Exploring the Pitfalls in Clinical Cosmeceutical Research

Zoe Diana Draelos, MD

Clinical research is an important method for validating the usefulness of technologies and products in the dermatologic cosmeceutical armamentarium. There is no doubt that the double-blind, placebo-controlled study is the standard study design employed in the search for truth. However, this validated research methodology may not always yield accurate results in the cosmeceutical realm. Why? This is an important question because cosmeceuticals do not lend themselves to the same type of scrutiny as pharmaceuticals, yet validation of casual observations, supposition, and in vitro observations require confirmation.

After evaluating the studies I have designed and administered over the past 20 years, it has become apparent that there are many pitfalls in clinical cosmeceutical research. I can design a topical over-the-counter product study that will fail, and I can design one that will succeed. I can make a poor product appear worthwhile and a good product appear ineffective. I can take a superb study design and make the data inconclusive. I can turn a failing study into a success. I can tell the truth, or I can lie. This means intellectual honesty is of paramount importance in clinical cosmeceutical studies.

This article explores the pitfalls in clinical cosmeceutical research based on my experience. I think this is a valuable exercise, as it will allow the reader to critically evaluate cosmeceutical studies for scientific validity.

What Is Wrong With the Placebo-Controlled Study?

In pharmaceutical trials, there can be no doubt that the placebo-controlled trial is the ideal study model. One group of subjects receives the actual medication, while the other balanced group of subjects receives a sugar pill. The study is clear cut since there is no chance that sugar can improve the appearance of acne. Any amelioration of acne seen as a result of the sugar ingestion can be ascribed to the enthusiasm of the investigator, the subject, or both. This is not the case in cosmeceutical research. In many cosmeceutical formulations, the main active is the vehicle, not the patented ingredient complex. Although it is important to study the vehicle on its own and the vehicle combined with the patented ingredient complex, the vehicle effect may be so profound as to make it impossible to achieve statistical significance. A combination of petrolatum, glycerin, and dimethicone in the cosmeceutical vehicle has profound cutaneous effects, including a reduction of transepidermal water loss, enhanced skin hydration, improved tactile smoothness, increased skin shine, reduced pruritus, and minimized wrinkles. These effects are so dramatic that the vehicle cannot be classified as inactive.

Perhaps cosmeceutical studies should be designed to compare the vehicle containing the patented complex versus no treatment. However, this, too, would be a faulty study design. Even the poorest-functioning moisturizer is better than nothing at all. A cosmeceutical study must include a comparative product, but the choice of the comparative product can yield an artificially positive or negative result. Many companies will test their new moisturizer formulations against the leading market competitor. This design is not valuable for clinical purposes. The idea that the new product is better than the best-selling consumer product will provide marketing value, but not medical value.

This has led to the idea of developing standard comparative products. For example, most new skin cleansers are compared with 2 standards. If the purpose of the comparison is to demonstrate superior moisturization, Dove® soap is used as the control; if the purpose of the comparison is to demonstrate basic cleansing, Ivory® soap is selected as the control. New moisturizer formulations are frequently compared with Vaseline® Intensive Care Lotion, which is the market leader in the body lotion category.

Picking the proper comparison product is part of the art of cosmeceutical study design. The appropriate selection depends on the claims that are to be made and

Dr. Draelos is Consultant and Researcher for the Skin Care Industry and Primary Investigator, Dermatology Consulting Services, High Point, North Carolina.

The author reports no conflict of interest in relation to this article.

the unique strengths of the formulation. Sometimes the market leader is not a well-formulated product, which may cause the study product to appear artificially efficacious. However, if the purpose of the study is to validate the usefulness of a new ingredient complex, a vehicle-controlled study must be performed. A good scientist wants to isolate the single active ingredient. There is no doubt that the formulation provides certain benefits only when used as a whole, but I would argue that the problem with most cosmeceutical studies is the failure to compare the "new" ingredient with the vehicle. Are vitamin C products effective because they contain vitamin C or because of the moisturizing vehicle in the vitamin C product? A good clinical cosmeceutical study should answer this question conclusively.

Why Do My Patients Not See the "Up to 60% Improvement" Observed in the Study Data?

Many seemingly well-designed studies produce results that do not translate into commonly observed phenomena. This is the case with the "up to 60% improvement" claim. This is an interesting statement because it sounds convincing, yet relates statistically irrelevant information. The "up to" claim means the observation was the absolute best achieved in the study, placing it statistically outside 2 standard deviations above the mean. For a more realistic result, the mean data should be presented along with the standard deviation. The subject with the best result may have been a poor enrollment choice not representing the average product user. Furthermore, if the sample size was small, such as less than 20, the standard deviation may be so large as to make the "up to" claim completely worthless.

Does the Collagen Regeneration Observed In Vitro Have Any Clinical Applicability?

Perhaps the biggest clinical cosmeceutical study pitfall I observe is when in vitro results are extrapolated to predict in vivo effects. Most cosmeceutical ingredients are initially studied in cell culture models. The cultured fibroblasts, keratinocytes, or melanocytes are exposed to the ingredient, and observations are made either visually or through gene chip array analysis regarding a beneficial effect. For example, many substances are touted as decreasing pigmentation. Is this due to the suppression of tyrosinase or a toxic effect on the melanocyte? Since effective pigmentation suppression must be long lasting for a clinically relevant effect, is it possible to maintain ingredient levels long-term in order to achieve skin lightening? It is easy to achieve high ingredient levels in a culture plate, but very difficult to do so in humans without experiencing other associated toxicities. Lastly, the stratum corneum is a formidable penetration barrier not present in cell culture work.

Does This Mean That Cell Culture Work Is Worthless?

No, cell culture work is not worthless. All cosmeceuticals should be first studied in cell culture. This is an inexpensive method of screening compounds that might be clinically valuable. However, the research cannot stop at the cellular level. The in vitro observations must be confirmed with actual human use. Sometimes this confirmation involves invasive tests, such as skin biopsies. Many would argue that cosmeceuticals should not be evaluated with skin biopsies, since they are not prescription medications, but I disagree. Cosmeceuticals can never become medically respected products unless properly studied. It is unwise for dermatologists to accept in vitro study validation for cosmeceutical efficacy.

Does a Good Product Earn Ratings of "Superior" From 100% of Subjects?

The design goal of some cosmeceutical studies is to achieve uniformly positive subject product assessments. I have read studies where 95% to 100% of subjects felt that a given formulation was the best they had ever used; all subjects reported tremendous wrinkle improvement of 70% to 80%. It is impossible to get a random sampling of people to unanimously agree on anything. These results are simply unbelievable, yet the study center will confirm their accuracy. This is due to the failure to select a random panel.

Many research centers unknowingly develop panels of subjects who report results not representative of the larger population. For example, a panel of subjects may be selected for its ability to judge facial moisturizers. These subjects may have completed more than 20 studies at the research center and are continually invited back to participate because of their positive responses. In time, the subjects become conditioned to answer questions regarding product performance and aesthetics in the manner desired by the researcher. This nonrandom sample leads to faulty data.

Another example of unreliable study data surrounds repeat-insult patch testing. This is a standard test performed on skin care products to determine if they produce undue irritation. I have studied products that had a completely clean repeat-insult patch test that caused irritant contact dermatitis in 25% of the subjects enrolled. Why did this happen? Many research centers specialize in repeat-insult patch-testing studies and maintain

panels of subjects who participate in this type of research. Imagine if you had enrolled in one of these studies as a college student to earn extra money and developed a horrible itchy dermatitis at the site where the patch tests were placed. It is unlikely that you would participate in another repeat-insult patch-testing study. Many of the subject panels developed over time by research companies are composed of people with extremely tough skin; therefore, topical products applied to their skin will not cause problems.

These examples point out the importance of proper randomized subject selection. The results of a cosmeceutical research study can be unreliable if a balanced panel were not selected. Research presenting unbelievable results should be further evaluated for appropriate panelist selection.

Is a Negative Study Valuable?

This brings us to the last topic for consideration: the value of a negative study. I believe that negative studies are just as important as positive studies and merit publication in the cosmeceutical realm. If a drug does not work during phase 3 of clinical testing, it will not receive approval from the US Food and Drug Administration, and the practicing physician will never see the medication in the marketplace. On the other hand, many cosmeceuticals containing ineffective ingredients are sold on a daily basis. It is for this reason that negative cosmeceutical studies are extremely valuable and should be published.

Negative studies are of value both to the dermatologist and the manufacturer. Unfortunately, one of the problems with publishing negative cosmeceutical studies is the right of the sponsor to withdraw support and prohibit publication. For monetary reasons, the sponsor does not want data released that might decrease product sales. Some manufacturers even state at study initiation that they want research designed to show that the product works. This is not an intellectually honest undertaking; however, I must say that most of the products that were launched despite negative study results from my research site were withdrawn from the market after 12 to 24 months because of poor sales. Instead of research studies designed to show that a product works, perhaps studies should be designed to see if a product works before further corporate investment.

Summary

There are many pitfalls in cosmeceutical clinical research. Some of these pitfalls will always be present. For example, many cosmeceuticals contain antioxidants that are intended to decrease oxidative collagen damage. A good antioxidant study would need to run for 10 to 20 years, since antioxidants treat what *might be*, not what *is*. A study on antioxidants would not even run for 5 years, since research in this area advances so quickly. Better products would become available before the study was complete. Improved short-term measurement methods for the evaluation of antioxidants are necessary before this research can be validated. Yet, in other areas, cosmeceutical research protocols that are intellectually honest can be designed. An intellectually honest protocol can only be designed if the pitfalls are recognized and understood. ■