebaceous gland. This enables the anaerobic bacteria, *P. acnes*, to proliferate. The follicular

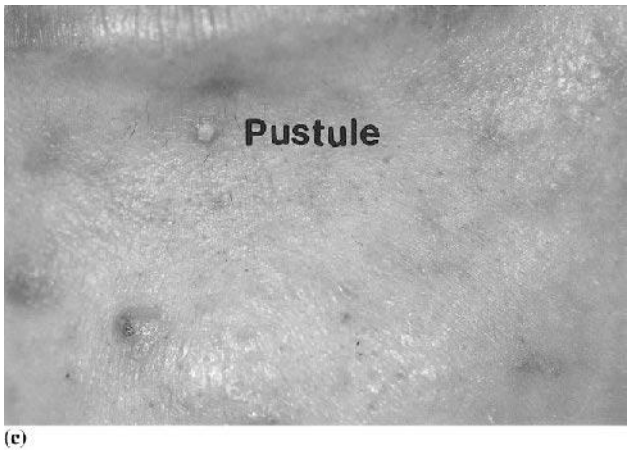
Marcel Dekker, Inc.
270 Madison Avenue, New York, New York 10016



(a)



(b)



(c)

FIGURE 1 Examples of the different forms of comedones and acneform eruptions. (a) Closed and open comedones. (b) Papules. (c) Pustules.

MARCEL DEKKER, INC.
270 Madison Avenue, New York, New York 10016

duct will expand until it ruptures, releasing the bacteria and their metabolic products into the surrounding dermis. Both immunological and irritant reactions occur. In severe acne, elevated levels of anti–*P. acnes* antibodies are detected. Irritation may come from the fatty acids that are the result of sebaceous triglyceride digestion by bacteria. Certainly at the histological level, the classic signs of inflammation such as a neutrophilic infiltrate are observed. The consumer recognizes this as acne pustules and papules.

PUSTULOGENIC POTENTIAL

Upon first using a new cosmetic, papules and pustules are sometimes observed after a few days. To the consumer this is acne, although this type of papulopustular reaction may be a form of follicular irritation. Certainly, as Mills and Berger pointed out, it occurs more quickly than can be accounted for by the formation of hyperkeratotic plugs and its sequalae [7]. Thus, Mills and Berger suggested that pustulogenic potential—the ability to cause inflammatory lesions—should be differentiated from comedogenic potential—the ability to cause the formation of hyperkeratotic plugs.

TEST METHODS

Human and animal models have been used to assess the comedogenic potential of cosmetic products and their ingredients. Both models require repeated applications of the materials to the skin for 2 to 4 weeks. The number of hyperkeratotic impactions produced is compared with positive and negative controls.

Animal Models

The rabbit ear is the most commonly used animal model. The rabbit's ear follicle has many structural similarities to the human sebaceous follicle. In 1941, Adams et al. [8] showed that the rabbit ear would respond similarly to human skin when exposed to chlorinated hydrocarbons, the most common cause of acne in industrial accidents. Until the rise of the animal rights movement in the late 1980s, cosmetic products were routinely screened using rabbits. Briefly, the method entails that the test product or ingredient is applied daily to the inner surface of one ear. This site is left open. The other ear is used as the negative control. At the end of either 2 or 4 weeks, the animal was sacrificed, and the degree of follicular hyperkeratinization is assessed. Frequently this was done by taking histological sections and giving an overall score based on the number and degrees of compacted follicles. Occasionally the impactions were removed from the ear using cyanoacrylate glue on a glass slide. This method is now frequently used in human testing.

The results from the rabbit-ear studies show that some cosmetic ingredients have comedogenic potential. These include branched chain esters and compounds that have solubility in both oil and water (hydrophile-lipophile balance (HLB) of 10–12) [9]. However, if these materials are chemically modified, or included at low levels with other ingredients in cosmetics, then their comedogenic potential is nullified. This was shown by Fulton [9] as well as Kligman and Mills [2]. Fulton showed that chemically modifying cosmetic ingredients can greatly effect their comedogenic potential. For instance, PEG-16 Lanolin gives a severe comedogenicity score of 4 on a 0 to 5 scale, whereas the higher–molecular weight and more water-soluble PEG-75 Lanolin yields a score of 0 under the

same test conditions (Table 1). Furthermore, Fulton showed that the comedogenic potential of fatty acid solutions was greatly reduced by replacing sunflower oil with acetone or ether as the solvent. Kligman and Mills reported that the comedogenic potential of vegetable oils is dose dependent, being abolished by diluting to 25% with mineral oil. From these observations, it appears vital to assess the comedogenic potential of the final product [7]. Additionally, Fulton and his colleagues screened the comedogenic potential of many ingredients and products [9,10]. Some of these results are included in Table 1.

It is interesting to note that the primary irritation potential does not correlate with comedogenic potential. For instance, sodium lauryl sulfate, which is frequently used as a model irritant, is noncomedogenic [11]. Conversely, many esters—such as isopropyl isostearate—that are highly comedogenic are relatively nonirritating.

To detect weak comedogens, the rabbit-ear assay was modified. Product applications were increased from 2 to 4 weeks, enabling the assay to detect products that cause comedones in a small but sensitive groups of consumers. The method may be overly sensitive for the average consumer, so there is a risk of false positives. Conversely, products that are noncomedogenic in the 4-week rabbit ear test are unlikely to cause comedone formation even in acne-prone consumers.

Many adverse reactions that consumers describe as breakouts or blemishes are not attributable to comedone formation. This is readily appreciated from the rapid onset of the blemish (a few days), which is too rapid for the formation of hyperkeratotic plugs in the follicular ducts. Furthermore, the formation of open or closed comedones frequently occurs without skin redness, whereas breakouts and blemishes described by consumers do have an inflammatory component. To better understand pustule formation, Wahlberg and Maibach developed a model to assess pustulogenic potential [12]. The test materials were placed on rabbits' backs and occluded for 24 hours. For some ingredients, the skin had to be abraded with a sterile needle to produce pustules. Pustule formation is dose dependent. Irritants such as sodium lauryl sulfate can elicit pustules even though they are noncomedogenic.

Human Models

Human models have been developed for looking at both acnegenic and comedogenic potential. Mills and Kligman first described the human comedogenic model in 1982 [13]. It is becoming more extensively used in the cosmetics industry as companies continue to avoid animal testing. In the human procedure, up to six test materials are applied to the upper back for 48 to 72 hours under an occlusive or, if necessary, a semiocclusive patch. Patches are applied three times a week for 4 weeks to give the 28 days of continuous exposure.

After induction, the test sites are sampled using an epidermal biopsy. A glass slide coated with cyanoacrylate (e.g., Crazy Glue®) is briefly applied to the skin for 1 minute. After it has dried the slide is removed, taking the follicular plugs and much of the stratum corneum with it. The size and number of follicular impactions are assessed using a 0-to-3 scale and compared with positive and negative controls. Positive controls include acetylated lanolin and coal tar. Mills and Kligman showed that the human model gave similar results to the 2-week rabbit-ear method (Pearson $r = 0.944$, $n = 32$ cosmetic ingredients or products). However, the rabbit-ear model appears to be somewhat more sensitive than the human assay (Table 2).

Recently, a new method for epidermal biopsies has been validated. The method uses

TABLE 1 Comedogenicity and Irritation Potential of Cosmetic Ingredients in the Rabbit-Ear Model

Ingredient	Comedogenicity	Irritation
Oils		
Cocoa butter	4	0
Coconut butter	4	0
Evening primrose oil	3	2
Soyabean oil	3	0
Peanut oil	2	0
Castor oil	1	0
Sunflower oil	0	0
Mineral oil	0–2	0
Lanolin and derivatives		
Acetylated lanolin	0	0
Acetylated lanolin alcohol	4	2
Anhydrous lanolin	0–1	0
Lanolin alcohol	0–2	0
PEG-16 Lanolin	4	3
PEG-75 Lanolin	0	0
Fatty acids and esters		
Lauric acid	4	1
Myristic acid	3	0
Palmatic acid	2	0
Stearic acid	2–3	0
Butyl stearate	3	0
Cetyl acetate	4	2
Cetyl ester NF	1	1
Isopropyl isostearte	5	0
Isopropyl lineolate	4	2
Isopropyl myristate	5	3
Alcohols, sugars, and their derivatives		
Isopropyl alcohol	0	0
Cetyl alcohol	2	2
Isocetyl alcohol	4	4
Oleyl alcohol	4	2
Stearyl alcohol	2	2
Sorbitol	0	0
Sorbitan laurate	1–2	1–2
Sorbitan oelate	3	0
Sorbitan stearate	0	0
Oleth-3	5	2
Oleth-5	3	2
Oleth-10	2	1
Oleth-20	1	0

Source: Ref. 9.

MARCEL DEKKER, INC.
270 Madison Avenue, New York, New York 10016

TABLE 2 Comparison of Human-Back and Rabbit-Ear Comedogenicity Scores

	Mean comedogenicity score	
Material	Rabbit*	Human
Acetylated lanolin alcohol	3	2
Cocoa butter	3	2
5% crude coal tar**	3	3
Isopropyl myristate	1	0.4
Safflower oil	1	0
5 or 8% sulfur**	3	2
2.5% sulfur**	2	1.2
Hydrophilic ointment	0	0

* Comedogenicity scored on a 0–3 scale, n = 3 rabbits and 5 humans.
** These test material were diluted with hydrophilic ointment. All other test materials used at full strength.
Source: Ref. 13.

commercially available cosmetic strips that are designed to remove impactions from the face without damaging the skin [14]. The bioré pore strip®, which uses a cationic polymer, preferentially interacts with the proteins of the hyperkeratotic plugs but not the stratum corneum. The plugs have more acidic amino acids and are therefore more negatively charged than the surrounding stratum corneum. An example of the bioré pore strip removing impactions from the nose is shown in Figure 2. Rizer et al. showed that the bioré pore strip® removed over 70% of the impactions that cyanoacrylate removes, but without the damage of the latter. The bioré pore strip is more effective than other cosmetic strips in removing plugs from the follicles. The other strips use nonionic polymers, which are not able to preferentially interact with the follicular plugs.

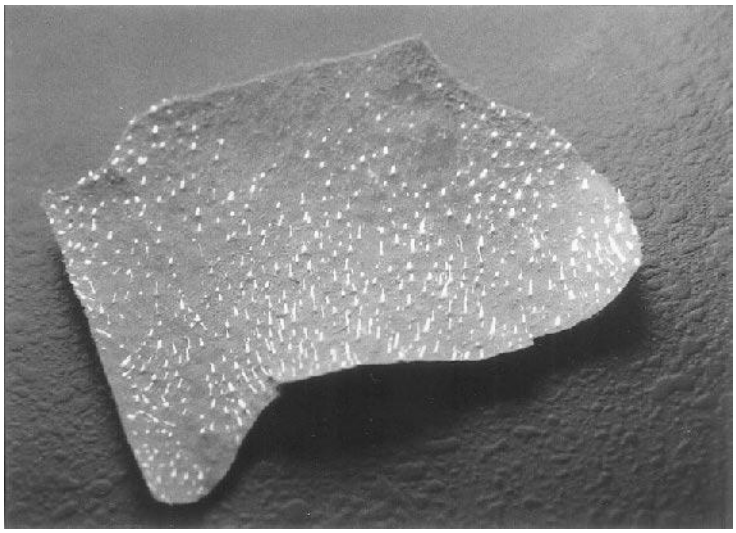


FIGURE 2 Half a bioré pore strip under UV light. The hyperkeratotic impactions on the strip fluoresce due to the *P. acnes*.

One advantage that the bioré strips have over the cyanoacrylate glue is that it is far less damaging and irritating to the skin. Therefore, it can be used to measure comedone formation on the face at the end of usage studies. Cyanoacrylate is too damaging to be acceptable to most test panelists for use on their face. The bioré pore strip has been shown to be acceptable to panelists; indeed, this is its intended use.

Human Usage Tests

Ultimately, all predictive models must be related to back to the consumers' experience in the marketplace. Consumers who experience an adverse reaction will report it in terms most familiar to them. Most consumers do not differentiate between acne and comedone formation, or blemishes and breakouts.

One approach to assessing the rate of adverse reactions is to have a group of consumers use a product for several weeks [15–18]. Test subjects should be evaluated for comedones, pustules and papules at the beginning of the study, and then at set intervals. A 1-week evaluation will reveal any propensity to cause irritation, including follicular irritation that panelists may recognize as breakouts. Any sensory irritation will become evident during the first week. Three and six-week evaluations are used to detect comedogenicity and acnegenicity. This design is consistent with the recommendation of the American Academy of Dermatology's consensus panel on acnegenicity testing [19].

A sizable proportion of adverse reactions is experienced by vulnerable subgroups. A subgroup for irritation may include panelists whose skin is readily irritated by surfactants. Another subgroup will include panelists with acne-prone skin. Both subgroups should be identified and form a significant part of the test panel. This will enable the investigator to identify potential problems before the product reaches the marketplace.

SUMMARY

The induction of comedones and acneform eruptions is a significant concern to many consumers, especially those with acne-prone skin. Any product that has a propensity to produce these eruptions will be unsuccessful in the marketplace. Indeed, many consumers expressly look for products that are labeled noncomedogenic and/or nonacnegenic.

Cosmetics manufacturers are meeting this consumer demand by showing that their products do not cause comedones and/or acne breakouts, and label their products accordingly. Consumers judge a facial cosmetic on whether it causes breakouts, blemishes, bumps, or blackheads. There are multiple causes for the reaction, including comedone formation and follicular irritation. Consumers do not differentiate between the biological mechanisms; they are only concerned with the results they produce.

Today, human models have replaced animals for testing both comedogenicity and acnegenicity. Comedone formation is determined by continuously patching the material on the human back for 28 consecutive days. Comedones are quantified by extracting the plugs from the follicle using cyanoacrylate glue on a glass slide or a bioré pore strip® containing cationic polymers. The degree of the impactions is compared with positive and negative controls.

Acnegenicity is assessed by human-use testing, where a panel of consumers uses the product under normal conditions. The skin is evaluated by a trained observer for comedones, papules, and pustules at the beginning of the study, and then after 1, 3, and 6 weeks of usage. Cosmetic pore strips such as bioré® can be used to assess comedone formation.

These can remove follicular plugs from the face without the skin damage associated with cyanoacrylate.

REFERENCES

1. S MacDonald Hull, WJ Cunliffe, BR Hughes. Treatment of the depressed and dysmorphophobic acne patient. Clin Exp Dermatol 16: 210–211, 1991.
2. AM Kligman, OH Jr Mills. Acne Cosmetica. Arch Dermatol 106: 843–850, 1972.
3. G Plewig, JE Jr Fulton, AM Kligman. Pomade Acne. Arch Dermatol 101: 580–584, 1970.
4. C Berlin. Acne comedo in children due to paraffin oil applied on the head. Arch Dermatol 69: 683–687, 1954.
5. JS Strauss. Sebaceous Glands in 'Dermatology in General Medicine' Eds. Fitzpatrick TB, Eisen AZ. Wolff K. et al., 4th Edition vol 1, p 709–726 (1993).
6. G Plewig, JE Jr Fulton, AM Kligman. Cellular dynamics of comedo formation in acne vulgaris. Arch Dermatol 102: 12–29, 1971.
7. OH Jr Mills, RS Berger. Defining the susceptibility of acne-prone and sensitive skin populations to extrinsic factors. Dermatol Clinics 9: 93–98, 1991.
8. EH Adams, DD Irish, HC Spencer, et al. The reponse of rabbit skin to compounds reported to have caused anceform dermatitis Ind Med 2: 1–4, 1941.
9. JE Jr Fulton, Comedogenicity and irritancy of commonly used ingredients in skin care products. J Soc Cosmet Chem 40: 321–333, 1989.
10. JE Jr Fulton, SR Pay, JE III Fulton. Comedogenicity of current therapeutic products, cosmetics, and ingredients in the rabbit ear. J Amer Acad Dermatol 10: 96–105, 1984.
11. WE Morris, SC Kwan. Use of the rabbit ear model in evaluating the comedogenic potential of cosmetic ingredients. J Soc Cosmet Chem 34: 215–225, 1983.
12. JE Walhberg, HI Maibach. Sterile Cutaneous Pustules: a manifestation of primary irritancy? In: Models in Dermatology 2: 297–302 Maibach HI and Lowe N, eds. Pub Krager (1985).
13. OH Jr Mills, AM Kligman. A human model for assessing comedogenic substances. Arch Dermatol 118: 903–905, 1982.
14. RL Rizer, JK Woodford, FA Simion et al. 'The follicular bioposy for assessing comedogenicity: a comparison of a pore strip and cyanoacrylate.' Poster at the International Society of Bioengineering and the Skin, June 1998.
15. EM Jackson, NF Robillard. The controlled use test in a cosmetic product safety substantiation program. J Toxicol Cut & Ocular Toxicol 1: 117–132, 1982.
16. OH Jr Mills, RS Berger, TJ Stephens et al. Assessing acnegenic and acne aggravating potential. J Toxicol Cut & Ocular Toxical 8: 353–360, 1989.
17. EM Jackson. Clinical assessments of acnegenicity. J Toxicol Cut & Ocular Toxicol 8: 389–393, 1989.
18. A Ghassemi, R Osborne, KA Korman, et al. Demonstrating the ocular safety of an eye cosmetic product using alternatives to animal eye irritation tests. Poster at the Society of Toxicology Meeting in Cincinnati, OH, March 1997.
19. JS Strauss, EM Jackson et al. American Academy of Dermatology Invitational Symposium on Comedogenicity, J Amer Acad Dermatol 20: 272–277, 1989.

Marcel Dekker, Inc.
270 Madison Avenue, New York, New York 10016

71

Sensory Testing

Linda P. Oddo and Kathy Shannon
Hill Top Research, Inc., Scottsdale, Arizona

Although most individuals don't realize it, they conduct sensory tests on a daily basis. Throughout the day, personal sensory assessments are made about the taste acceptability or liking for different foods for many different attributes. Individuals evaluate haircare products for various properties, not only while using the product but also for feel of the hair after shampooing. Every time a cosmetic product, moisturizer, or any other skincare product is applied to the skin, sensory assessments are made.

In very simple terms, the field of sensory testing applies controls, data-collection skills, and reproducible methodology to these types of assessments for the purpose of collecting not only viable but valuable information about the test materials. Civille [1] states that ''[t]he primary function of sensory testing is to conduct valid and reliable tests, which provide data on which sound decisions can be made.''

Although the roots of sensory testing as a discipline exist in the food industry, its applications have steadily earned respect in the consumer and pharmaceutical industries. Science now plays an important role in the field of cosmetology by providing guidance to the product formulator in predicting consumer response and by supporting or defining product claims.

Currently, trained sensory judges are routinely used in the development of many cosmetic products. Trained panels are used to evaluate the skin-feel properties of not only topical cosmetic products but any product that is applied to the skin. Trained judges are used to evaluate haircare products, and sensory judges are critical to determining oral and axillary malodor.

Using carefully controlled sensory applications, sensitive-skin subjects can be identified and selected. The response of these subjects can then be trusted to aid in the development of and/or to identify acceptable sensitive-skin cosmetic products. Finally, by applying sensory scaling techniques and controls to the design of self-assessment questionnaires, they are often added to clinical studies to provide a potential insight to consumer response. The primary focus of this chapter is to describe some of the sensory methods and tools that are currently being applied in these areas.

Marcel Dekker, Inc.
270 Madison Avenue, New York, New York 10016

AXILLARY MALODOR EFFICACY

Generally, axillary malodor efficacy tests are conducted on three product types: antiperspirants, deodorants, and soaps. Although antiperspirants are primarily designed to inhibit sweat production, they are also considered deodorants because they inhibit sweat, which acts as a culture medium for bacteria to produce, degrade, and form malodor. Deodorants are formulated to control malodor only, through absorption, fragrance masking, and/or by reducing antibacterial activity. Some soap products may also reduce axillary malodor by fragrance masking and/or inhibiting bacterial growth.

The use of sensory testing applications is the primary methodology used to establish deodorant efficacy. In 1987, the sensory testing division of The American Society for Testing and Materials (ASTM) published the Standard Practice for the Sensory Evaluation of Axillary Deodorancy [2]. This document recommends that a product meet the criteria presented in the standard in order for it to qualify as an effective deodorant.

The basic design for conducting a deodorant study consists of selecting subjects with high axillary malodor, applying or using the test material at selected intervals, and then measuring the level of axillary malodor using a panel of trained odor judges. The test material is considered effective if there is a statistical difference between it and a placebo or untreated control. Factors that are critical to the test design and consequently a successful study are the subject selection, subject restrictions, odor-judge selection and training, selection of a suitable test location, and using appropriate scaling techniques.

Subjects should be selected from the user population and have a distinct axillary odor. Those with extremely high or low odor and those with large differences in odor level between the right and left axillae are usually disqualified. It is important that potential subjects participate in a washout or conditioning period before selection to prevent a carry-over effect from the use of other products. For a minimum of 7 days, subjects are not allowed to use any axillary products and are instructed to wash the axillae only with a mild, nondeodorant soap. Other restrictions that are known to interfere with sensory odor assessments require the subjects to abstain from swimming, excessive exercise, and from using any fragranced products. They must also avoid spicy foods, and before an odor evaluation they are restricted from smoking.

Perhaps the most crucial factor in a well-executed malodor efficacy test is the selection and training of qualified odor judges. This process requires management support and commitment, a sensory staff or analyst to conduct the training, and a pool of available and interested candidates. Other factors to consider are the time commitments to not only select and train the judges but to continually maintain and validate their performance. The odor–judge selection and training process involves four basic steps: 1) interviewing candidates, 2) conducting screening tests, 3) training, and 4) validating performance.

During the interviewing process, candidates who have conflicting commitments or interfering health problems should be discontinued. A description of the test and the odor-judge process must be explained to each individual. If possible, because of the unusual nature of the intended task—sniffing the axillary region of subjects—a video of the process should be shown. Through interaction and discussion, those candidates who show a sincere interest and are willing to commit to the program are identified.

In vitro screening tests are administered to the potential odor judges to determine their olfactory acuity and ability to discriminate and reproduce results. Because it is possible for some individuals to be insensitive to some of the odors generated by the human body, potential judges should also be screened for this inherent lack of sensitivity.

Marcel Dekker, Inc.
270 Madison Avenue, New York, New York 10016

An inability to recognize some body odors is commonly referred to as ''specific anosmia.'' The anosmias that have been identified in axillary odor include sweaty, urinous, musky, and hircine odors, with the primary anosmia being a urinous smell. It has been reported that as many as 46 to 50% of the population are insensitive to the urinous odor [3]. Because of this high percentage, judges should be screened for the insensitivity using the odorant androstenone. Individuals who can smell this compound will rate it extremely strong and often find it offensive, whereas those who are anosmic will rate it low or may not smell it at all.

The odorant used most often to represent a sweaty smell is isovaleric acid. It is therefore often used in odor, judge acuity screening tests. Potential judges are often given a series of paired comparisons and at least one ranking test of the five established levels of isovaleric acid (Table 1) [4].

For the paired comparison tests, potential judges are given at least eight different combinations of concentrations. The pairs should represent different levels of difficulty between samples, e.g., 0.013 versus 0.87 and 0.053 versus 0.22. In the ranking test, a sample of each concentration is presented. Before administering the tests, the samples should be placed in identical bottles or jars. Each bottle is identified by a unique three-digit number. The pairs and ranking test should be randomly presented to the candidates with a distinct rest period between each test. When presented with each pair, the odor-judge trainee is asked to identify which sample has the stronger or more intense odor. For the ranking test, they rank the samples from the least to the most intense odor. It is always very important to control the conditions of the test area when administering any sensory test [5].

In addition to determining acuity, reproducibility should also be considered. This can be accomplished by administering the same tests one or two more times on separate days. The order of set presentation, bottle order, and coding system must be changed between days.

Training is initiated once individuals who show a high olfactory acuity and consistency are identified. Several steps are involved in the training process, including establishing a standard method for evaluating, identifying judge restrictions, providing reference standards that represent the scale, and conducting training sessions.

The method frequently used to evaluate the axillary region involves placing the nose near the surface of the skin located in the center of the axilla and taking several short bunny sniffs. Judges clear the sinuses by breathing into a cotton material or toweling between evaluations. The evaluation method should also include an established rest period between evaluations and/or subjects (e.g., 30 or 60 sec). The judges must also avoid touching the subject with either their nose or hands. In addition to avoiding contact, the

TABLE 1 Five Established Levels of Isovaleric Acid

Odor level	Concentration of aqueous solution of isovaleric acid (mL/L)
Slight	0.013
Definite	0.053
Moderate	0.22
Strong	0.87
Very strong	3.57

judges are restricted from wearing any personal products with a distinct fragrance. They should also be restricted from eating certain foods before evaluating.

Critical in the evaluation process is identifying a scale. To evaluate the intensity or express the degree to which axillary odor is present, two types of scales are usually considered. One is a line scale, which consists of a standard-length line on which the judge makes a mark. The primary disadvantage to this approach is that judges may have difficulty establishing consistency without a number to remember [6]. Category scale methods are perhaps the most frequently used. This type of scale involves using sets of words and/or numbers to identify established intervals on the scale.

Among the available category scales, a 0-to-10 numerical scale has been used to evaluate or score malodor intensity. Although some descriptive language may vary slightly, the zero on this scale consistently represents no malodor while the 10 represents extremely strong malodor. Table 2 is a complete example of a 0-to-10 numerical scale.

In addition to being used in judge-acuity screening tests, isovaleric acid is often used as a reference standard when training judges to use malodor intensity scales. The five concentrations previously identified can be used to represent various points on the selected scale or other concentrations can be used. After introducing the reference points, the judges should practice until they can repeatedly assign the correct score to each reference under blind conditions.

New judges being introduced to human axillary odors should, if available, train with an experienced judge. The new judge observes the score given to a certain subject then evaluates the same subject. After participating in this capacity for a period, the judge in training evaluates the subject first and then observes the scores given by the established judge. Finally, the judge trainee evaluates independently until statistical analyses of his/her data correlates with the established judges.

Training new judges without the benefit of established judges can be accomplished by using a couple of different approaches. In one approach, the sensory scientist training the group can determine the odor level of selected subjects then introduce the new judges to these odor levels using the previously discussed techniques. Another approach allows the new group of judges to standardize their scores through consensus. After each evaluation, the group discusses their scores, and repeats the process until they agree on the odor level for that subject. This process is repeated until independent evaluations correlate.

TABLE 2 A 0-to-10 Numerical Scale

Numerical value	Description of malodor
0	None, no malodor
1	Threshold malodor
2	Very slight malodor
3	Slight malodor
4	Slight to moderate malodor
5	Moderate malodor
6	Slightly strong malodor
7	Moderately strong malodor
8	Strong malodor
9	Very strong malodor
10	Extremely strong malodor

MARCEL DEKKER, INC.
270 Madison Avenue, New York, New York 10016

Although this approach is more time consuming, it often establishes a strong sense of commitment and involvement in the process for the new judges.

Once established, odor judges can be used to evaluate any personal-care product used in the axillae to control malodor. By using combinations of subject selection, product-treatment techniques, post-treatment evaluation times, and controlling environmental conditions, an almost endless number of possibilities can be evaluated by the judges. In addition to directly evaluating human subjects, odor judges can also be used to evaluate axillary odor that has been transferred to some other medium such as a t-shirt or a cloth worn against the axilla.

ORAL MALODOR EFFICACY

Currently, oral malodor efficacy studies are conducted with toothpastes, cleansers, powders, mouth rinses, toothbrushes, breath mints, tongue scrapers, and any oral treatment whose primary or secondary function is to reduce or control halitosis. Oral treatments are designed to control, mask, or eliminate sulfur-producing bacteria, the primary component of bad breath.

To accommodate the large variety of consumer products currently available for treating halitosis, clinical studies vary in their design. Variables include the profile of the target population in their medical and dental history, current health conditions, and personal practices. Other elements considered when designing oral-malodor clinical studies include the number of treatments, evaluations, and post-treatment evaluation intervals. Evaluations may include any combination of professional examinations, microbiology sampling, oral-malodor assessments, and instrumental measurements.

Instruments that have been used to measure levels of malodor include gas chromatograph (GC), which has been used to analyze oral volatile sulfur compounds. In a clinical study comparing the GC with sensory odor judges, the instrumental measurements showed good correlation with the organoleptic assessments. The GC, however, is considered large, cumbersome, and difficult to use in a clinical setting [7]. A portable sulfide monitor, easier to use in a clinical environment, has also been investigated and found to fall within the range observed with the GC. When compared with odor judges, the Halimeter® also significantly correlated ($p < 0.001$) with sensory ratings [8].

Although good correlation has been established, the manufacturer of the Halimeter® states that the data independently cannot confirm the existence of oral malodor because volatile sulfur compounds are not constant in any one person. They recommend using the instrument with other procedures, such as bacterial cultures and organoleptic measurements to assess levels of oral malodor.

Organoleptic measurements or assessments are generally conducted by judges specifically trained to evaluate oral malodor. The selection and training of these judges is similar to the techniques used to select and train axillary malodor judges. Differences include the use of reference standards more appropriate to oral malodor and training the judges in a different process of evaluation. In the oral-malodor evaluation process, the judge and subject are separated by a solid partition. The partition has a small circular opening in which the subject inserts a glass rod. During the actual assessment, the subject places his/her mouth around the end of the glass rod and either holds his/her breath or exhales into the tube while the judge places his/her nose near the other end of the tube.

Currently, two very different types of sensory scales are being used to measure oral malodor. One applies hedonic measurements whereas the other approach uses category

TABLE 3 The Peryam and Pilgrim Scale

Numerical value	Hedonic description
1	Most pleasant
2	Very pleasant
3	Moderately pleasant
4	Slightly pleasant
5	Neutral (not bad/no odor/not good)
6	Slightly unpleasant
7	Moderately unpleasant
8	Very unpleasant
9	Most unpleasant

scaling. Typically, hedonic measurements are used by untrained consumers to indicate a level of liking for the material in question. For example, Tonzetich used a panel of eight ''observers'' to rate their responses to different oral-cleansing treatments on a 0-to 6-point hedonic scale. On this scale, 0 represents an absence of odor, while 6 represents a strongly objectionable odor [9]. By using the term ''objectionable,'' the scale becomes a measure of displeasure or disliking for the odor.

Hedonic measurements have been successfully used by a smaller group of judges who have been trained to score the presence of oral malodor as unpleasant and the absence of malodor as pleasant. These judges use the 9-point hedonic scale developed by Peryam and Pilgrim (Table 3) [10]. This scale has a neutral midpoint with degrees of pleasant or unpleasant increasing in opposite directions.

Judges trained to use a category scale are instructed to rate the intensity of the odor present. The pleasantness of the smell is not considered. Various lengths or sizes of the scales can be used if the judges are trained to identify the different intensities, and if they not only correlate to each other but are also reproducible. Examples include a 0-to-3 numerical scale, in which each score represents a range of odor (Table 4) [11].

Each point on the following 0-to-5 scale (Table 5) is designed to represent one level or intensity of oral malodor.

In a paper presented at the 4th International Conference on Breath Odor (IADR), intensity judges using the 0-to-5 category scale were compared with hedonic judges who applied the 9-point hedonic Peryam and Pilgrim scale. The purpose of the research was to determine if both types of judges were able to assess oral malodor under an identical clinical setting. Results found a positive treatment effect from baseline compared with the control when either the hedonic scale or intensity scale was used ($p = 0.0001$), with similar percent reductions for each set of judges. The intensity scores had a reduction

TABLE 4 An Example of the 0-to-3 Numerical Scale

Numerical value	Description of malodor
0	None to low odor
1	Low to moderate
2	Moderate to high
3	High malodor

MARCEL DEKKER, INC.
270 Madison Avenue, New York, New York 10016

TABLE 5 An Example of the 0-to-5 Scale

Numerical value	Description of malodor
0	No odor
1	Questionable odor
2	Faint odor
3	Moderate odor
4	Strong odor
5	Very strong odor

from baseline of −4.51, −2.32, −1.19, and −0.23 (for immediate, 30-, 60-, and 90-min post-treatment respectively) as compared with the hedonic scores which had a reduction of −4.29, −2.49, −1.65, and −.88 [12].

Frascella also used both types of judges to compare the effect of a chlorine dioxide treatment on mouth odor. In this research, the intensity judges used a 0-to-4 category scale and the hedonic judges used a 7-point bidirectional scale. In this research, both judges showed significant treatment effects at the 2- and 4-hour evaluation intervals [13].

In addition to using trained-judge assessments and instrumental measurements to determine levels of oral malodor, some work has been done to better understand the role of self-perception or self-assessment of oral-care treatments. Most agree that individuals have trouble detecting their own halitosis because of adaptation or dulling of sensations that result from continued exposure [14,15]. Because of this, adaptation attempts to accurately conduct self-evaluations using methods such as cupping the hand over the mouth, licking the hand, smelling dental floss, and breathing into fabric have not correlated well with more objective assessments [16]. Regardless, there still remains a potential value to understanding when and how individuals perceive their breath as offensive. This may be better understood by focusing on other self-perceptions rather than self-assessments.

Recently, a self-perception questionnaire was administered to 32 subjects participating in an oral-malodor study that included hedonic and intensity organoleptic evaluations. In addition to assigning a breath-odor score, subjects were asked to rate other experiences. These perceptions included current pleasantness of taste, freshness of the mouth, clean mouth feel, general feeling of offensiveness, a bitter taste, and feel of teeth. Finally, subjects were asked to rate the overall effectiveness of the product. At each post-treatment interval, responses to each question showed statistically significant differences among treatments favoring the positive control ($p < 0.001$). These findings supported the trained-judge assessments [17]. This is an area of thought that deserves further exploration and understanding. As individuals or consumers ultimately decide when they need to freshen their breath and their subjective evaluation that often determines the effectiveness of the treatment when used in a personal setting.

DESCRIPTIVE SKIN FEEL

Skin feel is an important sensory area for bodycare and cosmetic products. These sensations directly affect the consumer's perception about the efficacy of the product. Products that are efficacious may not be successful when marketed because of negative reactions to how quickly they absorb, smell, feel during application, or feel and look on the skin after use. Whereas a clinical study can show that a lotion or cream can alter the surface

of the skin, only a sensory test can predict if this alteration will be perceptible to the consumer and give dimension and value to these perceptions.

These studies are conducted using descriptive sensory analysis. Descriptive analysis is perhaps one of the most sophisticated techniques used in the field of sensory testing. With this approach, participants or panel members describe the perceived characteristics of a material and then measure the strength of selected attributes on a scale. Formal descriptive analysis started in the food industry in the 1950s with Flavor Profile, which is a process for describing the aroma and flavor of various food products. A Texture Profile method was later developed in the 1960s to focus on the textural aspects of foods that were omitted in Flavor Profile [18]. Although this method was eventually expanded by Schwartz to include terminology specific to the skin feel of products, it remained based on the underlying principles of the original Texture Profile method [19].

Quantitative Descriptive Analysis (QDA) was perhaps one of the first descriptive approaches developed to investigate both foods and other consumer products. This method uses panel members who are users of the specific product being evaluated. Unlike other methods, these panelists spend only 5 to 6 hours in training sessions during which they develop a language for the product. According to Stone, there is ''no attempt to standardize responses, scores or train to score a particular attribute to some standard.'' Products are tested over several days using a repeated trials design collecting at least three responses from each participant for each parameter. Supporters believe this approach frees the methodology from dependence on the same panel and allows the language to be dynamic [20].

To capture the effect of time on the release of various attributes, time-intensity descriptive analysis was developed. This approach provides information on the dynamic nature of the response by monitoring the intensity of specific attributes over time. For example, the panel member may be asked to rate the intensity of several attributes every 10 to 15 seconds after use. With products that have a tendency to noticeably change over time, this technique has the potential to provide significantly more information than the more traditional sensory methods that measure attributes at specific intervals [21].

In the Spectrum descriptive analysis method, panel members rate the intensity of a product in relation to absolute or universal scales that are constant for all product types. This approach provides tools to custom design a panel for a specific product category and can be applied to a variety of product areas, including personal care. The final panel of approximately 15 members is carefully selected from a large group of individuals who participate in two screening phases. Once screened, the identified candidates then participated in at least 3 months of training during which they review samples that represent the product category, review references, define terminology, evaluate products, and discuss results. The performance of the panel must be established before they evaluate unknown test materials [22].

The DermatoSensory Profile approach to descriptive skin-feel analysis was introduced in 1986. Originally this panel was trained to evaluate only lotions and creams, but was expanded to other products that affect the feel of the skin, e.g., soaps, facial cleansers, antiperspirants, powders, and shaving products. The original panel of judges was carefully screened and selected before spending approximately 6 months in training. During this training period, under the guidance of a moderator, the group worked with a wide variety of marketed products to establish key attributes, agree on definitions, determine evaluation procedures, and select reproducible reference standards. The outcome involves a process of applying a standard amount of the test sample to a circle marked on the inner arm. Most of the attributes are evaluated independently with appropriate reference products

Marcel Dekker, Inc.
270 Madison Avenue, New York, New York 10016

continually used to anchor the 0- to 10-point intensity scales [23]. Some of the lotion and cream attributes selected by the panel include the rate of absorption, shine, greasy oily, drag, stickiness, ease of spread, and residue at several intervals after application. When five marketed products described as four oil-in-water (o/w) formulations and one water-in-oil (w/o) formulation were evaluated, the panel was capable of showing significant differences among all of the products for different attributes [24].

In 1992, the ASTM published the Standard Practice for Descriptive Analysis of Creams and Lotions [25]. This practice identifies the elements of and the process for training a skin-feel panel. In addition to identifying the needed equipment, it presents a process for screening and selecting panel members. One section describes an evaluation procedure that, in addition to explaining an application process, also discusses sample preconditioning, conditioning aspects of the skin including skin temperature, and environmental conditions of the test area. Evaluation intervals, definitions, and suggested references are included for each listed attribute. The practice does state, however, that it should be used by individuals who have become familiar with the process and have previous experience with sensory testing.

Descriptive skin panels fill the gap between clinical and marketing data by providing information that can help predict or better understand consumer needs. It has been used develop a master profile of a product that is later used for quality-control purposes or to improve the product. Panel information has also been frequently used to determine differences in currently marketed products whereas the descriptive terms and results are often used to promote or market the product.

There were a number of product-performance trends in the skincare industry during the past decade that may have benefited from a descriptive profile, in either the product-development stage or in better understanding the competition. For example, hydrating agents were added to increase skin moisturization. Although the physiological improvement of these agents is established in clinical studies, descriptive sensory data is essential to identifying potential consumer perceptibility. The move to silicone emulsion systems to decrease the heavy, greasier feel created with oil systems was a natural application for descriptive skin-feel data. Finally, industry responded to the increase in consumer awareness of the cumulative effect of sun exposure on the skin by adding sunscreens to many bodycare and cosmetic products. However, the addition of sunscreens often affects the skin-feel properties of the product. For example, they can increase the rate of absorption, add a greasy feeling, and create a heavier texture to the product. Descriptive skin-feel analysis was and continues to be an appropriate tool to address and minimize the effect of these changes on consumer perception.

DESCRIPTIVE HAIRCARE

The competitive world of haircare products is very dependent on consumer perception. The success of a product often depends on whether the user perceives a positive change or believes the claims being presented. Regardless of what can be shown clinically, it is ultimately the consumer who decides if his/her hair is shinier, easier to comb, has more body, or holds a curl longer. For these reasons, descriptive sensory analysis plays an essential role in the product-development stage. When appropriate sensory tools are used, these characteristics can be confidently assessed in a controlled environment. Formula changes as well as new ingredients and ideas can be screened before substantiating product-performance claims with large-scale consumer studies.

Marcel Dekker, Inc.
270 Madison Avenue, New York, New York 10016

Currently, the majority of descriptive sensory analysis with haircare products is being conducted internally. Most manufacturers use licensed cosmetologists, trained panels, or groups of semitrained consumers. Some companies will use each of these tools depending on the stage of development or the product type.

Cosmetologists are perhaps the most sophisticated tool used, and sometimes the most challenging to the sensory scientist. Regardless of experience, these individuals must be screened and carefully selected. Once selected they must be trained to follow established procedures such as those for washing the hair, combing the hair, and touching the hair. They are also trained in what characteristics or attributes to evaluate and when. Because these individuals often have many years of experience working in a salon environment, they are often faced with the challenge of changing old habits. At the same time, the cosmetologist typically provides a certain level of knowledge or experience that even well-trained consumers don't possess.

Panelists for a descriptive haircare panel are selected and trained using techniques similar to those identified for developing skin-feel panels. However, these panelists are trained to evaluate hair rather than their own skin. Consequently, providing samples during the training process becomes perhaps the biggest challenge to developing the panel. Hair swatches are often used during the training process because using actual subjects can become costly and burdensome. They also provide a distinct advantage because the type and condition of the hair can be carefully controlled. This introduces one of the primary considerations of evaluating haircare products; subject selection. Because products will react differently on different hair types, subjects must be carefully screened and selected based on the type of hair they have. Some of the things that must be considered are hair texture, thickness, length, color, and amount of natural curl. The condition of the hair (e.g., dry or oily) and pretreatments, (e.g., permed or colored) must also be considered. A well-defined profile of the subject should be established before testing, and a method for selecting subjects using the trained panel, cosmetologist, or an independent person should be built into the program.

Depending on local laws and regulations, unlicensed trained panelists may not be allowed to handle the subjects. Because they may be restricted from shampooing, they are often only used to evaluate the feel or appearance of the hair. Some panels are trained only to evaluate hair swatches or a combination of both.

As mentioned earlier, because the majority of descriptive sensory analyses with haircare products is conducted internally, very little has been published in this area. To provide the industry with information, ASTM committee E18.0 on sensory testing is in the process of finalizing a standard practice for the descriptive analysis of shampoo performance [26]. This practice will present an overview of several options, some of which were previously described, that the sensory associate can follow to develop a hair descriptive program. Similar to the skin-feel standard practice, it will identify necessary equipment, a process for screening panel members and/or cosmetologist(s), as well as evaluation and application procedures. Although this document will focus on shampoo performance, it has the potential to provide an excellent panel foundation that can be expanded to the other haircare products.

ANTI-IRRITANT APPLICATIONS

Manufacturers fully understand the necessity of thorough safety testing before introducing a topically applied product. A battery of standard safety tests to determine the irritancy

potential of the product to contact sensitization and photosensitization are routinely conducted. However, often missing are tests to determine potential subjective discomfort to the product, such as stinging, burning, and itching. In 1977, Frosch acknowledged that products that meet standard safety parameters may still be rejected by the consumer if disagreeable subjective discomfort develops after application [27].

Early work in this area investigated subjective response to substances applied to skin that had been damaged by either blisters [28] or scotch tape–stripped skin [29]. However, these were considered measurements of pain rather than measurements of more transient subjective discomfort. In response to concerns that some substances, such as sunscreens, may cause delayed stinging, Frosch and Kligman developed a method for identifying potential ''stingers.'' This method involved applying lactic acid to the nasolabial fold and cheek area of subjects brought to a profuse state of sweating. The intensity of stinging was then measured by the subject using a 4-point scale at 2.5, 5.0, and 8.0 minutes after application. It was also established that a stinging response could be induced in nonsweating subjects by increasing the concentration of lactic acid. An arbitrary method for classifying the irritancy potential of substances was also developed that identifies if the substance has a slight, moderate, or severe potential to cause stinging [27]. Although this method was established over 20 years ago, modified versions of it remain the basis for identifying subjects that are unusually sensitivity to stinging.

Grove improved the method by defining the demographic profile of the subjects and recommending the exclusion of males and older individuals. He also established criteria limiting the frequency of use and determined that sensitive subjects often reported a history of problems with soaps, cosmetics, and other personal-care products. Subjects who repeatedly reported a stinging response to lactic acid applied under ambient conditions were also tested for a burning and itching response. A method for evaluating burning sensations using a 20:80 mixture of chloroform:methanol pipetted into a greased aluminum cylinder covered and placed against the skin was used. To elicit itching, a 4% histamine base was also loaded into a grease-ringed cylinder and placed against the skin. Results found good correlation between burning and stinging, but individual response variability was rather high. A distinct correlation between itching and stinging was not observed [30].

A new interest in subjective sensory responses was renewed with the impact of alpha-hydroxy acids (AHAs) in the marketplace. When applied to the skin, these acids often cause a burning, stinging, and/or itching response, often without a visible sign of typical irritation. Draelos states that users have been conditioned to believe that stinging or burning sensations are an indication that the product is working, whereas ''in fact, they feel pain because the acid has penetrated the dermis and is interacting with the dermal nerve endings [31].'' Manufacturers have responded to these concerns by introducing a second generation of AHAs that do not penetrate the skin to the same degree. Although the FDA has found AHAs safe at low concentrations, the need to routinely include independent sensory assessments in the standard battery of safety studies is apparent.

Obviously missing from the available literature is a clear understanding of the sensory principles that were followed, which opens the door to certain questions that need to be addressed. Were the testing environments controlled? How were the unknown materials presented to the subjects? Were different scales explored? How was the scale that was used presented to the subjects?

For example, it may be advantageous to screen individuals for current use of certain medications that may affect their response, such as cortisones and other anti-inflammatory medications. Subjects should be screened for obvious skin pathology or irritation as well

as a history of allergic reactions. When conducting the screening test or evaluating unknowns, the subjects should be preconditioned for several days. During this period, they should be provided with instructions informing them of, e.g., which cleanser to use, when males can or can't shave, what cosmetics are acceptable the day of the study, and when to cleanse the face. If the test is to be conducted at ambient conditions, the subjects should be preconditioned in a quiet, climate-controlled room. If the subjects will be brought to a ''profuse'' state of sweating, this state should be carefully defined to ensure that all are brought to the same level.

The actual screening probe should be administered in isolated areas to avoid subject interaction and influence. The subjects should not be told what the appropriate response is. For that reason, it would be wise to either eliminate the word ''stinging'' from the scale or not let the subject see the scale and ask them to verbalize their response. The ''purest'' way to approach it would be to administer the test and ask the subject to report any sensation they experienced. To increase the sensitivity of the results and decrease variability it may be worthwhile to explore a 0- to 7-point scale. Finally, the frequency of subject use should be limited with a distinct rest period (~48 hours) between evaluations. With these considerations incorporated with the methods previously developed, it would be possible to quantitatively assess the intensity of facial stinging. Once a reliable method is established, a database of responses to known ingredients can be collected that will allow unknown substances to be tested for subjective discomfort with confidence.

REFERENCES

1. Meilgaared M, Civille G, Carr B. Sensory Evaluation Techniques. Vol. 1. Boca Raton: CRC Press, Inc., 1987:1.
2. American Society for Testing and Materials. ASTM Designation E1207-87, ASTM Annual Book of Standards, Vol. 15.07. Standard Practice for the Sensory Evaluation of Axillary Deodorancy. 1987.
3. Labows J, Leyden J, Preti G. Axillary Odor Determination, Formation, and Control. Cosmetic Science and Technology. Vol. 20, 71
4. Wild J, Bowman J, Oddo L. Clinical Evaluation of Antiperspirants and Deodorants. Cosmetic Science and Technology. Vol. 20, 1999:318.
5. Chambers E, Wolf M. Sensory Testing Methods. ASTM Manual: MNL 26. General Requirements for Sensory Testing. 1996:3–5.
6. Meilgaared M, Civille G, Carr B. Sensory Evaluation Techniques. Vol. 2. Boca Raton: CRC Press, Inc., 1987:3.
7. Niles H, Gaffer A. Relationship between sensory and instrumental evaluations of mouth odor. J Soc Cosmet Chem 1993; 44:101–107.
8. Rosenberg M, Kulkarni G, Bosy A, McCulloch C. Reproducibility and sensitivity of oral malodor measurements with a portable sulphide monitor. J Dental Res 1991; 70(11):1436–1440.
9. Tonzetich J, Ng SK. Reduction of malodor by oral cleansing procedures. Oral Surg 1976; 42: 172–181.
10. Peryam D, Pilgrim F. Food Technol 1957; 11:9–14.
11. Schmidt N, Tarbet W. The effect of oral rinses on organoleptic mouth odor rating and levels of volatile sulfur compounds. Oral Surg 1978; 45:876–882.
12. Borden L, Oddo L, Bowman J. Correlation between Hedonic and Intensity Measurements for the Evaluation of Oral Malodor. Presented at 4th International Conference on Breath Odor, UCLA, Los Angeles, CA, Aug. 1999.

Marcel Dekker, Inc.
270 Madison Avenue, New York, New York 10016

13. Frascella J, Gilbert R, Femandex P, Gorden J. Effect on oral malodor of chlorine dioxide assessed by a hedonic panel. TKL Research Abstract.
14. Rosenberg M. Clinical assessment of bad breath: current concepts. JADA 1996; 127:475–481.
15. Tonzetich J. Production and origin of oral malodor: a review of mechanisms and methods of analysis. J Peridontal 1977; 48:13–20.
16. Rosenberg M, Kozlovsky A, Gelemter L, Cheriak O, Gabbay J, Baht, R, Eli I. Self-estimation of oral malodor. J Dental Res, 1995; pp. 1577–1582.
17. Oddo L, Borden L, Bowman J. Comparison of trained judge measurements to self-perception of oral malodor. Abstract, 4th, International Conference on Breath Odor, UCLA, Aug. 1999.
18. Chambers E, Wolf M. Sensory Testing Methods, ASTM Manual: MNL, General Requirements for Sensory Testing. 1996; 26:58–63.
19. Schwartz N. Method to skin care products. J Texture Studies, 1975; 6:33.
20. Stone H, Side J. Sensory evaluation for skin care products. Cosmet Toilet.
21. Lee W. Single-point versus time-intensity sensory measurements. J Sensory Studies 1989; 4: 19–30.
22. Meilgaard C. Descriptive Analysis Techniques, Designing a Descriptive Procedure: The Spectrum Method. Boca Raton CRC Press, 1991: 196–199.
23. Oddo L, Aust L. Applications of sensory science within the personal care business, J Sensory Studies 1989; 3:187–191.
24. Aust L, Oddo L, Wild J, Mills O. The descriptive analysis of skin care products by a trained panel of Judges. J Soc Cosmet Chem 1987; 38:443–449.
25. Standard Practice for Descriptive Analysis of Creams and Lotions, ASTM E1490, Annual Book of ASTM Standards, Vol. 15.07.
26. Standard practice for descriptive analysis of shampoo performance. ASTM Committee E18.0, publication pending.
27. Frosch P, Kligman A. A method for appraising the stinging capacity of topically applied substances. J Soc Cosmet Chem 1977: 197–209.
28. Armstrong D, Dry M, Keele C, Markham J. Methods for studying chemical excitants of cutaneous pain in man. J Physiol 1951; 115:59.
29. Laden K, Studies on irritancy and stinging potential. J Soc Cosmet Chem 1973; 24:385–383.
30. Grove G, Soschin D, Kligman A. Adverse Subjective Reactions to Topical Agents, Cutaneous Toxicity. New York: Raven Press 1984, pp. 203–211.
31. Brewster B. MD's address sensory irritation from AHA's. Cosmet Toilet 1998; 113:9–10.

Marcel Dekker, Inc.
270 Madison Avenue, New York, New York 10016

Marcel Dekker, Inc.
270 Madison Avenue, New York, New York 10016