EVALUATION OF EFFICACY AND SAFETY OF COSMETICS*

By Albert F. Guiteras
Foster D. Snell, Inc., New York, N. Y.

It is with considerable hesitation that I address this Society, because I am not a cosmetic chemist, and I realize full well that all of you know more about cosmetics than I do. My only excuse for accepting the kind invitation of the chairman of your Program Committee is that I believe many of you may be more familiar with formulating cosmetics than with testing them.

As one phase of our work as consultants, we are frequently called upon by outside firms to evaluate different products, many of which are cosmetics. The purpose of these tests is usually two-fold: first, to establish the safety of the product under conditions of use, and second, to determine its effectiveness in accomplishing its intended objective. Since it is difficult and often impossible to establish definite absolute standards, many of the tests, particularly the so-called performance tests, are relative. In this case, comparative tests are usually run against several competitive products purchased on the open market.

Safety tests are designed to determine the safety of the product under normal conditions of use by normal individuals. Thus, in determining the safety of a lipstick, for instance, we attempt to establish whether or not it is apt to produce any irritation when applied to the lips of a normal, average, healthy woman. We are not concerned with its possible effects on a woman suffering from some particular disease, nor with any possible allergic manifestations on hypersensitive individuals, nor with its effects on the gastro-intestinal tract if accidentally swallowed by a child. The best method of determining the probable safety for use of most cosmetics and raw materials is the patch test. When working with new or unidentified materials, concerning the safety of which we can find nothing in the literature, we prefer to run some preliminary tests on rabbits, followed by a small number of humans working in our laboratory. This screening test is usually done on ten subjects and is for the purpose of discarding products that may be strong primary irritants.

Usually, these screening tests can

* Presented at the December 6, 1946, Meeting, New York City.

be dispensed with if we know the formulation of the product. The final patch test is done on at least fifty, and preferably one or two hundred subjects. The conditions of the patch test vary with the product being tested and its method of use. Some materials, such as lipsticks, are applied to the skin and left uncovered, although usually the patch is covered with an Elasto-patch, which is a strip of elastic adhesive tape, the centre of which is protected with a cellophane disc to prevent the adhesive material coming in contact with the product being tested. We ordinarily use the flexor surfaces of the arms, and although we prefer to use the lower arms because they present a more rigid surface and the patches remain in place better, we are often obliged to accede to the demands of female subjects that they be patched on the upper arm. Occasionally we use the back instead of the arms, especially when a relatively large test area is desired. Some products are applied as is, others are moistened with water or artificial perspiration. The patches are worn for periods of time ranging from several hours, in the case of certain solvents and plasticisers, to five days, in the case of textile materials. But for most cosmetics, the patches are worn for one or two days. The subjects are examined for any irritation, erythema or other dermatitis immediately after removal of the patch. The test area is then thoroughly cleaned with a suitable non-irritant solvent, followed by soap and water, and the subjects again examined after one or two hours. They are again examined the following day and at two-day intervals for one week, in order to catch any delayed reactions. If performed on a large enough number of subjects, this test will indicate whether or not the product is apt to produce an irritation on primary contact, that is whether it is a primary irritant. Obviously, the greater the number of subjects patched, the greater is the accuracy and dependability of the test. If the results are questionable, or if they are apt to involve court testimony, they are checked by our staff physician or preferably by a consulting dermatologist. Some products are not primary irritants but they are sensitizers. In other words, a person may not react to a product when he first comes in contact with it, but he may thereby be sensitized to it so that subsequent exposure may produce a reaction. Although this is really an allergic reaction, the manufacturer should guard against the use of materials that may be sensitizers to even a small percentage of the population. The period during which a person is most apt to show maximum sensitivity is ten days to two weeks after the original exposure. For this reason, if we wish to determine whether a product is apt to be a sensitizer, the same subjects are repatched after ten days to two weeks with the same products in the same place as previously. The technique is exactly the same as that used in

the original patch test. We owe a great debt to our next speaker, Dr. Louis Schwartz, who is the outstanding authority on this subject, and in whose presence I hesitate even to mention patch tests. His publications and written advice have been of inestimable value to us, and before tackling any unusual problem, we always ask ourselves, "How would Louis Schwartz do it?"

Some substances, particularly certain dyes, may not be either primary irritants or sensitizers, but they may be so-called photosensitizers, that is they may cause a reaction only when activated by ultra-violet light. Thus any pigmented cosmetic to be used on the face or hands and which may be exposed to the action of sunlight during use, such as lipstick, rouge, face powder, cake make-up, nail lacquer, etc., should be tested for photosensitization. This is done by exposing the patched area to the action of ultra-violet light for a period of time calculated to produce a mild erythema on the subject. This varies from five to fifteen minutes, depending on the complexion of the subject. It is preferable to use natural sunlight, but because it is impossible to standardize, we always use artificial ultra-violet light. We use a 1 per cent solution of eosin as a positive control, and we find that anywhere from 10 to 25 per cent of the subjects are photosensitive to eosin.

We must bear in mind that the results of patch tests are much more dependable when they are positive than when they are negative. In other words, it is much easier to say that a product should *not* be used because of positive reactions, than it is to say it is *probably safe* to use because of negative reactions. All products passing the patch test should be followed up by actual use tests before the product is placed on the market.

So much for safety tests. Now let us turn our attention to performance tests. There are no set rules to go by, but we must adapt our tests to the product and what is expected of it. Stability tests are of prime importance. The product should not change its characteristics on aging for a reasonable period of time. If it is a cream, it should remain as a cream and not harden or cake. If it is an emulsion, it should not break into two phases. Accelerated aging tests are run at 52°C. (about 125°F.) when the product will stand that temperature. In general, we have found that if a product is stable for three weeks at 52°C., it will probably be stable for one year of normal shelf life. Stability at low temperatures is also a factor. These tests are run at 4°C. (about 40°F.). It is often advisable to determine whether a product that has been frozen will regain its original consistency on thawing. Comparative accelerated aging tests are also run on emulsions in a centrifuge, where different competitive products are centrifuged at a definite speed and the time required for each emulsion to break is determined.

Stability to bacterial and fungous

decomposition is an important factor, particularly for creams. Is the product actually bacteriostatic and fungistatic? Will it inhibit the growth of typical bacteria and fungi under optimum conditions for their growth? Failing that, will these organisms grow in the product itself if it is inoculated with them? Many headaches due to mold growth and rancidity in jars could have been prevented if these relatively simple tests had been run. We have had liquid products that were perfectly stable as long as the bottle was unopened, but that soon developed a foul odor on opening and partially removing the contents, due to aerobic, air-borne organisms which would only grow when air was admitted.

Pigmented cosmetics are tested for their relative opacity and covering power. Nail polishes and lacquers are tested for their light reflectance and for their drying time, adhesion to metal and other surfaces, and for flexibility of the dried film. Cake make-up is tested for its adhesion to leather, simulating skin, and for its ease of removal by soap and water as well as by cleansing creams. Many other tests are devised as a method of evaluating the comparative performance of various products.

Sun-tan lotions, although legally drugs, are more often thought of as cosmetics. Their effectiveness depends on their ability to absorb the erythema-producing ultra-violet waves of sunlight, in the band between 2900 and 3200 Ångstrom units. It is only the rays within these wave-lengths that produce sunburn and which should be screened out. Since these same wave-lengths are the ones that produce tanning, it is not desirable to exclude them entirely, but only to reduce them by about 75 to 90 per cent.

Ultra-violet absorption curves are run on these products, in a quartz cell in a film thickness of 0.02 mm., which is the usual thickness of the film on the skin. Some sun-screening agents may be very effective in absorbing ultra-violet rays, but may be totally unfit for use in a sunburn preventive, because they are rapidly destroyed by the longer wave-lengths in sunlight and would rapidly lose their effectiveness on the beach. For this reason, the ultra-violet absorption curve is again run after exposing the product in the quartz cell to the action of four hours of sunlight. These tests are comparative and must all be run simultaneously. Results obtained on one day cannot be compared with those obtained on another day.

Deodorants and antiperspirants are other products which we usually think of as cosmetics, although antiperspirants are actually drugs. A deodorant which merely masks the odor of perspiration with a more pleasant odor is difficult to test. This can be done by use of the osmoscope, which is an instrument that utilizes the principle of diluting a given concentration of odor with air until a concentration is reached where the objectionable odor is

barely perceptible. This is called "the threshold of perception of odor" and is assigned a numerical value known as the pO value. If two test pieces of gauze are worn under both armpits of several subjects for the same length of time, one axilla with and the other without a deodorant, it is possible to evaluate the effectiveness of the deodorant by the difference in the pO values. These tests are open to subjective influence, and great care must be taken to conceal the identity of the test sample from the observer. Most deodorants that are not antiperspirants depend for their effectiveness on the fact that they prevent bacterial decomposition of perspiration. The objectionable odor is due almost entirely to metabolic products produced by bacterial decomposition, since fresh perspiration has practically no odor normally. This type of deodorant is tested for its bacteriostatic properties. Antiperspirant preparations may be evaluated *in vitro* by measuring their capacity to precipitate protein. Several *in vivo* methods have been purposed for determining antiperspirant effectiveness on human beings by measuring the amount of perspiration transpired from a given area in a given length of time. We have not had much success with any of these methods and we find it extremely difficult to duplicate results.

Any test requiring a diagnosis or evaluation of a skin condition is run in conjunction with a dermatologist. In order to avoid subjective influence, the products are given code numbers. Neither the subjects nor the examining physician have any knowledge of the identity of the products or of the code system. Such tests are used to evaluate the relative effectiveness of hand creams, lotions, soaps, etc., where the original condition of the skin is compared against its condition after the use of a product or products for a definite period of time. In cases of doubt, the results are submitted to a consulting mathematician for statistical evaluation.

I have attempted to give you some of the highlights of safety and performance tests which consultants are frequently called upon to make on cosmetic products. There are others not mentioned here.