FOOD AND DRUG ADMINISTRATION

CENTER FOR DRUG EVALUATION AND RESEARCH

MEETING OF THE

DERMATOLOGIC AND OPHTHALMIC DRUGS ADVISORY COMMITTEE

8:35 a.m.

Monday, November 4, 2002

Versailles Ballroom
Holiday Inn - Bethesda
8120 Wisconsin Avenue
Bethesda, Maryland

ATTENDEES

COMMITTEE MEMBERS:

ROBERT S. STERN, M.D., Chairman
Professor of Dermatology
Beth Israel Deaconess Medical Center
330 Brookline Avenue
Boston, Massachusetts 02215

KAREN M. TEMPLETON-SOMERS, Ph.D.
Acting Executive Secretary
Advisors and Consultants Staff
Center for Drug Evaluation and Research
Food and Drug Administration
5630 Fishers Lane, HFD-21
Rockville, Maryland 20857

ELIZABETH A. ABEL, M.D.
Dermatology
2660 Grant Road, Suite D
Mountain View, California 94040

ROBERT KATZ, M.D.
Dermatology
11510 Old Georgetown Road
Rockville, Maryland 20852

LLOYD E. KING, JR., M.D., Ph.D.
Professor of Medicine, Dermatology Division
Vanderbilt University
1301 22nd Avenue North
3900 TVC (The Vanderbilt Clinic)
Nashville, Tennessee 37232-5227

PAULA KNUDSON, Consumer Representative
Executive Coordinator for IRB Research and
Academic Affairs
The University of Texas
Houston Health Science Center
1133 John Freeman Boulevard
Jesse Johnes Library, Room 322
Houston, Texas 77030

ATTENDEES (Continued)

COMMITTEE MEMBERS: (Continued)

R. TODD PLOTT, M.D.
Acting Industry Representative (non-voting)
Vice President, Clinical Research
Medicis Pharmaceutical Company
8125 N. Hayden Road
Scottsdale, Arizona 85258

SHARON S. RAIMER, M.D.
Professor and Chairman
Department of Dermatology
University of Texas Medical Branch
Galveston, Texas 77550-0783

KATHLEEN Y. SAWADA, M.D.
Dermatologist
U.S. Naval Reserve Medical Corps
Alpine Dermatology Associates P-LLC
1785 Kipling Street
Lakewood, Colorado 80215

MING T. TAN, Ph.D.
Professor of Biostatistics
Room N9E36
University of Maryland Greenbaum Cancer Center
22 S. Greene Street
Baltimore, Maryland 21201

CONSULTANTS: (voting)

WILMA F. BERGFELD, M.D.
Head of Clinical Research
Department of Dermatology
Cleveland Clinic Foundation
9500 Euclid Avenue, Room A61
Cleveland, Ohio 44195-5001

S. JAMES KILPATRICK, JR., Ph.D.
Professor of Biostatistics
Medical College of Virginia
Virginia Commonwealth University
1101 East Marshall Street
Sanger Hall, Room B-1-039-A
Richmond, Virginia 23298-0032

ATTENDEES (Continued)

CONSULTANTS: (voting) (continued)

THOMAS R. TEN HAVE, Ph.D.
Department of Biostatistics and Clinical Epidemiology
University of Pennsylvania
School of Medicine
607 Blockley Hall
423 Guardian Drive
Philadelphia, Pennsylvania 19104-6021

GUEST SPEAKERS: (non-voting)

ALBERT KLIGMAN, M.D.
University of Pennsylvania
Department of Dermatology
226 Clinical Research Building
415 Curie Boulevard
Philadelphia, Pennsylvania 19104

HAROLD LEHMANN, M.D.
Division of Health Sciences Informatics
2024 East Monument Street, 1-201
Baltimore, Maryland 21287-0007

JAMES J. LEYDEN, M.D.
Professor of Dermatology
University of Pennsylvania
2nd Floor Rhoads Pavilion
3400 Spruce Street
Philadelphia, Pennsylvania 19104

PETER E. POCHI, M.D.
Emeritus Professor of Dermatology
Boston University
715 Albany Street, J-205
Boston, Massachusetts 02118

ALAN R. SHALITA, M.D.
Chair, Department of Dermatology
State University of New York
Downstate Medical Center
450 Clarkson Avenue
Brooklyn, New York 11203

ATTENDEES (Continued)

FOOD AND DRUG ADMINISTRATION STAFF:

MOHAMED ALOSH, Ph.D.
JONCA BULL, M.D.
BRENDA CARR, M.D.
MARKHAM C. LUKE, M.D., Ph.D.
JOSEPH PORRES, M.D., Ph.D.
JONATHAN WILKIN, M.D.

ALSO PRESENT:

JOANNE M. FRASER, Ph.D.

C O N T E N T S

C O N T E N T S (Continued)

P R O C E E D I N G S

(8:35 a.m.)

DR. STERN: Good morning, everyone. I'm Robert Stern. I'm chair of the advisory committee for dermatology to the Food and Drug Administration.

Today and tomorrow morning, we'll be working with everyone here to try to come up with the advice concerning six areas, as listed on questions, to help the FDA in its production of a draft guidance document on evaluating therapies for mild to moderate acne. So our purpose here is really to see how therapies for this class of acne are currently measured, learn about that, think about how which ones work well and poorly, and try to come up with suggestions about what are the best ways so that we can understand which agents are in fact effective, and then also how information about how effective and in what types of acne they're effective can be best transmitted to practitioners for drugs that are subsequently approved for this indication. So that's what we're trying to do.

I'm looking forward to it because acne is one of my interests, but certainly not my core interest, and I'm hoping to learn a lot today from our very august and learned speakers.

And I'd like to start with going around the room, starting on my left, if everyone would introduce

themselves and tell me and the audience a little bit about where they're from and what their background is.

DR. PLOTT: My name is Todd Plott. I'm from Medicis Pharmaceutical Company in Scottsdale, Arizona. I'm the Vice President of Clinical Research and Regulatory Affairs. I am the Industry Representative to the committee.

DR. ABEL: I'm Elizabeth Abel, Clinical Professor of Dermatology at Stanford University Medical School, and I'm in the private practice of dermatology in Mountain View.

DR. TEN HAVE: Tom Ten Have. I'm Professor of Biostatistics in the Department of Biostatistics and Epidemiology at the University of Pennsylvania. My collaborative experience has been more in the areas of psychiatry and disparities research focusing on clinical trials and issues regarding dropout and noncompliance, nonadherence. This is a new experience for me. I am also hopefully going to learn a lot here today. Thank you.

DR. KING: I'm Lloyd King. I am Professor of Dermatology at Vanderbilt University, and I'm a member of this FDA board.

DR. KILPATRICK: Jim Kilpatrick, biostatistics, Medical College of Virginia, Virginia Commonwealth University. I'm known as the joker of the pack, and so I'm

neither learned nor august.

(Laughter.)

MS. KNUDSON: That's a hard act to follow. I'm Paula Knudson, and I'm an IRB administrator at the University of Texas in Houston. And I've learned a lot already just by reading the material that was sent. It was fascinating.

DR. SAWADA: And I'm Kathleen Sawada. I'm from Lakewood, Colorado. I am a practicing dermatologist in private practice, and I am also a recent graduate -- or I like to think recent -- of the Medical College of Virginia.

DR. TEMPLETON-SOMERS: Karen Templeton-Somers, acting Executive Secretary to the committee, FDA.

DR. BERGFELD: I'm Wilma Bergfeld from the Departments of Dermatology and Pathology at the Cleveland Clinic, and I'm acting as a consultant to this advisory committee, and I've been previously on it for many years.

DR. TAN: I'm Ming Tan. I'm a practicing biostatistician and a professor of biostatistics at the University of Maryland School of Medicine. I've been with the committee for several years.

DR. RAIMER: I'm Sharon Raimer. I'm Professor of Dermatology at the University of Texas in Galveston and also a member of the committee.

DR. KATZ: I'm Robert Katz. I'm a practicing

dermatologist here in Rockville, Maryland, Clinical Assistant Professor of Dermatology at Georgetown, and a consultant at Walter Reed Army Hospital.

DR. CARR: I'm Brenda Carr. I'm a medical officer in the Division of Dermatologic and Dental Drug Products, FDA.

DR. WILKIN: Jonathan Wilkin. I'm Director of the Division of Dermatologic and Dental Drug Products, FDA.

DR. BULL: Good morning. Jonca Bull. I'm the Director of the Office of Drug Evaluation V.

DR. TEMPLETON-SOMERS: The following announcement addresses the issue of conflict of interest with respect to this meeting and is made a part of the record to preclude even the appearance of such at this meeting.

Since the topics to be discussed at the meeting will not have a unique impact on any particular product or firm, but rather may have widespread implications with respect to an entire class of products, all committee participants have been screened for interests in products indicated for use in the treatment of acne vulgaris and their sponsors.

In accordance with 18 U.S.C. 208(b)(3), Dr. Thomas Ten Have and Dr. Robert Stern have been granted particular matter of general applicability waivers which

permit them to participate fully in the matters at issue.

A copy of the waiver statements may be obtained by submitting a written request to the agency's Freedom of Information Office, room 12A-30 of the Parklawn Building.

Because general topics impact so many institutions, it is not prudent to recite all potential conflicts of interest as they apply to each member and consultant.

FDA acknowledges that there may be potential conflicts of interest, but because of the general nature of the discussion before the committee, these potential conflicts are mitigated.

With respect to FDA's invited guest speakers, there are reported interests that we believe should be made public to allow the participants to objectively evaluate their comments.

Dr. Albert Kligman is a consultant and scientific advisor for Allergan, Dermik Laboratories, and Medicis Pharmaceutical, and receives $10,000 annually from each company for his services. He also owns stock in each firm.

Dr. Peter Pochi owns stock in Pfizer.

Dr. James Leyden has participated in clinical trials, served on advisory boards, given lectures, served as a consultant, and received research grants from Bertek

Pharmaceuticals, Dermik Laboratories, Pharmacia and Upjohn, Galderma, Medicis Pharmaceutical, Lederle Laboratories, Oclassen, and Ortho Dermatologic.

Lastly, Dr. Alan Shalita owns stock in Johnson & Johnson, Medicis Pharmaceutical, and Allergan. In addition, he is a researcher, consultant, and scientific advisory for Allergan, Medicis Pharmaceutical, and Stiefel. He is also a consultant and scientific advisor for Dermik Laboratories and a researcher for Johnson & Johnson. Lastly, he lectures for Galderma, Dermik Laboratories, Medicis Pharmaceutical, and Allergan.

We would also like to note for the record that Dr. R. Todd Plott is participating in this meeting as a non-voting acting industry representative, employed by Medicis Pharmaceutical Company. Medicis Pharmaceutical is one of the many firms which could be impacted by the committee's discussions.

In the event that the discussions involve any other products or firms not already on the agenda for which FDA participants have a financial interest, the participants' involvement and their exclusion will be noted for the record.

With respect to all other participants, we ask in the interest of fairness that they address any current or previous financial involvement with any firm whose

product they may wish to comment upon.

Thank you.

DR. STERN: We'll begin this morning with the open public hearing. Dr. Fraser from Stiefel Research Institute.

DR. FRASER: Dr. Stern, members of the committee, FDA representatives, and invited guests, good morning. My name is Joanne Fraser. I'm the Director of Research at Stiefel Research Institute which is the research arm for Stiefel Laboratories.

This presentation concerns the use of acne lesion counts in clinical trials.

Acne vulgaris is characterized by the presence of papules, pustules, open and closed comedones, nodules, and cysts. In clinical trials, investigators are asked to count inflammatory lesions and non-inflammatory lesions. A total lesion count is then calculated as the sum of the two. Total lesions is used in an attempt to represent the patient's overall acne condition.

In this presentation, I hope to convince you that the variable, total lesions, is not useful in assessing the efficacy of acne products and can lead to misconceptions about efficacy.

In determining the treatment for a patient with acne vulgaris, the types of lesions present is an important

factor. There are specific drug products to treat inflammatory and non-inflammatory lesions, and there are some agents that affect both. These lesions are physiologically different and respond to drugs differently.

Currently the requirements for an approval for a drug product for the indication of acne vulgaris are that a significant difference from control be shown for two out of three lesion types, inflammatory, non-inflammatory, and total, and global severity. So where the circles are intersecting represents meeting the requirement of two out of three.

If a product is only active for the treatment of one type of lesion, then the only requirement for approval should be for that lesion type, plus global. There is a concern that the patient's overall acne should look better as a result of treatment, and therefore if the total lesion count improves, there's some assurance of the overall effect. But global severity could be used to address this concern. Using total lesions for this purpose adds no information about the efficacy of the product and can lead to misconceptions about efficacy.

This was a study of a combination product. The results of the combination, each of the single agent controls and vehicle are shown for inflammatory lesions, non-inflammatory lesions, and total lesions. The use of

total lesions has no advantage over the separate analysis of inflammatory and non-inflammatory lesions. In many cases, the percent reduction of total lesions is essentially the average of the percent reductions of non-inflammatory and inflammatory lesion counts.

This slide shows hypothetical data for two subjects. The first subject has more non-inflammatory lesions and the second subject has more inflammatory lesions. The percent reductions for inflammatory and non-inflammatory are the same for each subject, 60 and 20. For subject 1, percent reduction for total lesions, 30, is similar to the non-inflammatory lesion percent reduction, 20, the more numerous lesion type. For subject 2, total is closer to the inflammatory percent reduction, the more numerous lesion type. In a study of subjects similar to subject 1, a large reduction in inflammatory lesions is canceled out in the total lesion percent reduction because of the small change in non-inflammatory lesions.

This slide shows two subjects from one of our clinical trials. The entry criteria was at least 25 inflammatory lesions and 12 non-inflammatory lesions. In a subject with both inflammatory and non-inflammatory lesions, non-inflammatory lesions are usually more numerous. In our clinical trials, approximately two-thirds of subjects have had more non-inflammatory than

inflammatory lesions despite similar entry criteria. For these subjects, the percent reduction of total lesion count is similar to the percent reduction for non-inflammatory lesions, the more numerous lesion type. For subject 2, substantial efficacy for inflammatory lesions was canceled out in the total lesion variable because of no efficacy in non-inflammatory lesions. Applying the rule of two out of three, a product with results like for subject 2 would not be approvable even though it has substantial efficacy toward inflammatory lesions. The product with results like subject 1 might be approvable for acne vulgaris with only modest efficacy for inflammatory lesions.

This slide shows two more subjects. The first subject has more inflammatory lesions than non-inflammatory lesions. The same is true, that the percent reduction for total lesions is similar to the lesion type count that is more numerous. Subject 2 has approximately equal numbers of inflammatory and non-inflammatory lesions, with substantial efficacy for inflammatory and modest efficacy for non-inflammatory. Percent reduction for total lesions is approximately the average. The exact average is 59.

This is data from a recently approved product. All three lesion types were significantly different from the vehicle control for percent reduction. The total lesion count data adds no information about the efficacy of

the product. This product was approved for the treatment of acne vulgaris.

This is data from the first of two studies from a recently approved product. In this study, all three lesion types were significantly different from vehicle control. Again, the total lesion count data adds no information about the efficacy of the product.

This is the data from the second study for this product. In this study only inflammatory and total lesion counts were significantly different from the vehicle control. The use of the total lesion count data masks the lack of efficacy for non-inflammatory lesions. This product was approved for the treatment of acne vulgaris because it met the two out of three lesions requirement and global for both studies. Perhaps this product would have been more accurately labeled for treatment of inflammatory acne based on these studies.

This data is included in the package insert which is then available for the clinician to decide for themselves how best to use this product, but regardless of the indication, it seems useful to include all the data on the labeling. But again, total lesion data does not add any real information.

Two products were recently approved, both containing the same active ingredients at the same

concentration. Product A was approved for inflammatory acne, and product B was approved for acne vulgaris in general.

Five studies were completed for product A and two studies were completed for product B. Here are the percent reductions in inflammatory lesions for each product. They are quite similar in the effect on inflammatory lesions. And here are the percent reductions in non-inflammatory lesions for each product. Again, the results are quite similar. And here are the percent reductions for total lesions. Again, very similar.

As these products were combination products, the control of interest and challenge to find a statistical difference was the comparison to the benzoyl peroxide alone control. For product A, three of five studies showed a significant difference compared to BPO, and for product B, both studies showed a significant difference for inflammatory lesions.

This is the difference for the non-inflammatory lesions. Neither product is more effective than benzoyl peroxide for the treatment of non-inflammatory lesions. The labeling for product A, which was approved for the treatment of inflammatory lesions only, has a statement that the product is not more effective than benzoyl peroxide for the treatment of non-inflammatory lesions.

The labeling for product B does not include the same statement.

And the reason product B was approved for acne vulgaris is the differences for total lesions compared to benzoyl peroxide. The differences are significant in both studies for product B and in only two of five studies for product A. The results of the total lesions has masked the lack of effect of product B for non-inflammatory lesions compared to benzoyl peroxide.

The data in the previous slides were for the comparison to benzoyl peroxide control since those were combination products, but both products have substantial efficacy compared to vehicle for inflammatory lesions and for non-inflammatory lesions.

In summary, product A was approved for inflammatory acne only. It did not meet the two out of three requirement when compared to benzoyl peroxide. An exception was made for the indication of inflammatory acne. Product B met the two out of three rule with inflammatory and total when compared to benzoyl peroxide and so was approved for the indication, acne vulgaris. Both products were effective against both types of lesions compared to vehicle or clindamycin.

The labeling for product A includes percent reduction results for inflammatory lesions and the

statement that the product is not more effective than benzoyl peroxide for the treatment of non-inflammatory lesions. The labeling for product B includes the percent reductions for all three lesion types. There is no statement about product B not being more effective than benzoyl peroxide for the treatment of non-inflammatory lesions. And the difference in labeling for these two products with essentially identical activity is due to the results of the derived variable, total lesions. Use of the variable, total lesions, has masked the lack of effectiveness of product B for non-inflammatory lesions compared to benzoyl peroxide.

In conclusion, we need the option of three target lesions for products to treat acne, inflammatory, non-inflammatory, and acne vulgaris when a product is effective for both. And I hope I've convinced you that total lesions is not a useful variable in assessing the efficacy of an acne product.

Thank you.

DR. STERN: Could I just ask you one question?

DR. FRASER: Sure.

DR. STERN: Or two questions. One is, are you then saying that you're advocating that products, when they go to phase III, there should be an advance hypothesis that we will prove efficacy for inflammatory acne or non-

inflammatory acne or both, and if it's for both, is it going to be that unless you get it for both, the product is not approved? Or are you advocating that if you say we want to do this for both and it only makes criteria by one, that in fact, since you put forward three hypotheses, that there be some correction, some change in the requirements of the p value for multiple comparisons?

So those are sort of two related questions. The first is, do you just pick one of the three indications and you've got to go with that to the end, meet the criteria statistically? The second, if you're going to allow a fall-back by another criteria other than the one you put forward, how are you going to correct for the multiple comparison problem?

DR. FRASER: Right. I believe that's correct that if you set your hypothesis just for one lesion type when you're going into the study, that would be the best way to do it, but if you want the option of either one, you're going to have to adjust for that statistically.

DR. STERN: Any other questions from the committee?

DR. KILPATRICK: Thank you, sir.

It seems very obvious to me that since total equals inflammatory plus non-inflammatory, total depends on these two. Therefore, from a purely statistical point of

view, you can only have two of these three things, whatever they are. So it was a given to me, before you started, that you use either inflammatory or non-inflammatory because total is the sum of the two. I mean, it's so obvious.

DR. FRASER: Right.

DR. KILPATRICK: So I don't know what the fuss is about. But Dr. Stern asked the difficult question.

DR. TEN HAVE: Isn't there also a multiple comparisons problem with the current approach, if you're choosing two out of three?

DR. FRASER: Right. Currently there's no statistical adjustment for the multiple --

DR. TEN HAVE: Comparisons problem with the current --

DR. FRASER: Right.

DR. STERN: Thank you very much.

Is there anyone else who would like to comment during the open public hearing?

(No response.)

DR. STERN: Seeing no one who wishes to do so, we will go on to Dr. Jonathan Wilkin who will give an introduction to why we're here today and tomorrow.

DR. WILKIN: Well, we are here today because there are over 50 million people in the United States with

acne and many of these are adolescents and young adults. The burden of acne, especially in this population, the physical, the psychological, the quality of life issues, impels the public health need for safe and effective products for acne.

What we're asking the committee to consider today and tomorrow is how should we look at the evidence for effectiveness of these products in a way that we can craft this into a guidance document so that industry and academics and the regulatory folks at FDA can all be working from the same page.

To help the committee in thinking about the six questions, which I should say are actually essay questions, not yes or no questions, we have multiple speakers. We've asked Dr. Bergfeld who, as she mentioned, is an alumna of DODAC, to give an overview of acne, and the dermatologists always gain something from her insights, but especially helpful I think will be for the statisticians and others on the committee who might need an acne 101 so that they know what the different lesion types are.

I'll follow up with sort of an historical view of how FDA has viewed the two primary efficacy endpoints of lesion counts and global and also give some work that I did before I came to FDA which actually looks at the relationship between acne counts and global.

And then the speakers who follow immediately will be primarily talking about the global severity scale, Dr. Carr, Dr. Pochi, and then Dr. Leyden, Dr. Shalita, and Dr. Kligman will be talking about severity scales but also lesion counts and what their views are.

One of the important aspects of all of this is not just what the primary efficacy endpoints might be but how do we analyze the data, what are the statistical methodological issues, and Dr. Alosh will be presenting that.

Dr. Luke will speak to some of the interesting aspects of combination topical products and how we look at efficacy.

And then we will end up the FDA's portion with Dr. Porres describing what kind of information gets crafted into the package insert which describes efficacy outcomes, and we'll be asking the committee for suggestions on how we might improve that to better convey to the clinician and to the patient and to improve the patient-clinician communication on what might the expectations be for acne therapy.

Then finally this afternoon Dr. Lehmann, who has conducted research under a contract to the Agency for Health Care Research and Quality, which is a sister organization in our Department of Health and Human

Services, will have some thoughts on how to get some useful information out of acne trials that might even be in addition to what we're going to talk about earlier in the day.

And then we're looking to tomorrow to actually have the questions deliberated.

DR. STERN: Thank you very much.

Now I'm very pleased to have Wilma Bergfeld speak to us about acne.

DR. BERGFELD: Thank you very much. I'm delighted to be back at the FDA. I always love coming back. This is a very important committee activity.

What I've been asked to do is to paint a picture of acne today and perhaps reflect a little bit about what was going on yesterday.

It's important to realize that acne represents 4 percent of all dermatological disease and it, as you heard, involves a population group that is very large, basically 50 million. This represents the demographics of acne, mainly a disease of youth, as you can see here in the white, 12- to 24-year-olds representing 40 million plus, whereas 25- to 35-year-olds, about 3.5 or 3.8 million, and a very large growing group is the adult group which is usually women 35 to 44 years of age.

Now, you heard from Jonathan Wilkin that it is

very important that we address acne, being a major disease for us in dermatology and as a health issue, but also it's very important because of the psychological and economic impact. There have been numerous studies done over the last 20 years that display that those who have moderate to severe acne greatly suffer in their life, psychologically as well as economically. You will note here that they have reduced self-esteem, confidence, and body image, which then reflects in their ability to perform, to reach the essence of their life and their desires for success, but it also limits their lifestyles, their interpersonal relationships, and interestingly enough, has been noted to reduce their employment. They're more unemployable. And certainly adults are more affected than the young, but all are affected.

Now, the problem that we see today in dermatology is that there's a growing desire for the patient, the parents of the patient to reach dermatologists, and there's a growing need for more dermatologists to be in practice. And this is reflected by patient preference as well as the growing addition of dermatologists to a variety of HMOs and other medical groups. And patients now have great access to dermatologists through a variety of a different health care programs. So we are seeing that acne is one of our number

one diseases to treat. We are seeing a growing population that's affected, one that is growing in its age as well, and also the fact that we do not have a great enough work force to take care of these patients.

What we know about acne. Again, here is another graph or table demonstrating it is a major disease for dermatologists, but there are other physicians who care for the disease, but the dermatologists are the key caretakers.

Now, the acne classification is rather classic. comedones, which is blackheads, papulopustules, which are erythematous papules and pustules, and then cysts. And the dermatologists have classically defined these as being mild, moderate, and severe and also include the sites of involvement, which are usually face and trunk and occasionally arms and buttock.

I'd like to show you a number of pictures of mild to moderate acne and then end with some very serious forms of acne. This is a comedonal acne in an African American black young athlete showing both blackheads, comedones, as well as inflammatory papules.

A caucasian with comedones and milia which are closed comedones, whiteheads, around the mouth, cheeks with cheek scarring.

An Indian young woman demonstrating a number of

features, namely hirsutism as well as acne, with inflammatory papules and scars on the cheek.

A little less well demonstrated here, but a lot of inflammatory lesions on the cheek and around the chin.

A male demonstrating the inflammatory form of acne and the classic distribution on cheeks and chin.

A cystic form of acne in a little bit older individual who has excoriated these lesions.

And a more severe form which is the erosive pustular form which is a very serious disorder for us.

Now we know that acne affects almost all age groups and it certainly has been noted in the neonate. Usually they are comedones and they're non-scarring. In the young infant, especially the male infant, we can see papulopustular lesions. These do leave scarring, and the teenage acne usually is face and trunk and is male dominant and it can induce scarring. And now the adult acne which is mainly in females, but males do also have this, and this is a late onset usually or it can be chronic from teenage through their mid-years up to about 60.

Now, it's important when a dermatologist or a physician sees a patient with acne, that they take the appropriate history. There's no doubt that it's familial. We do see it run in families. It's important for us to examine the patient and ask some very pertinent questions

around family history, as well as androgen excess and diabetes.

As you've already heard, we do do lesion typing as well as location of lesions, and we do grade these acne patients. This then evolves into developing therapeutic options, which are discussed with the patient, along with the adverse events that might occur, as well as the expectation, and the therapy is then given.

Now, the therapy is aimed at a variety of different areas of the acne pathogenesis, namely getting rid of the blackheads and whiteheads which are thought to be the primary lesions, especially what we call the microcomedones, getting rid of the microorganisms that live in these lesions, getting rid of the inflammation. And a group of these, at least one-third of these patients, especially the female, have androgen excess, and they have androgen stimulation of the sebaceous gland which then induces or exaggerates the acne. And certainly external irritants can either worsen the acne or, in some instances, can actually induce acne.

Now, if we look specifically at how we do this and why we do this, we want to get right of the P. acnes because it produces inflammatory lipids, which are fatty acids, which then release cytokines. We want to get rid of the inflammation because there is a cascade of cytokines

which then ends up with tissue destruction. We attempt to get rid of the keratinizing defects which are in the hair follicle canal way plugging the follicle, thus inducing the blackhead, the micro-blackhead. And we also want to reduce the size and function of the sebaceous gland from putting out its oil, or sebum. And we certainly want to reduce, when present, the hormonal influence on the oil gland, the sebaceous gland, and in doing so, we can improve the acne. So as you can see, when we look at all these various targets, we may be using multiple therapies to achieve this end.

So what might we use for the blackheads, whiteheads, or even milia, which are the closed blackheads? We would use a variety of agents, the retinoids being the leading ones usually used topically. They can reduce the size and the function of the oil gland, reduce the microorganisms, reduce the inflammation. Benzoyl peroxide can be used as well, which has similar effects in reducing the organism. And there are a number of other acids, both fruit acids, natural acids, that can be used for similar purposes.

When we're looking at inflammatory acne with papules and pustules, however, we're looking at using more, I guess, important drugs in some aspects in the fact that they're mostly antibiotics and they may also include the

use of oral Accutane. But for antimicrobial, we can use the benzoyl peroxide agents because they certainly do have some activity in that area, as well as some of the natural topical acids, but we do use commonly topical antibiotics in the form of erythromycin and clindamycin, and we also use oral antibiotics in the form of minocycline, tetracycline, and more recently zithromycin.

We use, as I said, oral and topical retinoids.

We also use, in very severe forms, anti-inflammatory agents which would include corticosteroids in the very, very severe forms of this disease.

We do also use anti-androgens to reduce the testosterone or androgen effect on the oil gland, and these would fall into groups such as estrogens in the female, spironolactone, and flutamide. Mainly those are used in the female.

We also identify in this group, especially the female, an androgen excess syndrome related to insulin resistance, and this leads us into other therapies such as metformin.

And we can also use vitamins and minerals for some of their anti-inflammatory as well as anti-androgen activity.

Now, the tretinoin effects. I'd just like to go over them because they are so broad and affect many of

the targets that we need to hit. We can reduce the scaling that occurs in the hair follicle which plugs it up. We can alter the microorganisms by reducing them. We can resolve the early comedones and the microcomedones, the milia, with these particular agents. We can prevent new lesions, and we can enhance, which is very important, penetration of other drugs.

Now, here is the list of the topical retinoids that we do have available to us, and as you can see, there are numerous ones and they come in all concentrations and vehicles, all of which assist us in treating topically these microcomedones and comedones.

Now, when we look at their efficacy, using two different ones -- not to discuss their comparison, but using two different ones -- adapalene and also Retin A, we can see that we can get greater than 50 percent reduction of lesions, which is very important. You can see that some are better at inflammatory and some are better at non-inflammatory, but the bottom line is that they reduce greater than 50 percent the inflammatory and non-inflammatory lesions.

But we also have a problem with topical retinoids in the fact that they are irritants, and we have had a hard time reducing the irritancy of these because over time, using these two same drugs, we can see that the

irritation is about the same. And irritation, as I mentioned, is, one, painful but also it can induce more acne.

Using some of the natural acids -- and this happens to be one, dicarboxylic acid -- we can also have some effect on bacteria anti-inflammatory activities, as well as reducing keratinization. So we have other options other than the tretinoins, but the tretinoins have been our base therapy.

As I mentioned, antimicrobial therapy would include benzoyl peroxide. It is a potent bactericidal agent. We also use it as an agent that kills all in my practice. And you can use it up to 10 percent, and it can reduce blackheads and also papules and pustules. It reduces the infectious agent P. acnes, but it also can induce irritation to the skin. And that reflects in dryness and pain, scaling. We use topical antibiotics, again erythromycin, clindamycin, specifically for the same reasons, and oral antibiotics.

This is a study done very early by Kligman, and this demonstrates the activity of benzoyl peroxide on P. acnes in red, reducing it basically 60 percent plus, as well as the fatty acids which are produced by the sebaceous gland. So it is an effective therapy too.

Now, one of the problems that we've had and, in

fact, discussed here at the FDA is the bacterial resistance to some of the antibacterial agents that we use in dermatology, and this is a growing problem for us today in practice because we are having more patients present to us who fail to respond to what we consider our basic regimens and this is something that we're striving to overcome.

Now, I wanted to touch very briefly on androgen activity because the circulating, as well as the androgens present in the tissue and the target organ, namely the sebaceous gland and the hair follicle, do stimulate acne. We know that the sebaceous gland in particular has androgen receptors. So using anti-androgen therapy selectively in both males and females can be exceedingly helpful, especially in the more resistant forms of acne.

Now, there have been some studies, and the classic studies have been looking at circulating androgens. And one done by Lucky in the 1980s demonstrated that females with very persistent papulopustular acne had elevations of free and total testosterone and less commonly elevated DHEAS, which is an adrenal androgen.

This followed a study done by Ortho regarding the Ortho Tri-Cyclen that's used in acne in females, and this was a study in 250 female acne patients with moderate acne. What it demonstrated was that 83 percent versus the control which had 63 percent improvement -- that 83 percent

improvement of acne was seen in this study. When measuring circulating androgens, it was noted that the testosterone levels were reduced. As I just previously mentioned, these testosterone levels are elevated in some of these acne females. And there was also an increase in sex-binding hormone which is important because it binds the testosterone.

At the Cleveland Clinic, we too have studied androgens and androgen excess presentation, one being acne. And we noted that it was common for us to have elevations of total and free testosterone, as well as the adrenal androgen. And the reason for pointing this out at this time is that testosterone can be made by either the ovaries or the adrenal gland, and the birth control pills would affect mainly a suppression of the ovarian testosterone. However, if the acne was stemming from the adrenal gland, one would have to suppress the adrenal gland as well.

So, hormonal therapy is generally reserved only for females, and we use a variety of therapies, namely the low dose birth control pills. We can use anti-androgens in the form of spironolactone, and we can use corticosteroids, especially if the adrenal gland is involved. We also have the opportunity in selected patients of using Accutane. It is more commonly used today in males than females for this form of acne. And we also would be using anti-

inflammatories because this is an inflammatory disease and one needs to also address that.

So when we look at the therapeutic options that we have in acne, one has to address the fact that we are after multiple targets that induce the final lesion. So we have a number of agents that fall under getting rid of the blackhead or the whitehead, or the milia, the closed comedone, and these include the retinoids, benzoyl peroxide, sulfur, and some of the natural acids.

We have a number of agents that we have available to reduce sebum, or oil production by the oil gland, namely the retinoids, the anti-androgens, the low dose birth control pills, and we could add corticosteroids here.

We have agents to reduce the main organism that produces acne. At least in our belief it produces acne. And there are a variety of topical and oral antibiotics, the retinoids, benzoyl peroxide.

And the inflammation can also be reduced by oral antibiotics and retinoids.

Now, what we are looking at today is the fact that because of the bacterial resistance, we are looking towards what are the effects of combining benzoyl peroxide with a number of antibiotics, and they seem to be very good. In fact, not only are they combined with oral

antibiotics, but also zinc. So this is the future for us in dermatology, at least in the topicals, because of bacterial resistance. There is very little resistance to benzoyl peroxide, in fact, none to date, but there is resistance to erythromycin and the tetracycline-like products. So combining them, we then get rid of our resistance.

Now, what is important to us in dermatology is the fact that no one gets better with one or two prescriptions, go off, and come back never again. We need to see these patients again and they need to understand what's going on with their disease, why they have it, and why we are giving certain medications.

They also need to know what the time frames are for improvement, and certainly we never promise anyone any marked improvement under a couple of months.

And they need to know that their therapies might be changed on each visit depending on what their clinical response is and what their skin irritation is. So each time a patient returns, their therapy is reevaluated.

We also need to have patient compliance.

Now, patient compliance is important because most patient, if you give them a load of prescriptions aimed at a variety of these targets, will not do any of it or do too much of it. So it is an active agreement that

the physician dermatologist has to have with the patient as to what they will do and what you want them to do, and somehow you have to mesh these choices so that there is something active being given to this patient to improve their acne.

It's important for physicians, as well as parents, to remember that no one can remember more than three things. So you need to write down instruction, or greater than that, we need to have patient educational materials for both the parent as well as the young person, and we need to provide written instructions for our patients.

Now, what I see as the acne treatment pitfalls is not just the diagnosis, not just establishing the therapies, but if the visit is too quick and the educational piece is not given, as well as the instructions, and the compliance pledged. I also see a problem in over-treatment. When there is too much skin pain and irritation from the therapies, the patient is not compliant. And then we have the problem of giving therapies that are non-compliant with the lifestyle of the patient.

So what does the patient do? He gets irritated if he overwashes, too many medical facials, too many medications, lack of education, and fear of the therapies.

And certainly there are patients who want to get better with no therapy.

So we the dermatologists, specifically the dermatologists, have a real medical problem that faces us with acne. This is not just a superficial disease and a cosmetic problem, but this is a profound disease that needs attention. And as you can see, it has many aspects of both diagnosis and therapy, follow-up, compliance, and safety.

So thank you.

DR. STERN: Thank you very much, Wilma.

Our next speaker will be Dr. Wilkin who will speak to us about evidence of effectiveness of acne products.

DR. WILKIN: Many years ago I participated in an acne trial as an investigators, counted lesions, and I noticed that at the end of the trial, that the lesion counts by themselves didn't seem to actually be as meaningful as what the global looked like or what the patients felt they had accomplished in the trial. Their sense of how better their acne got actually seemed to me to be related to the global and not directly, at least all the time, to the difference in lesion counts.

So I thought about this for over a decade, and it seemed like a paradox, at least to me. How could you have a system that inherently had a lot more information in

it -- that is, all these different lesion counts and very precise, very unbiased, very accurate -- how could that really not have as much clinically meaningful information as just the simple 0 to 4-plus subjective ordinal scale, sort of an estimate?

Now, acne is too complex to ask the question about how this would happen with all the different kinds of lesions.

So I chose a model. And a model, when you're going to look for mathematical relationships, is the system that has the relevant properties, but only those properties, and everything else has been removed. So it's an oversimplified model. It doesn't have many of the things that we look at when we're looking at acne severity like halos of erythema around the inflammatory lesions. It doesn't have the different size kinds of lesions. It doesn't have elevation.

So that's why you'll see acne in quotes because what I chose to do is to have acne lesions literally painted on faces of human models who didn't have acne so that I could characterize the relationship between the actual number of these painted-on lesions and the perceived severity of the acne lesions. Since again, there was no variation in the size and morphology of the lesions, what really is perceived severity is judged numerosity. How

numerous did the lesions appear?

So to do this, we recruited 33 research subjects who were the evaluators. They came into a dark room and looked at kodachromes of two models, and the models had lesions painted on their face for acne severity. The two models had up to 200 of these acne lesions painted on their face by a professional theatrical cosmetic artist. And then the research subjects, the observers, looked at these kodachromes and scored on a 10 centimeter linear horizontal visual analogue scale what they thought was the acne severity. And the visual analogue scale was scored by digimatic calipers which are quite precise.

This is the visual analogue scale. You can see here where if this were one of the research subjects marking it, they would have marked a 35-millimeter deflection from clear, and so that would be one-third as bad as the acne could be.

So this is the basic paradigm of the study. The input is the actual number of the lesions that have been painted on by the theatrical cosmetic artist. The test subjects are the human subjects that came in and looked at the kodachromes. And then their mind processed it, and then they wrote on a horizontal linear visual analogue scale. They made a mark which was the judged numerosity, if you will, of the acne.

This was the first model they looked at. This was stated as clear.

And this was stated as bad as can be. It was intended that there would be only 100 lesions, but it turned out the cosmetic artist was not majoring in mathematics and there are actually 101 if you count them all.

I'll only show a couple. I won't show you all 48 slides.

This is nine. If you look at it, you can actually count that.

Next is 49.

Now, for the committee, there's going to be a quiz after this. So I'll show you the anchors at the beginning. This is clear. This is as bad as can be, which is in this case 200. This is 50, 100, 20. Okay. Here's your unknown. How many think there is less than 150 lesions here? How about more than 150 lesions?

(A show of hands.)

DR. WILKIN: Actually there are 120. So there is a nonlinearity.

What we have here, the output is judged numerosity, and so it is the millimeters of deflection on the horizontal visual analogue scale, again, of judged numerosity. The input is the actual number of lesions

painted on the face. So you can see we've got two series. The blue line is the subject that had from 0 to 200, and the yellow line is the subject that had 0 to 101.

What we're showing on this slide is input, which is the actual number of lesions painted on the face and seen on the kodachromes, given as a fraction of the maximum input so that we can bring the 101 and the 200 into the same kind of scale. And then judged numerosity is likewise presented as millimeters of deflection from clear or 0, represented as a fraction of the maximum judged numerosity, or as bad as it can be.

What we've done on this slide is we've added some very fine lines. Those I think at the table may be able to see these. So we've broken up this curvilinear relationship into three segments, and I would just point out that in this segment, you can see that for every increase in lesion count, you actually get twice as much impact on judged numerosity. If one is up in the range above one-half maximal lesion count that is painted on the face of the subjects, then in that range you get only half of the judged numerosity for each increased number of lesions at the upper end.

Now, the one thing that's been added to this slide is that the output domain, judged numerosity, has been broken up into an ordinal scale so that this would be

4 plus, 3 plus, 2 plus, 1 plus, and 0. What you can see is that for the maximum number of lesions painted on the face, if you reduce that in half, that is appreciated by the human subjects who were judging numerosity as a drop in one grade, so from, say, 100 lesions to 50 lesions. That's a drop from a grade 4 to a grade 3. If you go from 50 lesions to 25 lesions, which is another half drop, then that's going from a grade 3 to a grade 2. And if you go from 25 lesions to about 10 lesions, that's again approximately a drop in half, and one drops another rank on the ordinal scale.

So what I believe this to be is that the ordinal scale is actually an empiric attempt at a ratio scale, and we know that that is sort of the psychometric wiring of the human mind. That's what happens with decibels when one is considering loudness. It's not really a linear function. It's a logarithmic function. When one goes down 10 decibels, you're reducing loudness literally by 90 percent.

Likewise stellar magnitude. You go out at night. You look up at the constellations. You see first magnitude stars the brightest and so on down to sixth magnitude. It's not equal differences in terms of the photon energy coming in the starlight. It's actually a ratio function.

So, I think this is the way people look at acne lesion severity, at least the part of judged numerosity, in a manner that is a cognate of stellar magnitude and the decibel system.

Having said that the psychometric model provides a curvilinear relationship between the more clinically relevant acne global severity scale and the more precise acne lesion counts, I would like to come back and again emphasize the disclaimer I gave at the beginning. I've stripped away an awful lot of the reality of acne. I've taken away the difference in size, the many different kinds of lesions. Certainly inflammatory lesions have more of an impact on judged severity than non-inflammatory lesions. Some have that erythema halo. So again, I'm not offering this as a very simple way of looking at real acne, but I think this relationship, nonetheless, exists. It's probably too complex to ever convert acne lesions per se into a global, and Dr. Alosh will mention that later.

Now, I did this about three years before coming to FDA. Once I came to FDA, I learned from the people who were already at FDA, in the usual oral tradition, how they had looked at acne lesions. I learned this from the clinicians and the statisticians that were on the team.

So I'm describing actually what was happening before 1994 when the division was created, and as my

colleagues at FDA know, I refer to that as the paleo-regulatory era. I can't really give all of the discussions that happened at that time, but it is clear that the folks at FDA and industry were using lesion counts which was total plus either inflammatory or non-inflammatory, and also an investigator's global assessment, which early on sometimes wasn't dichotomized into a success and non-success, but more frequently later on was dichotomized into a success and non-success.

Over time the total became, I think, changed to two out of three, that is, the total, the inflammatory, and the non-inflammatory, because it was thought that if you won with two out of three, one of them was going to be total. It would be pretty hard to win on inflammatory and non-inflammatory and not win on total.

What I learned from the statisticians and clinicians of '94 and '95 is that they viewed the lesion counts to be more accurate, more objective, harder data, if you will, I think was the line. The investigator's global was imprecise, subjective, might vary among investigators, especially with some of the less morphologically defined global scales.

And then over the last decade, we've seen a lot of differences in the NDAs that have come in. We've seen very different baseline lesion counts from one study to

another, even within the same sponsor's package. We've seen different lesion count analyses. Dr. Alosh will be talking about this. We've seen absolute change studied in some, percent change, a whole variety of transformed values, and then also a lot of different global investigator scales.

So we'd like to have one consistent way where we can approach the evidence for effectiveness for these acne products, that is, the mild to moderate kind of acne vulgaris products.

And our first question to the committee will be, should the current success criteria using co-primary endpoints be retained? Of course, that's not meant really to be a simple yes or no because if the answer is no, we'd like an essay question telling us how to fix it and which parts we need to preserve.

How should lesion counts be analyzed?

What investigators' global severity scale should be used? At what level should it be dichotomized?

I really cannot recall any sponsor initially coming in saying that they wanted only inflammatory lesions or non-inflammatory lesions of acne as their indication. All of the applications that I've seen, sponsors have come in saying that they want the indication of mild to moderate acne vulgaris as monotherapy. I think Dr. Bergfeld

indicated that while dermatologists may focus on different lesion types, it's not clear that non-dermatologists actually make a distinction between inflammatory and non-inflammatory.

So I think that's going to be one of the questions that we need to work with, and that is, should acne lesion types, inflammatory or non-inflammatory, be medically acceptable indications? I think there are two products out there right now that actually have this. Maybe there is a third. But is it something we want to continue that practice?

What we can do is we can always craft into the package insert outcome measures for both lesion types so that a more elite kind of dermatologic practice that wants to use a particular, say, topical for a particular lesion type can still find that information in the package insert. But again, the question is going to be, do you want something less than acne? Do you want lesion types as an indication?

Number five, should lesion counts be assessed at multiple time points late in the study and averaged to increase power? What we know and what Dr. Kligman has actually written about is that acne lesions, inflammatory and non-inflammatory, surprisingly fluctuate in size and appearance and even number in very short periods of time.

So one of the ways to reduce intra-subject variability and hence increase the power is to go out to that time in an acne study when you're on that horizontal asymptote of efficacy, which may be 8 to 12 weeks, and instead of just capturing one lesion count or one global assessment, do these assessments at, say, week 8, week 9, week 10, and week 12, and then take the average, and by doing that, you can substantially reduce the intra-subject variability. You can increase the power.

The other side of that, though, is that you can drive some very impressive p values within some very small lesion count deltas. But that will be one of the questions for the committee.

Then how should the efficacy outcomes of clinical trials be portrayed in the package insert to be maximally effective in communicating, especially so that physicians can communicate with patients? And we'll be presenting some information on that later today and, again, hope to hear from the committee on that point as well.

Then as Dr. Stern mentioned, the ultimate goal is a guidance document on the evidence for effectiveness for products for mild to moderate acne vulgaris. What we hope to gain over the next two days is the pieces of information that we can put together to craft a draft guidance document, which then would be published. We would

get some comments back, and that would get us going in the process.

Thank you.

DR. STERN: Thank you very much.

We'll be having questions after our next speaker, Dr. Carr, who will talk with us about the FDA perspective on global evaluation. Thank you, Dr. Carr.

DR. CARR: Again, I'll be speaking on the FDA perspective on the global evaluation in facial acne.

I'm going to begin by describing some challenges associated with the design of a global evaluation scale, move on to discuss benefits of a standard scale, then discuss proposed attributes of a scale, and close by giving examples of scales that have been proposed for use to the agency.

A number of different scales have been published in the literature and a number of different scales have been proposed by sponsors for use at the agency. It begs the question, what is it about acne that makes it so difficult to design a scale that's universally accepted?

The American Academy of Dermatology convened a consensus conference in 1990 which considered acne classification, and one of the conclusions was that the difficulties in large part related to the pleomorphic

nature of acne pertaining to the mixture of lesion types, inflammatory and non-inflammatory, the variability in the clinical presentation of those lesions, how they can vary, particularly inflammatory lesions, in size, the papules, the pustules, the cysts, and how they can vary with regard to the extent of inflammation associated with the lesions. Also, there's variability in how the lesions evolve over time.

Additionally, there's no consensus as to what should be assessed in the global evaluation of acne. Some consider that only inflammatory lesions should be considered. Some consider that nonfacial sites should also be factored into the global evaluation.

The potential benefits of a standard scale would include that for clinicians it could serve as an objective basis for treatment choices, as well as assessment of treatment responses. In the investigational setting, a standard scale could potentially increase consistency across centers as to enrollment of subjects who more closely fit the enrollment criteria as well as increasing consistency of assessments of study treatment response. And for clinicians and investigators, a standard scale could serve as a common system to aid in the interpretation of clinical trial results.

Now, the proposed attributes of a scale would

include that it have a limited number of levels -- we'd suggest no more than five or six -- that each of the levels be described sufficiently so that intra-observer and inter-observer variability is minimized; that the scale include levels which indicate the clear state and the almost clear state because these are the most clinically meaningful treatment outcomes; that it be of a static design so that the assessment reflects the clinical picture at a particular time point; and that the scale have a high degree of correlation with lesion counts.

I'm going to give some examples of a few scales that have been referenced in applications that have come to the agency and make a couple of comments about each of the scales.

The first one is the Leeds scale, sometimes referred to as the Cunliffe scale. And it's presented as a 10-grade scale where grade 0 represents no acne and grade 10 the most severe acne. But it actually is a 26-point scale because with this scale, grades 0 to 2 are subdivided so that there are nine possible grade assignments between grades 0 to 2. Similarly grades 2 to 10 are subdivided by increments, making for a total possibility of 17 grades. So this makes for a possibility of 26 grades on this scale, and a case could be made that that's a bit cumbersome.

Additionally, the only two levels that have

word descriptors on this scale are the grades 0 and 10. So this scale would be considered to be perhaps lacking in definitions.

The Cook scale presents five definitions. However, it's a 9-point scale because with use of this scale, investigators can assign grades to points that aren't identified on the scale. So investigators can assign grades of 1, 3, 5, and 7, and that makes for a problem, or potentially so, because those levels aren't defined which means assignment to those levels is completely arbitrary.

Additionally, if we look at some of the definitions, we see that there's no level that represents the clear state. Grade 0 permits for some lesions, albeit few lesions, but lesions nonetheless. And then if we step down to grade 4, we see that it begins by being described as being between grades 2 and 6. So it's considered that perhaps reworking of some of the definitions might make this scale more useful.

Now, this is another proposed scale and this is an example of a dynamic scale. The problem with dynamic scales is that their memory-dependent requiring that investigators have some recollection of the baseline status of a subject in order to make the assessment. Additionally, there are no clinical descriptors given to

any of these levels, so it's not clear really what's being scored. If you are told that a subject scored a slight improvement or a moderate improvement, that doesn't bring any particular clinical picture to mind.

A variation on the dynamic scale would be where the improvement is reported by percent change, and the same argument could be made that if you say a subject is 25 percent improved or 50 percent improved, that doesn't bring a particular clinical picture to mind.

Now, this is an example of a scale that begins to meet the criteria presented so far. It has a limited number of levels, namely five. But when we look at the definitions, we see that grade 0 which is said to be none is not none because it's defined as having occasional comedones. And then if we examined a definition for minimal acne, the question is, is this definition really minimal acne or might it be too severe to be considered minimal?

This scale, similar to the one before, has a limited number of levels, again five. It does have a level which identifies the clear state. However, if we look at almost clear, the same question could be raised. Is this definition really one that would be considered almost clear or is this too severe to represent the almost clear state?

And the last example is a scale that's

considered to meet the proposed criteria. It has six levels, so the number of levels is limited. It does have a level which defines the clear state. The almost clear state is defined by rare inflammatory lesions and papules are permitted, but if present, they can't show any signs of active inflammation. I'm not going to go through all the levels, but they are considered to be sufficiently defined so as to minimize observer variability. The scale is of a static design, and it does have a correlation with lesion counts.

So with that, I'll close my scaly presentation, and we look forward to the comments from the committee.

DR. STERN: Thank you very much.

I guess I'd like to take the chair's prerogative and ask one question of any of the three presenters who would like to answer. We've been hearing about comedonal/noncomedonal, about various scales in terms of what are usually descriptors of number of lesions and type of lesions. What I haven't heard about in terms of approvability of products is -- and we've been seeing only faces. The question gets to be, is the criteria for approving a product that is only assessed on the face necessarily applicable for other anatomic areas? At least in my clinical experience, what works on the face may not necessarily either be tolerated or acceptable for use or

effective on the trunk, another site of mild to moderate acne.

So none of these scales have broken it down into -- or do we want to break down products into those that, yes, they work on the face but we don't know whether they work on the trunk or other acne-prone areas, or yes, they work on both? Or if they work on one, we'll assume they're safe and effective on another. That's one other dimension of the scale business, be it counts, but particularly for the kind of scales Dr. Carr just alluded to.

So I'd be interested in knowing both the agency's position on that and Wilma's feeling about it as well.

DR. WILKIN: Well, we haven't required that, for example, a topical product be active on acne lesions of the back and chest in order to get approval. All a sponsor really needs to do is demonstrate success on those criteria on the face alone. However, we do encourage in the trials that the medication, which may be the active or the vehicle, be applied to lesions elsewhere on the body so that especially if we can find that it's clearly not working in some other area, we could put that advice into labeling. But we've pretty much limited it to the face. That's what the sponsors are requesting when they come in.

Their labeling is directed in that way, and we've only asked for the face.

DR. STERN: So the labeling actually says approved for facial acne mild to moderate, or does it just say --

DR. WILKIN: It wouldn't say that necessarily in the indications section, but that may be a suggestion of the committee that we want to craft that into the indications section of labeling. I think the place where one would find it would be in the clinical studies section.

DR. STERN: Questions.

DR. BERGFELD: I'm not sure I have too much to add to you, Rob, but I will agree with you that the truncal lesions, the extremity lesions sometimes are a little bit resistant, and they do require oral medications, rather than topical even though topicals are used.

I would also mention that to use topicals on the trunk and the extremities for broad generalization of acne is a very expensive deal. These are very costly products and to spread them over the body in that nature is hard to do cost-wise.

DR. ABEL: I would also like to bring up the issue of resolving acne lesions. There is an element of they may not be completely clear, but they may be significantly improved. The lesions may be smaller. They

may be resolving toward a post-inflammatory, hyperpigmented state and still might be counted as lesions, but yet they are almost clear. How does one take that into account?

DR. STERN: Dr. Wilkin, Dr. Carr?

DR. CARR: That is one of the factors that is raised as a question as to what should be counted on the global severity scale. Some people have raised the question to what extent should resolving lesions be counted in the scale.

DR. BERGFELD: I'm sorry. I'd like, Elizabeth, to have you define resolving. Hyperpigmentation for me is a resolved lesion with residual hyperpigmentation which I would not count as an active lesion.

DR. ABEL: Well, I see varying degrees of inflammation. In new severely inflamed papules, papulopustules, as they resolve, they may still be elevated. It's not just the hyperpigmentation, but they may be less inflammatory, be significantly less inflamed, but they are still papular. I have patients who come to me and say, well, their acne is not that much better, but yet when you look at it, there are many lesions in the resolving stage, maybe not completely resolved. They'll have some mild erythema, and yet they won't be inflammatory papular. They are resolving but are not completely clear, but yet they're definitely, to my assessment, improved.

DR. STERN: Along that line, we're going to be hearing after the break from a number of true acneologists, if there are such things.

I think one question that speaks to that is, do we believe that acne therapy in fact treats prevalent lesions when you start the therapy or does it reduce the incidence of new acne lesions. I think, at least in my probably, as usual, wrong concept, when we treat acne, with the exception of using things like oral steroids or anti-inflammatories, for the kind of agents we're largely talking about, we're trying to reduce the incidence so that in time, as prevalent lesions resolve, eventually the prevalence will go down as the new incidence is lower than the old.

I'd really like to hear from perhaps Dr. Pochi and Dr. Kligman and Dr. Leyden and Dr. Shalita, any of you or all of you, about is that your concept for most of the products, that we're treating incidence and not prevalence. The ideal thing would be to measure incidence.

DR. LEYDEN: I could answer it now if you like.

DR. STERN: Could you, Jim? Jim, would you introduce yourself?

DR. LEYDEN: Yes. My name is Jim Leyden. I'm Albert Kligman's personal valet.

(Laughter.)

DR. LEYDEN: I think all of us would agree the answer is both. The primary mechanism of action is working on one of the multiple areas of pathophysiology for most drugs. Most drugs only work on one area. There's one drug that works on all of them. We call it Accutane. Most drugs only work on one area and slightly on another and basically help to prevent the formation of new lesions and also to a certain degree -- and the vehicle also to a certain degree has effects on speeding the resolution of more superficial, less inflamed lesions. So it's primarily the prevention of new lesions.

DR. STERN: Well, I'm glad I got that one right for once.

Dr. King.

DR. KING: Under the concept of beauty is in the eyes of the beholder, is the FDA going to look at the global assessment by the patient? We're talking about the operation was a success and the patient died. You can reduce comedones by a lot sometimes and we all have experience of the patient not necessarily thinking it was a great therapy. So is that somehow going to be in this discussion or not?

DR. CARR: At present the subjective evaluation is not part of what we're considering. Part of the problem with quality of life or patient perception of improvement

is two subjects can have the same extent of clinical improvement, but there can be other factors that might make for different conclusions. And their assessment to treatment response such as an adverse event that one subject might rate in one way and another subject might rate in a different way so that you can have the same clinical outcome, but because of other events might have two totally different assessments as to their overall impression of treatment. So right now we're just looking at the objective assessment.

DR. STERN: Dr. Plott.

DR. PLOTT: I have two questions. First for Dr. Bergfeld. I'd like to ask when you see a mild to moderate acne patient in your clinic, what is your expectation for treatment over the first 12 weeks of your therapy with the whole armamentarium that you have to throw at them?

DR. BERGFELD: My expectation for the therapeutic response in the 6- to 8-week period would be a moderate improvement. Over a 3-month period, though, I would expect to be at 60 to 80 percent improvement. So moderate might be defined as 30 to 50 percent with a mixture of combined therapies. It might be combined topicals as well as combined orals.

DR. PLOTT: How many patients would you expect

to get clear or almost clear in 12 weeks?

DR. BERGFELD: Clear or almost clear in 12 weeks? 70 percent maybe of the mild to moderates.

DR. PLOTT: And my next question to Dr. Carr. In your example number 6, the score number 3 and number 4 -- it appears that they really differ by the type of lesion that predominates, the inflammatory in number 3 and inflammatory. It suggests that inflammatory lesions are a more severe type of lesion. I wonder if you would comment on if you believe that inflammatory lesions are more severe.

DR. CARR: Well, the inflammatory lesion does seem to drive the global evaluation. They do seem to predominate in the global picture. So I don't know if it would be termed a more severe lesion necessarily, but in terms of the global evaluation, they do have more impact.

DR. STERN: Dr. Kilpatrick.

DR. KILPATRICK: Thank you, sir. I have a number of questions coming after Dr. Plott.

Wilma, what I heard you describing was an ideal treatment of a patient. That may not be what actually happens with non-dermatologists. But what I was hearing seemed to imply that there were limitations on actually trying evaluating in clinical trials because how can you treat the patient at the same time if you're going to be in

a double-blind clinical trial? Basically perhaps I'm indicating my ignorance of the natural history of the disease. Does it allow for the intercession of a clinical trial to answer these questions while preserving the rights of the patient?

DR. BERGFELD: I think that most dermatologists would agree that with combined therapies, the responses are quicker and more long-lasting. In a clinical trial, it's a solo monotherapy. So those patients who were picked for that would have some limitations on their full responses. But perhaps Alan Shalita and Jim, Peter, you might want to respond. Al?

DR. SHALITA: I think a very important question has just been brought up and I was actually going to bring it up later in my talk. We do have an IRB member on your advisory panel.

But increasingly we are seeing IRBs, particularly community representatives, who are opposed to the concept of vehicle control or non-treatment control, et cetera. I know that this creates enormous problems for those that rely on evidence-based medicine and the concept of using a vehicle or placebo, but it is contrary to the best interest of the patient to be treating them with something other than an active, even the concept of treating them with monotherapy when you have strong

inclinations that more than one therapy would be best.

And then finally, Todd just brought up a concept. We don't use monotherapy generally to achieve a clear or almost clear status, and to use that then as a criteria becomes self-defeating if you're talking about monotherapy.

DR. KILPATRICK: Dr. Wilkin wants to get in.

DR. WILKIN: If I could speak to the issue of vehicle control. I think in virtually every study that we've gone back and looked at the data, people who were assigned vehicle or an oral placebo get better in acne trials.

I would say that the second piece is we're talking about mild to moderate. We're not talking about something that is going to damage someone for years if it turns out they're assigned to one of these so-called inactives.

And the third thing is you'll have to look at some of the data and see what the actual differences are between the contribution of the active over the vehicle. I think you may from that decide that it really is informative to have a vehicle control.

And then if I could come back to an earlier question, and that is do we ask for the patient's perception of how well things happened during the acne

trial. And I think Dr. Carr answered that we don't request that information. Often we get it as a secondary kind of an endpoint, and we'll look it over.

But for the exact reasons that she mentioned, I would like to lift up for the committee's consideration a very thoughtful editorial that appeared in Lancet by Mark Lebwohl. It's not on acne. It's actually on psoriasis. He was referring to a paper in the British Journal of Dermatology by Fountain on psoriasis. What they found out was that looking at objective measures of the severity of the psoriasis didn't really correlate very well with the patient's perception of quality of life change during therapy. In Dr. Lebwohl's thoughtful account, he indicates what Dr. Carr was saying and that is that patients bring an awful lot to that equation, what they want out of something, what their expectations are, what others' expectations are, around them.

Our thought is that that is important to that person in that trial. I don't want FDA to ever sound like we're not interested in quality of life. We're enormously interested in quality of life. But our thought is if we can somehow craft into the package insert some fairly objective measures of outcome, then we actually convert the quality of life discussion to the clinician's office where he or she is sitting with the patient and can say, well,

you could expect this sort of thing, and then it's that patient in real time that can come up with the quality of life assessment. But clearly, we're all interested in quality of life. That's actually a big part of the mild to moderate acne indication.

DR. KILPATRICK: Sir, may I continue because my light is on?

(Laughter.)

DR. KILPATRICK: I find myself in the position of disagreeing with my friend and colleague, Dr. Wilkin. As a non-M.D. but as a statistician, I'm interested in the accession of information, and the subjects I think can bring information to a clinical trial in terms of their subjective, albeit it subjective, evaluation of their improvement or lack of improvement over time.

The fact that this may not be highly correlated with scores leads me to a second question directed at Dr. Carr. I'm not surprised that in the global evaluation one of the conditions for a scale is that it is highly correlated with the score. I would have thought that they would want it not correlated with the score in order to get some different perspective. If it's highly correlated, if you go to the extreme, if it's a correlation of one, then the two are redundant. So I'm looking to broaden the evaluation of acne therapy not limit it. If we have two

things that are measuring the same thing, let's take the simpler one.

Finally, since I'm on the microphone, let me ask again a simplistic question to Dr. Wilkin. This must be done. Why cannot we take photographs and literally count the number of comedones rather than evaluate them in a patient-doctor contact? Jon?

DR. WILKIN: I would actually like to defer the photography question to the acne numerology experts who do the counting. There is a published system of getting really very well-controlled photographs and then doing counts.

DR. STERN: Would you introduce yourself first, Dr. Kligman, just for the record?

DR. KLIGMAN: Al Kligman from Philadelphia.

Jonathan, in the first group when we met to lay out rules for assessing the efficacy, at that time we denounced and made light of photography. It wasn't meticulous enough. It missed little lesions, especially comedones and closed comedones.

All that has changed. The improvement in photographic procedures now is unbelievable with digital photography, with video microscopy, with the ability to look at UVA photography, fluorescent photography, PRIMOS imaging. An enormous amount of bioengineering skill and

resources are now available.

Of course, they're expensive and the lighting has to be defined. The film has to be defined. It's a very rigorous procedure, but in my opinion it's going to offer much more believable, credible, and objective results of what we are actually seeing considering the fact that we have a mixture of lesions and they all have their own history and their own outcome.

So I think that's a very good idea. Those resources are now available and they could be put into place by anyone with money.

(Laughter.)

DR. STERN: Ms. Knudson.

MS. KNUDSON: It's Paula Knudson.

I would like to speak to the IRB issue. I do know that over the years placebo-controlled trials have become an anathema to many IRBs.

However, I would say that one of the things that we would be asking is for mild acne would the acne resolve by itself most usually, in which case I think a trial with placebo would certainly be countenanced. For moderate acne, we would ask what is the likelihood of scarring, and the other thing that we would ask would be what's the length of time for it to resolve. So those would go into the makeup as to whether a vehicle-controlled

trial would be approvable or not for mild to moderate acne.

But I wanted to ask a different question of Brenda Carr and that is, is it anticipated that at every visit that a patient comes to the dermatologist, the scale would be used?

DR. CARR: You're speaking of in the clinical trial?

MS. KNUDSON: Yes.

DR. CARR: Yes.

DR. STERN: Are there other questions? Yes.

DR. TEN HAVE: I'd just like to make one comment about the monotherapy versus combined therapy issue. In other areas such as psychiatry where therapy is usually done in a sequential, complicated way, people are thinking about enhancements to the simple clinical trials design in terms of using adaptive randomization as opposed to a single baseline randomization to possibly attempt to make a more realistic comparison and evaluation.

DR. STERN: I'd like to make a statement and ask a couple of questions, one at least of Jonathan. In the issue of combination therapy, one of the things that to my knowledge has not been looked at is by combination therapy I think we all agree that using multiple agents seems to be more effective than using one agent alone for mild to moderate acne, whether it be a combination of a

topical and an oral agent or combinations of appropriately used topical agents.

Sometimes when people think about combination therapy -- and if you look at a number of the recent approvals, they are in fact taking two agents that are available individually, putting them together and marketing them and approving them as being better than the individual agents. The question gets to be then one of frequency. We learned from topical steroids and from topical antifungals where the paradigm was you always had to do everything at least twice a day, and in fact for many agents once a day is sufficient. So some of the question gets to be can you just use the individual agents as well or better in terms of tolerance than the combined agent as opposed to combination therapy.

So I think there are some added complexities of combined agents, that is, an agent that take two active agents known to be independently therapeutically active and puts them together in terms of what should be the criteria of approving a combined agent as opposed to having those two individual agents available separately. What are the real advantages of that agent? Do they really work better than the individual application? Is there anything that makes them better?

And then for Jonathan I wanted to ask just a

question. One of the interesting things to me about your results were that the anchor point was 101 lesions for the worst ever or 200 lesions. If you looked at the two curves that essentially said once we overestimate the number of lesions through most of the interval, they were almost superimposed on each other. That to me, being the victim of one of those curves in terms of overestimating the number of lesions, was interesting. You're saying at least within this spectrum, a lot is a lot and how we view that a lot in terms of estimating, once we're given the anchors, is subject to the same kind of biases.

Now, if you're looking at lesion reduction, the worse the patients you have, it may impact on how many lesions you have to reduce when on your last curve, I believe it was, you showed how much down the scale you have to go to get one level of improvement by your non-quantitative scale.

So could you talk a little bit more about that? Because I found that interesting in terms of what it might mean for evaluating agents with these non-quantitative scales.

DR. WILKIN: Yes. I think maybe what you're leaning towards is what actually happens in an acne trial. You can imagine that those who come into the trial -- there will be inclusion criteria and there will be a range

of the non-inflammatory lesion numbers that one can have to be in the trial and also the inflammatory lesion numbers. People who are at the upper end often are the folks that drive success on the lesion count analysis. Those who come in, they just barely had enough acne to get into the trial, they are the folks that drive the global. Is that the point you were --

DR. STERN: That's the data I took away from it, and it seemed to me that a system like that was less than desirable on the one hand. To Dr. Kilpatrick's point, it did allow two independent measures, one of which was in a sense active and robust at the low end of severity and the other perhaps more active and robust at the higher ends of severity within the spectrum. But somehow that lack of correlation in what sort of we think should be correlated across the spectrum of people coming in the study is a bit bothersome.

Dr. Kilpatrick?

DR. KILPATRICK: Well, yes, again I heard earlier from was it Dr. Fraser who talked about specificity of objective in going into a trial, and I'm all for that. What I'm hearing now is stratification. But that has to be very carefully crafted between the FDA and the sponsor beforehand.

DR. STERN: I'm sorry. Dr. Tan.

DR. TAN: Yes. I'm still trying to get to what is the real problem here. Can Dr. Wilkin and Dr. Carr clarify for me how exactly you define the percent of reduction? Dr. Fraser presented that the percent of reduction is patients from the baseline to 12 weeks, for example.

I think one of the problems is the number of lesions because all the lesions are different. And when you just lump them together that causes all this problem. I think you have these stratifications, non-inflammatory, inflammatory. In molecular biology these days they're counting different cells, but this is all related. There are different clusters that are related. They should be weighted a little bit differently when you consider them together to derive a global scale. So there should be a weighted type of scale that you should use for the final endpoint.

And another problem I have is -- that's why I asked the percent of reduction.

The last thing is percents, that is between 0 and 1. Right? So when you analyze this kind of data, I was remembering in the past several Derm meetings, from what I remember, it's just a comparison, ANOVA type of comparison using normal distribution comparing the percent of reduction for the control versus the active treatment.

And there is a profound problem if it's a percent, as we say, it's a ratio, and that percent, if it is a ratio -- if the numerator and denominator are normally distributed, mathematically you can prove that the ratio is not normally distributed. So actually a lot of these things are -- you're assuming it's normally distributed and there is a problem with that. So I don't know how that ratio is really analyzed. Probably we'll hear more in a later presentation.

DR. STERN: Dr. Wilkin.

DR. WILKIN: I think those are important questions. Actually Dr. Alosh this afternoon has some material that he can present some numerical analyses that I think will help. They'll be very responsive to that. We were thinking that the first part would be sort of to go over clinically what the different lesions look like and whether or not we want different lesions, and then the analytical part and whether there's normal distribution -- you'll get to see data from NDAs that have been suitably anonymized this afternoon.

And I would like to just add a third disclaimer. Once again, I gave a disclaimer at the beginning and at the end of mine. I want to emphasize again that was a model. That was not real acne. It was intentionally simplified. The curvilinear relationship,

while it looks kind of neat when you're looking at little dots painted on a face in kodachromes, real acne is not that simple. I think the acne experts will indicate that you really can't predict where someone is going to fall out in the global scale based on the lesion counts.

DR. STERN: I think with that last comment, perhaps we'll end questions here since we'll be going on to this in greater detail as the day goes on. Thank you very much. We'll resume at 10:45.

(Recess.)

DR. STERN: I think we're particularly fortunate this morning to have our four next speakers with us. In my mind they represent certainly the majority of individuals who have made a substantial contribution. Notice, Dr. Kilpatrick, I did not say significant contribution.

(Laughter.)

DR. STERN: A substantial contribution to our understanding of acne, and in fact, I know significant is okay in that non-statistical usage as well.

DR. KILPATRICK: I'd like to make a comment about the difference between clinical significance and statistical significance.

(Laughter.)

DR. STERN: But they're all clear thinkers and

inspiring teachers, and I'm very much looking forward to hearing from them. Our first speaker will be Peter Pochi who knows not only how to do the research, how to teach, how to practice, but also where to live, and Peter will be talking to us about the American Academy of Dermatology. He is Professor Emeritus at the Boston University School of Medicine and lives in Boston, the right place to live.

DR. POCHI: Thank you, Dr. Stern. When Dr. Wilkin invited me to speak today, I accepted with some trepidation since I hadn't given a lecture in 11 years, and I hope I have not forgotten how to talk.

In 1990 the American Academy of Dermatology sponsored the convening of a consensus conference to look at the problem of the classification of acne. I'll just read for you, for those who don't have the article before you, the first sentence or so. "A number of systems have been described for the classification of acne vulgaris, but there's no universally accepted method for assessing gradations of acne severity. This lack of uniformity from one classification system to another has made it difficult to compare therapeutic efficacy among different studies."

It's 12 years later and the issue is still being addressed.

The academy prefaced the report. The proceedings of the conference were published subsequently

in 1991 in the Journal of the American Academy of Dermatology, and the report was prefaced by the academy saying that the results of future studies may require alteration of the recommendations as set forth in this report.

The proceedings that were reported were not really proceedings. They did not go into any detail of the various presentations that were made on the first day of that day-and-a-half conference. A number of speakers, including Professor Cunliffe and Professor Plewig from abroad talked about their classification systems, and as the day droned on, it became evident to most of us at least who were interested in the subject -- and among the participants were, beside myself, Dr. Kligman, Dr. Shalita, and Dr. Leyden who are here today -- that trying to define acne is not a walk in the park and that it might be better to present it in almost a global sense, which I'll come to ultimately. But first I want to go over what the conference intended to provide.

The purpose of the conference was twofold. The first was, as I've already indicated, to review and to assess the suitability of the grading systems that were in place at that time, and there were a number of them. I'm not going to go into detail at all, not discuss them at all really except to allude to one or two as I go along. It

became evident, as I've already said, that it was very difficult to arrive at sort of a universality of a type of system that could be used in all situations.

The second purpose of the conference, which was really an outgrowth of the first, was to categorize what is meant by severe acne. It's very difficult to know when a moderate case of acne ends and a severe case of acne begins. Patients are treated with oral medications such as the oral tetracyclines, which are FDA approved as adjunctive therapy in individuals with severe acne, and oral isotretinoin, or Accutane, for not adjunctive therapy but prime therapy. It was hard to know just exactly what constitutes a patient with severe disease. So these were the two goals of the conference.

Now, in assessing acne activity I think there are two aspects to consider. One is the practitioner's assessment and the other, which you are more concerned with today, the investigative therapeutic trials. These are really two quite different areas of consideration. The practitioner assessment I think gets divided into two types of assessment.

One is the individual physician, dermatologist, pediatrician, or family practitioner, who sees the patient on every visit from the beginning of treatment until the treatment is concluded. Here the examiner has latitude in

assessing what the activity of the patient's acne is, creates his own grading system, as I did in my own patients -- I would grade the patients as mild, moderate, and severe, for example -- and then would have clinical descriptors for each of them, inflammatory predominates, non-inflammatory predominates, they're both present, is there scarring, et cetera. And when the patient is seen again by the same examiner, it is really easy to do an assessment in my experience and the experience of those to whom I have spoken to get a reasonable evidence-based, if you will, outcome of the disease of that particular patient.

The problem is that different examiners may see the same patient. This is particularly true in clinics and especially true in university clinics where there are resident physicians who rotate around, say, every month, and it's almost uncommon for a patient to be seen by the same physician on subsequent visits. And this really would relate to the problem that we have in investigative therapeutic trials wherein a system has to be established that's fairly objective with subjectivity intercalated among the objective observations.

Now, the oldest system I could found was this neolithic textbook of dermatology published in 1956. I'm being actually unkind. It was really the breakthrough

textbook of dermatology in this field by Pillsbury, Shelley, and Kligman, and they were the first to really attempt to give some sort of a subjective/objective, if you will, evaluation of acne. And they graded acne into four grades, and they gave descriptors: simple, banal; no significant inflammation. That really is simple. And then grade II, moderate severity, occasional inflammatory lesions. These are not my words. I've taken these directly from the text of that book. And grade III, more severe; grade IV, most severe.

Well, really this is okay, but really inadequate. One really has to fit in more describing attributes to the patient's acne. Nonetheless, this is what really is done in a global assessment of acne, is to try to divide the disease into several grades and then to give little descriptors of what one sees, and that should be adequate but is it?

Now, it's already been mentioned that acne is difficult to classify because it is pleomorphic. It's highly pleomorphic. Let me just go through each of these steps one by one.

First of all, as you'll recognize, there may be both inflammatory and non-inflammatory lesions. In a global or even in a counting technique, trying to integrate these together I think leads to specious information. And

I agree with Dr. Kligman. Perhaps he doesn't agree with himself any longer, but I agree what he has written that the inflammatory lesions and the non-inflammatory lesions really have to be considered separately and they need separate grading because you can have situations where the non-inflammatory lesions so predominate and yet the patient doesn't really look that bad with only mild inflammatory disease.

I noticed, if I recall, in one of the grading systems that Dr. Carr spoke about, she showed with increasing severity of the disease, an increasing number of comedonal lesions. In my experience usually the opposite occurs, that as the disease becomes aggressive, there are fewer non-inflammatory lesions. But, of course, there are many, many exceptions to that.

Secondly -- and this is the most important, the second point -- the inflammatory lesion which is really the hallmark of the disease, what brings 90 percent of the patients to doctors for their disease -- is variation in size, density, and severity.

Acne lesions vary greatly in size not just from patient to patient but within a given patient, and I'll show you some clinical photographs in a moment. If you look at patient, no one lesion looks -- well, they do look alike but they're quite different in their size. They can

be large, they can be small. And where to draw a line as to what is small and what is large is arbitrary but is subject to, I wouldn't say, misinterpretation but difficulty in classifying.

And they vary in density. There are two meanings of density. One is the number of the inflammatory lesions that are seen in a square area of involvement, and the other is the distribution, clustering versus a more even distribution. This latter aspect has never, to my knowledge, been considered in any classification of acne. Does an individual who has a lot of their acne concentrated in given areas in the face versus the patient with the same number of lesions but more evenly distributed look better or look worse? And this is another aspect that I think should be looked into.

And then the severity, the severity of the inflammation, not the severity of the disease. Some lesions are quite red. Some lesions are not as red. Some are only pink, and this is roughly the same for a given individual but can vary so much in the same region of the face. You have a variation of erythema even if the lesions are roughly of the same size. Of course, they're not. So the degree of inflammation is important, particularly in doing a global evaluation.

The patient's background pigmentation is often

not considered in global assessments. If an individual has light skin and has inflamed lesions, red on white looks much worse than red on dark. If a person is sunburned, the inflammatory lesions will look so much less intense, and this is why individuals probably improve when they go out in the sun. It's not that the acne improves from the sun, but it's globally they look better because it's red on red instead of red on white.

In some individuals who are darkly pigmented, the inflammatory aspect is quite difficult to see. In fact, people who are not familiar with seeing black patients at first they say it's very hard for them to perceive that a lesion is even inflammatory. So this is an important aspect again that I think has been largely neglected.

Individuals with black skin also, on the other hand, as Dr. Abel has pointed out, have the problem of pigmentation and this becomes a clinical problem. Does one assess persistent pigmentation as part of the global assessment?

Then there's finally the variability in the evolution and healing of lesions with or without treatment. Some patients heal quickly even without treatment. Their lesions just subside more quickly than others do. In some it is much more persistent, probably having to do with P.

acnes. Dr. Leyden I'm sure can address this far better than I can. And under treatment some patients just simply get better, and lesions can evolve more slowly. Unless you have significant numbers of patients who are being treated, this variability would be an important aspect.

Now, let's look at some acne. I don't know if you can see this in the not totally darkened room. This is a patient with mild disease, not maybe to the patient's eye, but to the physician's eye, just a few scattered erythematous papules.

This is a patient with terrible disease, large numbers of inflammatory lesions, pustules, nodules, sometimes referred to as cysts over the course of the face. These patients present no problem in global evaluation and certainly at baseline. The problem that comes up is the patients who are in between. If you call this grade V and you call the slide before grade I, how many grades in between are necessary to get an "accurate" assessment and what should be included in them? Well, this is what this conference is about, and I would hope that something will come of it in this regard.

Now, going back to the milder side, this is a patient, a little more severe than the one I first showed you, but still no scarring, and the lesions are all small. This would probably be called moderate. Some may call it

mild, but certainly not minimal and certainly not severe.

This is a patient with somewhat more severe disease. A few more lesions, but some of them are larger, not terrifically large, but they're certainly approaching nodular size which by definition arbitrarily is a lesion that is 5 millimeters or larger. These lesions may be 4, they may be 5. There are other lesions that are much smaller. There are a few areas which may show this post-lesional inflammation that Dr. Abel referred to as these flat, macular erythematous areas. When an acne lesion heals, it sometimes leaves no erythema; it sometimes leaves erythema that can persist for many weeks and months. Do we count these? Do we not? Would high resolution photography that Dr. Kligman suggested earlier today be able to discriminate papular lesions from these healed inflammatory lesions? Should they be counted? They're difficult to see by photography but perhaps with virtual reality photography they will be able to be seen.

This individual actually has more severe acne, and if you count the number of lesions that this patient had with the number of lesions the patient on the previous slide had, they're about the same. But this patient is worse. Why? Because several of the lesions are quite large. They're nodular, and so this patient has a more intense appearance. So counting lesions by themselves I

shouldn't say is hazardous, but it has to be taken with not a grain of salt but has to be appreciated.

This individual has obviously bad acne, not the type of patient that would be considered in topical therapeutic trials. I want to point out something and that is her lesions are quite clustered. She doesn't have any nodular lesions. She has a large number of small papular and pustular lesions. She also has scarring. A word about that in a moment. But one of the things that one sees in acne -- not commonly but it does occur -- is perilesional erythema, erythema surrounding the lesion and this can make a patient look much worse. If you have a patient that has, say, 10 inflammatory papules and another patient has 10 inflammatory papules but with surrounding erythema, then that patient looks worse. And here this patient has a lot of this and happens to have lesions concentrated in an area, so this looks like almost something other than acne. It's very highly inflammatory, but yet does not have a large number of lesions.

I mentioned scarring in a moment. This person has had disease for a long time. This should never happen to a patient nowadays. But in scarring, in global evaluation of a patient and when you're considering the type of therapy in a private setting or in a clinic setting, the presence or absence of scarring is very

important. While most scarring of this type that you see here will occur in individuals with severe acne, you can occasionally get scarring in patients with mild acne. In fact, the reverse of the case, you can get no scarring in patients with severe disease. So there's not a one-to-one correlation in individuals with mild disease and the prospective scarring.

I only mention this because if an individual is being considered for a study who has very minimal scarring, such scarring should be a contraindication. The individual should not have any scarring. It's not going to affect the outcome of the inflammatory component of the disease. Therefore, it should be excluded.

I'm afraid this doesn't show up too well, but it illustrates a problem. We have here the forehead of a young man with highly inflammatory lesions. They're actually not quite nodular in size. They're about 4 millimeters with pustular centers. So this would be a pustular lesion with surrounding erythema. And then there are some smaller lesions, and then there are some of these seemingly flat, erythematous lesions. If you were to count these lesions, you would have to count smaller lesions in the same count as lesions that are much more intense looking, and yet they would be classified as a papule or a pustule less than 5 millimeters. This is very difficult.

This narrow area of papular and pustular lesions. Should attempts be made to grade those?

I'm getting into lesion counts, which I don't want to get into, but Burke and Cunliffe back in the original report divided papules and pustules that were smaller than 5 millimeters, which is the definition of a papule and pustule, into two categories: active, larger, more inflammatory; less active, smaller, less inflammatory. Highly descriptive. And they mention that "some 40 percent of the lesions fell between these two types but in practice we assigned the lesion according to its major component." This statement is a direct quote. It's inscrutable to me, and I don't understand how they could arrive at this attempt at least to classify lesions smaller than 5 millimeters by more active, less active. I would have great difficulty doing this. It shows the problems and the tenacity with which this issue is approached.

Now, the last slide, which is literally the bottom line. From the result of the conference that I was supposed to discuss and have been, it was concluded by the members that it was very difficult to approve, if you will, or to recommend a grading system for acne dependent upon lesion counting and other aspects, and it was better felt that a grading system, at least on baseline in patients with acne, would be best achieved by what was called

pattern diagnosis. I think this term was suggested at the time of the meeting by Dr. Kligman.

Patients with acne would have either mild, moderate, or severe disease -- they were talking only about inflammatory acne, leaving non-inflammatory acne aside -- and describing the degree of papules and pustules and nodules. A patient with mild acne would have few to several papules and pustules, again no numerical definition, descriptive definition, and no inflammatory nodules, no cysts or nodules. Patients with moderate acne would have several to many papules and pustules, again no numbers, and few to several nodules. And patients with severe disease would have numerous and/or extensive papules and pustules and many nodules.

Let me preface my dubious comment about this slide and the conclusion of the conference. This is not applicable for treating mild to moderate acne in terms of successive assessments of patients because you would have to go from here to here or here to 0, which is not part of the grading. So this is not what is germane to the discussions at hand. However, I think that this is wrong. I think that there was a mistake in calling moderate acne as having few to several. This should have been only few, and several to many should be under the category of nodules.

So the conclusion of the consensus conference in 1990 was that one could not clearly identify a single classification system for grading acne or even for the global assessment of acne on a longitudinal basis, but this at least provides some guideline for the use of therapies in acne in patients seen in the office and in the clinics.

Thank you.

DR. STERN: Thank you very much, Peter.

Our next speaker is Jim Leyden who is a professor of dermatology at the University of Pennsylvania and another person with a long and illustrious track record in the evaluation and treatment of acne.

DR. LEYDEN: It's great to be here just to hear Peter come out of hibernation and give one of his usual very thoughtful presentations.

While we're doing that, I'll tell you a story about my oldest grandson who is just 5. About a couple of months ago he said, Pop-Pop, could you get me some cream? And I said, yes, sure, what for? He said, I got a couple of little red dots here that won't go away. They were two little inflamed milia. And I said, I'll get you some cream, but let me tell you why you get them. He likes to play chess with the computer a lot. I said, when you're playing chess and you're thinking, you're doing this all the time. If you stop doing that, you won't get them and

you won't need the cream. He said, okay.

And a couple of hours later, his mother called me and said, Jamie just came to me and said, I don't think Pop-Pop is a very good skin doctor.

(Laughter.)

DR. LEYDEN: Well, he told the story and he said, I'm not doing that. Why would he say that?

And then the dagger in the heart. He said to his mother, I want to talk to another doctor.

(Laughter.)

DR. LEYDEN: So I hope you won't feel that way when I'm finished.

(Laughter.)

DR. LEYDEN: I'm going to talk about global assessment primarily. I thought I'd begin by just reviewing what you've already heard, that currently the approval process involves what I like to refer to as the meatloaf approach, you know, two out of three ain't bad. You have to have reduction in non-inflammatory lesions, inflammatory, and total lesions, two out of three, plus some kind of evaluation, overall global assessment.

And this is where all the problems are as all of you are getting the sense. This has worked more or less reasonably well probably because the majority of drugs that we've had have been either topical antibiotics or topical

combination antimicrobial/antibiotics and topical retinoids, and then more recently oral contraceptives.

Oral contraceptives have enough effect on sebum that the overall severity of the disease, both inflammatory and non-inflammatory, goes down enough that this kind of system works.

Antibiotics work mainly by suppressing the organism that creates the inflammation, but we have also known for a long time that there is a modest but consistent effect on non-inflammatory lesions. We now understand the mechanism by which that occurs.

Topical retinoids work mainly on the abnormal desquamation and have the most obvious clinical effect on non-inflammatory lesions although they all have been shown to have effect on the inflammatory phase. And now we have some understanding, at least of some of the molecular mechanisms in terms of their effect on total receptor expression.

So the drugs we've had have worked well enough with this kind of system even though we have all kinds of issues dealing with the global assessment.

However, I think in your considerations, the drugs of the future may well work only on one area of acne pathophysiology to the exclusion of others. And I think to some degree that day is already here. We have very low

dose doxycycline. While an initial study showed some effect on non-inflammatory lesions, whether that effect will be great enough to make sure that two out of three is reached and whether that's reproducible needs to be seen. There are non-antimicrobial antibiotics that have anti-inflammatory effect. We're all familiar as dermatologists with the macrolide derivatives that have anti-inflammatory activity.

In a series of regional derm meetings that I've been involved in over the last three or four months, it's quite clear that many dermatologists have decided that Eladil, for example, and also to a certain degree, Protopic have effect in the inflammatory phase of acne. Whether or not that can be substantiated enough or whether or not the manufacturers will choose to try to substantiate that in terms of an approved FDA claim remains to be seen. But I would suggest to you that if and when that's the case, it's very unlikely that a pure anti-inflammatory drug will have any effect on the non-inflammatory phase. So the day of thinking about approval of drugs for aspects of acne I think is here and should be part of your overall considerations.

A couple of general issues before we get into the global assessment I'd like to bring up -- and you heard a little bit of it already. It's very clear from

investigator meetings that -- I try not to attend them. I try to send my nurse coordinator. It's very clear that recruitment of patients has become a big deal. It used to be relatively easy when there was not the kind of access that the population in general now has to recruit patients by telling them you're going to be in a study for 3 months or 6 months, if it's an oral contraceptive, or whatever, and you have a 50/50 chance of getting something that's not likely to be very useful, and at the end of that, you're going to get paid for your time and we're going to treat you free.

Now people say, well, I don't think I want to wait for that, particularly as we'll get into when you discuss about where the line is for mild and moderate. Right now the current guidelines suggest that you must have at least 20 inflammatory lesions, which means most of the patients have more than 20 and lots of them are at a point where you would have to say would you want your child in that study if that meant 3 months of no treatment. Leaving aside that their life is not going to be ruined, it's a difficult discussion particularly when people now have access.

So I think the time may well come -- and it has come -- with the recent study a year or two ago with the new formulation of systemic isotretinoin. That was a

positive controlled study because I think there it was easier to say, well, this is very, very bad acne that isn't likely to get better spontaneously, or if it does, we'll call the cardinal and tell him a miracle has taken place. So that study was a comparative between a new formula and an old formula. And I think you really have to consider that because I think the time is coming when our IRBs will be more and more like Europe and just not permit it unless it's very mild disease.

And vehicle for topical and placebo systemic controls are less and less acceptable to potential patients. This is something I would hope you would at least consider and that's a placebo or vehicle run-in. If you look at every study that's ever been done, as Dr. Wilkin said, the vehicle patients always got better, or at least as a group they got better. The mean goes down. Most of that is in the first visit after starting the trial. You can particularly see that most clearly in those where there's a relatively early first visit at week 2 or week 3 after stopping. So consider a placebo or vehicle run-in where everybody gets in and they're in. Then at a certain point no matter what they have, they're still in even if they're below the initial minimal inclusion criterion.

Let's get to the global assessment, and I was

asked by Jonathan to stress the inflammatory aspect in terms of global assessment.

One question you can ask is, is it needed? Actually as it stands now, a group of 9 or 10 of us was brought together at the academy meeting last year by a company new to dermatology who was somewhat perplexed by the requirements. And the group of us decided, as it stands, it probably should be removed. Should not lesion counting be sufficient? You'll hear from Dr. Kligman later how difficult lesion counting can be. With the imaging techniques that we have now, I think all of us agree that that can be greatly improved.

I'll also tell you a secret if you promise not to tell him that I said it. He's never counted pimples ever in the 35 years that I've worked with him. But as is often the case, he knows things without having to go through the work that the rest of us have to.

(Laughter.)

DR. LEYDEN: And he's rarely been wrong. So one has to just remember that.

In the past the global assessment was a so-called dynamic, a pre-post therapy, and the question of, well, how can you remember? Well, obviously you can't remember, but you can have images, large transparencies. Some companies now have very sophisticated ways you can

just type in a number and up comes a large, life-size image of the person, right side and left side, from the initial visit. That kind of analysis was done with the photo damage for the tazarotene clinical trial, for example. So you don't have to remember. You can have an image to compare with.

I would agree that in the past without an image, the global assessment was probably done mostly by "how are you doing" and seeing what the lesion counts were and then making some assessment, various so-called static global assessments with varying scales, and you heard of the difficulties with some of those scales.

But I just want to make sure you all know that success means 100 percent clear or near clear with no further treatment required as being part of the near clear. We'll get into that. Is that a reasonable, clinically relevant endpoint? It's a crisp endpoint. Nobody would argue that someone that's totally cleared up has gotten better unless they had practically nothing to begin with, but if they have at least 20 inflammatory lesions and they have none at the end, and they had, say, no comedones and they have none at the end, I don't think anybody would argue. They're better. The question is where should you draw the line in the sand to constitute a degree of improvement that's meaningful and should be part,

therefore, of the overall analysis. And one of the questions in your book is how to best present in the package insert information.

And I'll show you people who would qualify as not successful, failures. Not to include the fact that they achieved that kind of improvement with monotherapy I think is not fair and does not accurately present the benefit that a given monotherapy in this disease with multiple areas of pathophysiology. As I think all of us would agree, it's an uncommon patient that gets one drug for acne, and that reflects the fact that it's multiple areas of pathophysiology and you can counteract multiple ones.

Using this kind of facial diagram that Anne Lucky first came up with in making sure that you go into each quadrant means that if you take your time and are careful, you can count these individual non-inflammatory lesions and even count the most difficult ones, the ones that are best seen by stretching the skin, the so-called closed comedones. They can be counted on the hoof, so to speak, with the patient there. They can also now be visualized and counted without the patient sitting there and hoping you'll get finished quickly so they can get out of the room.

I'd just like to emphasize a couple of things

that Dr. Pochi said and others during the discussion. This is a patient who would qualify by today's -- this patient actually has 37 inflammatory lesions, but the quality of the inflammation is very, very different than this patient who actually has almost 100 inflammatory lesions because just about every individual follicle is involved, although the quality of the inflammation is quite different. I think by trying to put words to a description of how bad a patient is is part of the problem, which I'll say a little bit more about in a few minutes.

So, as I see it anyway, some of the problems with current success, meaning 100 percent clear or practically nothing such that a patient wouldn't need any kind of treatment, assuming they stayed at that point -- is very uncommon with a single mode of action treatment. Acne, as we all know, is a chronic, relapsing condition. Three months of therapy is almost -- that's it. You can go home now. Your acne is gone is just something I'm not personally familiar with. And to think that at the end of three months it's over -- or at least that's implied in the fact that you've gotten to a point where you're clear or near clear, not requiring further therapy.

The more inflammatory lesions you have, the less likely -- and I think you've heard from Jonathan's presentation that that makes sense from his point of view.

Again, certain drugs have more effect in one area, and drugs that have primarily effect in the non-inflammatory phase of the disease without influencing the precursor of inflammatory lesions, if such drugs are in development -- I would suggest they will be developed because we now understand some of the molecular aspects of comedogenesis -- could fail by today's standards.

I'll just show you one example of combination drugs that work on multiple areas of pathophysiology and seem to be susceptible to some statistical quirks that don't make sense to me when you have a low responder rate. When you take the endpoint of 100 percent clear, you end up with very low, but highly statistically different. You know, 6 percent versus 0. Even I can do the statistics. But when you have multiple cells, then there is the potential for very good drugs not showing a statistically significant difference while the clinical effects may be obvious.

So, these are not as good as they would be if the lights were completely out, but this is a patient who's got mild disease, and you could say, well, he's almost clear if the other side were the same.

But here's a patient with much more severity. Those up front can see the non-inflammatory lesions, a lot of inflammation. He's clearly, definitely better. But by

today's standards, he has failed.

And this patient who is not clear but really better has failed, as has this patient. This is a failure because it's not 100 percent clear nor almost 100 percent clear.

So it just seems to me that doesn't make good sense clinically. One could envision a drug that did this in 75 percent of patients failing because not enough patients reached total clearing or almost total clearing.

Now, for the statistical quirk. If one knows from some preliminary work that a global assessment was 18 percent clearing versus 11 percent in the vehicle, if you wanted to have an 80 percent power, you'd need somewhere in this neighborhood of patients, and then to allow for dropouts, something like 2,000 patients for a four-arm trial. What happens if the response rate was 18 percent versus 12 percent instead of 18 versus 11? You're down to 65 percent power apparently. That I think reflects this low responder rate can have influence on studies with multiple cells.

I personally like a scale called the Allen and Smith, which was not mentioned this morning. It's a validated scale that was published in the Archives in '82 or '83. It involves evaluating both the non-inflammatory and the inflammatory aspect of the disease separately

instead of trying to jumble them together with words, as you saw on some of those. That Cook scale. I always loved that one where one of them begins with loaded with comedones, whatever that means. That was the first line in the grade. So this has been shown that investigators can reproducibly give the same kind of grade for both phases of the disease.

I personally think that the pre and post use, the so-called dynamic evaluation by investigators, with either transparencies or digital images or, as I'll go into in a second, using the same kind of images for an external panel of judges makes it a lot easier than trying to come up with words that describe what we're trying to integrate.

This is practically no acne. You can see a pimple or two. As you start to get a little more pimples, if you want, you can put words. It's getting a little more. This is just looking at the inflammatory phase. Getting more intense inflammation, more, and then more severe.

Now, I did this with a company who eventually decided they weren't going to do it, but it was an oral contraceptive. And they had a group of potential investigators, gynecologists and their nurse coordinators. And I went through a series of pictures with grades for inflammation, and they had a little booklet with those

pictures in it. And then I showed maybe 30-35 patients and asked them all to grade it. Having never done it before, it was amazing how easy it was for them to look through and match up, with very little discordance, on their first attempt.

So I think you can use this kind of system if you have standardized photography. All of us who do studies know of the Canfield systems. And you can have these kind of images which, when you see them, the way they do it, they're much, much larger, and you can count individual lesions or you can look at whether they got worse a little, a lot. They got definite improvement, marked improvement, or they completely cleared up.

And you can begin to get a sense of it with these photographs which again are not as good as what you can actually achieve. But you can begin to, I think, say, well, that patient is a whole lot better, and maybe you would put them in the almost clear and maybe somebody wouldn't. This patient is clearly better but is not anywhere near totally clear.

So you can use this kind of system, and we have used it in the past. A group of us, Alan Shalita, myself, Diane Thibitot, Guy Webster, and Ken Washinik looked at over 600 individuals with inflammatory acne. We looked at a subgroup over three days, a subgroup every day to see how

reproducible we were and what our intergrading variability was. Fortunately, our concordance was very, very high, and we were able to clearly delineate drugs from vehicles, as well as to see some differences between various drugs within a category.

So I think those kinds of things which many people are aware of and have been using for their own purposes but have not really used them in clinical trials yet because they kind of get the feeling that, well, this is what you got to do to get your drug approved, and once you start talking about modifications of the way it's been done, then all kinds of legitimate questions. Well, how do we know that that method is better than what we're doing? And so people have not really pursued them.

So my final slide here. I would say the time has come or soon will be here even for moderate acne where you'll have to consider positive control studies and/or at least significantly unbalanced trials in order to get by IRBs. I think the real question is, would you want your daughter in this study if they're going to have 12 weeks of no treatment? I think we have to consider possibly setting not only lower but upper limits for mild to moderate if we're going to have vehicle controlled studies persist, and only in the most severe forms are positive controls going to be used.

Either we eliminate this global assessment we have now which picks out only that small handful of people with monotherapy who reach total clearing or we bring back a comparative or dynamic kind of assessment using some of the advances in terms of imaging that all of us have become aware of that add to the ability to do this in a way that's meaningful and also consider a vehicle or placebo run-in.

I believe that's my last slide, Rob.

DR. STERN: Thank you very much, Jim.

The panel will have an opportunity to ask questions of our experts at the end of the four talks.

Our next speaker is Dr. Alan Shalita who is the Chairman of the State University of New York in Brooklyn Medical School, and he will talk about considerations on success criteria in acne trials. Thank you, Alan.

DR. SHALITA: Thank you, Rob, Dr. Wilkin, colleagues.

First I would like to tell a couple of stories so that one does not think that I'm being facetious in some of my remarks.

I had the great privilege, when I was a resident at New York University, to be allowed to go periodically down to the University of Pennsylvania and sit at the feet of Professor Kligman. And I remember grand rounds where one of the residents gave an elaborate

description of laboratory values on a patient trying to make the point that the patient had lupus erythematosus, and Dr. Kligman said, is she sick? And the resident couldn't figure out what he wanted, and finally he got the point across that lupus did have some implications other than laboratory values.

Well, I think the same thing applies to our judgment of acne. The bottom line is are these patients getting better or aren't they. And we can go through all the statistical manipulations and evaluations of lesion counts. I think that Jim Leyden's grandson and mine are a month apart, and I think that if you show them the pictures that Jim just showed you, that they could both tell you whether those patients got better or not.

I know that we need numbers and we need objective criteria to be able to evaluate something to get formal approval, but I also think we make it a hell of a lot more complicated than we need to.

Now, because Dr. Wilkin mentioned this earlier, I hadn't intended to show this slide, but I wanted to show you what the background noise is in acne because you alluded to it. This was a group of student nurses that we looked at about 30 years ago without any treatment. They all had acne. And you can see that they were getting a little bit better and a little bit worse at roughly 2-week

intervals, and it absolutely had nothing to do with the menstrual cycle in spite of a paper that I co-authored a couple of months ago. So that's background noise in acne, and there is a high degree of variability.

The other thing, shortly after Dr. Kligman and his colleagues at the University of Pennsylvania described the effect of tretinoin in acne, there were a series of clinical trials initiated. To the best of my knowledge -- and please correct me if I'm wrong -- this is the first drug that was officially approved as a formal NDA for acne. Everything else had either been grandfathered or was being used without approval. For example, I think the tetracyclines are still adjunctive use for acne.

But at any rate, so we enrolled patients in clinical trials and we did this at the New York University skin and cancer unit. I'm sorry. I want to come back to this. I apologize.

This was, I said, the original formulation. You can see that there was significant improvement in lesion counts. We didn't know any better and that was the methodology that they used at Penn. The company that put the NDA together used that methodology. But notice that there's a very, very poor vehicle response in spite of the fact that this is a fairly sophisticated and irritating vehicle. The obvious question is why.

My hypothesis is that these were all patients that were coming to what in New York was considered the mecca, the skin and cancer unit at New York University, and they had all been to three or four dermatologists. Their philosophy was prove to me that you can get me better. They also put up with irritation that the average patient in a dermatologist's office would not put up with. That's a side issue. It shows the motivation that they had to find a new drug to treat their acne.

On the other hand, this was a study done many years later in which I understand -- and this is strictly hearsay -- one of the reviewers from the agency told the company, why don't you market the vehicle? This happened to be 2 percent erythromycin in one of the original vehicles, which actually happens to be probably mildly effective in acne because had polyoxyl lauryl ether is in it which is a fairly potent substance. In point of fact, they were violating somebody else's patent and never could market this drug.

But I think one of the reasons one sees this kind of so-called placebo or vehicle response, the exigencies of doing clinical trials today basically because of the short patent life, by the time preclinical trials are done and a drug gets to phase III clinical trials, when a company decides to do a clinical trial, they want the

data yesterday because then they have to submit it to the agency. There's time to review it till the drug gets to market to recruit what has been estimated as a $500 million minimal investment.

So what happens around the country, you'll see people advertising in local newspapers, college dormitories, student unions, looking for volunteers for acne. For many cases, these are not volunteers that are actually coming to the doctor seeking treatment for their acne. It's what I call drugstore acne. And I think that the proportion of vehicle response increases almost geometrically in relation to the motivation. If the motivation is strictly that they're going to get reimbursed for participating in a clinical trial, then you have created a real problem in terms of vehicle response and that's where the placebo run-in can come in.

On the other hand, with the so-called placebo run-in or placebo washout, we once conducted a clinical trial in a reform school in Hartford, Connecticut looking at zinc. And this was published in the Archives where we said that zinc was ineffective in acne, and it probably is not ineffective. But the reason for that was these were kids that were all incarcerated for crimes related to narcotics or drug addiction and therefore probably very susceptible to the effects of drug. Well, after lactose

capsules for a month, they had 50 percent improvement. And it was pretty hard to prove that zinc was going to do any more than 50 percent because that's the average of what you get with most acne drugs. So that can be, depending on the population, a very dangerous route to take using the so-called placebo washout.

These are all confounding factors. I don't have a simple answer for you because if you're going to use real patients that are coming to a dermatologist for treatment, you're pretty hard pressed to use a vehicle control. Now, if you're using a drug such as oral isotretinoin for very severe acne, it's obvious that you're not going to use a vehicle and you can use a positive control. But that gets much grayer, as we discussed before, when you're talking about drugs for moderate acne.

I don't know why we're discussing mild acne. I didn't know that the agency actually regulates the OTC drugs, or at least not this division. It seems to me that most of the approvals that are being sought are for a little bit more severe than mild disease, but maybe that's semantic.

Then the other point I wanted to bring up -- and this has, I think, been emphasized a few times -- in the concept of clear/almost clear, which I think Dr. Leyden has spoken very eloquently about, we tend to use

polychemotherapy in treating acne, particularly moderate to moderately severe acne. But the submissions are going to be for monotherapy drugs for the most part, although you have some combinations.

Here was a classic study by the late Dr. Sidney Hurwitz, which had been published in 1976, showing that using vitamin A acid, or tretinoin, and benzoyl peroxide at separate times a day produced exceptional results, actually better than I get, but he was treating more of a pediatric population. And in other parts of the study, he showed that it was better than you could get with either drug used alone. So you don't get to the clear or almost clear till you use a combination of drugs in the most part, not always.

Then finally, in terms of where we're at -- and I think Dr. Leyden has demonstrated this very clearly, so I'm not going to belabor the point, just to show you a couple of different formulas. This was that series of photographs that he talked about where we looked at over 600 patients. I think it's pretty clear that this patient has improved, although it's not clear/almost clear, but there is significant improvement.

Again, I don't think you need a rocket scientist to evaluate these. We've had medical students look at these photographs. We've had nurses look at them

and non-medical personnel, and they've all come to the same conclusion.

Dr. Kligman I think is going to refer to it and did earlier, about some of the specialized techniques. This is just one. I think this happened to be one particular retinoid, but that's not what's important. I think the progression of improvement over the treatment period is very obvious. Again, one could try to quantify this, but you don't need anything else.

Finally, there are several other advantages I believe in using the photographs as a method for evaluation. Number one, it gives you a record that is permanent and not fudge-able. I'm not talking about digital photography which can be altered. But it gives you a permanent record of what actually happened. It gives you a confirmation of the investigator's evaluation, and it also allows for an independent third party, including the agency, if you so desired, to examine the results and say, this is a drug that works, this is a drug that should be on the market.

Thank you for your attention.

DR. STERN: Our next speaker has already been introduced at least five times this morning because of his eminence in the field, and it's Dr. Albert Kligman, one of the true luminaries in dermatology. Among his

contributions are those in the field of acne. And he's also from the University of Pennsylvania.

DR. KLIGMAN: Well, Dr. Stern's remarks validate what I have learned. If you live long enough, people will start to say good things about you. It's just a matter of age.

(Laughter.)

DR. KLIGMAN: I am 86 years old, by the way, and it demonstrates that the practice of dermatology is life-giving.

(Laughter.)

DR. KLIGMAN: My talk is about counts and counts are the popular, traditional, so-called objective way of demonstrating and measuring efficacy. The popularity of counts, of course, are obvious. You get numbers. Numbers bring joy to the heart of statisticians. You can make statistical analyses which gives confidence to regulatory agencies. We approved this drug because there was a statistical difference in the comparative assessments. So this is regarded as the gold standard, one of the objective, unbiased ways of assessing efficacy.

And I will tell you forthrightly that the most that could be assigned in terms of standards is bronze, after silver perhaps, but not much better than that. And the limitations are enormous. The accuracy and precision

has never been looked at. Worse than that are the reproducibility and repeatability of lesion counts. I know of no instance in which five different observers were looking at the same group of patients and their estimates correlated. There is no such objective evidence. Even within observers, the variance may be extraordinary.

We did a test years ago which I undertook in kind of a mirthful, mischievous way. We had Otto Mills who spent most of his days counting lesions and considered himself an expert. We had 10 patients with a mixture of lesions, and all he could see of the patient was a hole in a sheet. He could not see the patient and only this template. And he made counts, and then we scrambled all the patients and he made the counts over again. I am ashamed to tell you what the results were. The variance was enormous. He did very well on open comedones, big black lesions. They were easy, but for inflammatory lesions he did really very badly. So this method, as it now exists, is certainly full of difficulties.

Well, another way of knowing that the counting is an imperfect and difficult method and very unreliable is to see what the literature says. When you read the literature on acne comparative trials, if you are young and sensitive, you could get nauseated. If you're old like me, you just get cynical. It's just unbelievable.

May I remind you? And maybe you know, Dr. Stern. I don't think there's ever been an NIH-supported acne protocol. It's all industry supported. I'm not here to bash industry, but we all know that the capitalistic system often does not produce honorable people or results which are meritorious depending on how the study is set up. And that makes a very big difference in what you might see.

A recent review of all the papers that have been published in the last 50 years, based on evidence medicine, double-blind, placebo-controlled, randomized studies, about 10 percent of the studies that were reviewed fulfilled even minimal requirements for assessing efficacy. There will be improvements and the endpoints all mixed up. So it's kind of a mess. Let me give an example of how bad it is.

Azelaic acid in several studies was shown to be as effective as benzoyl peroxide in suppressing P. acnes and in clinical improvement. Anybody with experience knows that's nonsense. Benzoyl peroxide is a powerful antibacterial agent. In 10 days you get a tremendous decrease in the P. acnes count, and Jim Leyden has certainly showed that. And there's no comparison. And yet these studies were apparently conducted by responsible physicians under reasonably good conditions. That's just not acceptable. I could give you innumerable examples in

which equivalence is achieved for drugs which are completely different.

Another example, for example, would be 2 percent erythromycin against 1,000 milligrams of tetracycline orally. Three studies show equivalence. That's nonsense. Oral tetracycline beats the hell out of 2 percent topical erythromycin certainly in inflammatory acne. So that's kind of silly stuff.

And then another issue here is what do you count. Do you count microcomedones which you can hardly see? Closed comedones, open comedones, nodules, papules? Which kind of papules? Little ones, big ones? Dr. Pochi has already gone into that.

And in fact you have to decide many other troubles. Do you count the whole face or do you do it regionally? You have counts based upon the forehead, cheek, chin, and nose as Anne Lucky has sometimes indicated. You get very different results.

You also get very different results when you divide acne into categories, and there are many categories. We have all heard about the pleomorphism and the multiplicity of expressions, the phases of acne are so variable. If you start with early acne in prepubertal girls, they just have a few comedones. Boy, they do swell with comedolytic agents. Then you get into adolescent

acne. That's a little more difficult, and you get variable mixtures. And then you get into post-adolescent acne in females, and they tend to get lesions on the lower part of the face, and those are deep, ferocious papules and they're damned difficult to treat. So the outcome of much of this is depending on what you start with.

We have also heard about the placebo effect. Let me emphasize what Alan has said. You can't imagine a more labile disease which involves psychosomatic aspects. The psychological factors are profound, and the placebo effects are profound.

Alan, nobody showed you this, and both of us like to say in some of the drug studies where you look, you use the eyeball test. I'm a great believer in the eyeball test. When I see two curves and they're pretty comparable -- you know, there's only a little bit of difference between them -- I don't give a damn what the statisticians say. They may have all the power in the world. The confidence limits are wonderful. But the fact is clinically and biologically there's no difference when the curves are almost superimposed upon each other. In fact it would be possible to sell the vehicle with a perfectly good outcome.

I can tell you for sure that using exactly the same procedure, double-blind, randomized, the whole

religious stuff on how to do a study, that Jim Leyden is always going to get better results than most practitioners all over the world. And the reason is he's Irish, he's romantic, he's optimistic, we know how to treat acne. I've seen 1,000 patients. You just do what I tell you to do. He gets more compliance and he gets much better results. These are all part of the emotional difficulty, in fact, impossible problems to measure, and yet, they come into our concerns all the time.

Well, another thing that I want to talk about is what's already been mentioned. Acne is an astonishingly mischievous disease. It's very labile. Lesions come and go very rapidly. The life cycle of individual lesions is remarkably unpredictable. We have done a study using target areas taking digital photographs every 3 days. And this is something that's really difficult to understand, why it's so fluctuating, why it's so episodic. Those of us that have experience know this to be true. Sometimes you see a pustule come up in 1 day and 2 days later, it's gone. I'm talking about one area which is a target and we're measuring what's happening to each lesion. Other times you see a papule come up and it stays there for 2 weeks. Comedones will suddenly disappear. I have no idea what controls this kind of uncertain behavior, but it is certainly something that we have to take into account. Not

only do we made a global estimate, a severity estimate, but we should be able to follow individual lesions.

There are many biological problems that remain. I don't know why two pustules or papules that look exactly the same, one leaves a scar and the other does not scar. What the hell determines that? There must be some way for us to qualitatively assess lesions and predict what would happen.

All of this lability leads to what most of us have been saying. Use modern, highly precise imaging devices and a lot of this difficulty of classifying lesions, about their size, their shape, their color will all disappear.

Duration of the study is also extremely important business. If you're going to do a 3-month study and that's the end of it, as Jim has pointed out, it's not the end of it. But as a result of the fluctuating course of lesions, if 3 months is acceptable because that's all the companies can pay for and you at least get to some kind of a result -- they're moderately improved, greatly improved, or cleared -- you have to take multiple assessments. I can tell you it's damned near worthless to do a pre-assessment and then for the next 3 months, you just wait till the end, and you take your photographs, so you do your counts. That's almost worthless unless you

have a fantastic drug which clears 75 percent of patients which would be a very nice endpoint.

What's happening in between is absolutely important in view of the fact that most of what we do -- as Jim pointed out, it's not what happens to existing lesions. Let me emphasize what Jim told you again. Most lesions, if you don't do anything, comedones and papules, you don't do a damned thing but watch them, they spontaneously regress. So in therapy what we are really doing is measuring the inhibition of the evolution of new lesions. The existing lesions are going to get better anyway over a period of time. It's the prevention of new comedones, the prevention of new papular pustules that is really extremely important. All of these, of course, become issues.

Now, let me tell you what the most important thing is in counting lesions and why it's so variable. It's a tedious, onerous, bitchy business. It takes time, and if you see a study being carried out in a clinic situation, an office with patients waiting, and you have a technician, let's say, who's counting the lesions, I can assure you it will take at least a half an hour per subject to do the cheeks, do the forehead, do the chin and get accurate lesions. What mostly happens is that people are not experts, they're not seasoned, they're not well trained, and it's easy enough when the doctor is saying,

you know, you can't take a half an hour for a patient. I can't make a living unless you cut it down to 5 minutes or 10 minutes. And that's a reasonable assumption.

So what very often happens -- I've watched this -- unless you're in the domain of Anne Lucky -- when Anne Lucky makes lesion counts, you can damned well believe them. When most other people make lesion counts, they're up for grabs. You start looking at them and then the technician says, my God, well, it looks like 15 comedones and 20 papules look like a good idea.

What Jim has said is absolutely right. I have never counted papules in my life. I've always depended upon other people who have better vision and more patience. It's an extremely difficult thing.

I just want to show you a couple of slides to highlight some of the things that I've told you. Well, this is to tell you that the so-called placebo effect -- incidentally, there are no placebos in dermatology. That's another story I'd like to tell you about some day in bar-like situation.

(Laughter.)

DR. KLIGMAN: There are no placebos. Everything you do to skin, Nivea cream, any lotion, goose grease has a beneficial effect because it improves the stratum corneum. It prevents injury. They even have some

anti-inflammatory effects in their own right. You have a lousy stratum corneum in acne. It's punctured and you've got inflammatory lesions. Just putting the Nivea cream down long enough, in 50 percent of the cases in 3 to 4 months, a pretty damned good result, no activity whatever.

Here's a study that was done for 4 months using Cetaphil lotion. It's a non-medicated lotion. And just looking at the general assessment, well, 10 percent got excellent. Look at the good results. And you see that's a pretty good number in terms of percentages. We've got 40 or 45 percent of people achieving a pretty good result with what amounts to a vehicle.

The spontaneous events, the placebo effect here again is very important. Here's a study done by Lucky, who I think is an extremely rational and meticulous and vigorous minded clinician. This is a study on ethinyl estradiol. You heard from Dr. Bergfeld about the estrogens. Well, these are cycles. The difference between the hormone and the non-hormone, the placebo pill, notice that they're getting steadily better as you get up to five cycles. This is part of the placebo effect. The minute that patients are put into a study, when they're recruited, their compliance becomes better. If the doctor is a very supportive, cheerful doctor, then the results get even better so that the temporal effects always have to be

considered.

Well, Jim mentioned this and I certainly second it. A run-in effect is a very, very good idea, and I think it should be incorporated in the published outcome of this meeting. Recruiting people, putting them in a study and just using either nothing or a vehicle, you see what happens here with comedones and papulopustules. Before the study starts, minus 4, minus 2. There already is a significant reduction. You need to know what the slope of that curve is, what you're starting with. So I think it's a very, very good practical strategy for doing controlled, comparative studies, a run-in period in which you do nothing or you use a non-medicated medication.

I just want to show you a couple of little tricks that add to the fun of being an acneologist, if there is such a category. This is crazy glue, and what you do is you simply put some glue down on the skin and you cover it with a slide and then you let it polymerize. It's a cyanoacrylate, and you lift it off, and you see all that stuff, all follicular contents, hairs and sebum and horn, and any debris in the follicle comes out. And you can look at the slide and make some judgments.

I want to show you this because it shows what a smart lady Anne Lucky is. It was Anne Lucky who made us really aware of adrenarche, the time when prepubertal acne

is a real phenomenon and important to make the diagnosis because if you can identify high-risk patients who are in an early stage of acne, which happens to be comedonal acne, in girls as young as 8 and 9 years of age, one way to recognize such people in the prepubertal acne due to the secretion of adrenal androgens which promote growth of the sebaceous gland -- here is an 11-year-old girl who is not at high risk. Neither parent has acne. Neither parent has scars. So she's normal. And this is what the cyanoacrylate looks like.

I'm hot on this subject of pre-acne and pre-rosacea and identifying diseases years before they become clinically apparent. It's a favorite thesis of mine called invisible dermatology. As far as I know, I'm the sole practitioner of invisible dermatology.

Here's a normal person. Here is an 11-year-old girl without visible acne, a few little comedones in the nose and the forehead.

And incidentally, the pattern of acne is another thing, which is troublesome. This damned disease behaves in pesky ways. When it starts, it tends to start up here, and then the older you get, it sinks down. You get down to the point in post-adolescent acne which is in the lower part of the face and it's a lot more difficult, for reasons unknown to me, why the lesions on the lower

part of the face are much more refractory to treatment than the upper part of the face.

Well, you can see in a moment this kid is in trouble. The time to treat her is right then and there with a comedolytic agent, and our preliminary study shows that that works very well.

Another way of doing that is to look at sebutape and just look at the number of dots. We can image analyze this, determine the density of sebaceous follicles, how much they're making, the size distribution and do all the statistics. This is the same girl I showed you who is cyanoacrylate positive. She's making sebum. If I showed you a 1-hour sebum excretion rate on sebutapes of the control person, you will see little or no droplets.

And here is looking at the sebutapes with a fluorescent light for porphyrins and you look at it with porphyrins. And that's another way, incidentally. Another possibility of looking at acne is to just turn out the lights. Let your eyes get accommodated and look at it with a Wood's light and see how many follicles are fluorescing. It's another attribute which is really quite useful. It's a nice little trick.

Here is post-adolescent acne, and I think now that there are more women with acne, troublesome, deep papules, than all other forms of acne. Post-adolescent

acne in females is increasing in prevalence and is a very important thing. Notice that she's got some lesions up here, but many of the ones are down below and they're tough to treat. And the reason that they're tough to treat is when you do a biopsy -- and we would like to avoid doing this. We now have, believe it or not, things that you have never heard about, optothermal coherent tomography. We can outline without touching the skin just what this lesion looks like from the surface down. Confocal microscopy does the same thing, and we can make cuts without touching the skin, all optically done, which is going to increasingly give us the kind of resources that will enable us to make the comparisons that we're interested in.

Finally, this was brought up. When you talk about acne, you have to define blacks, orientals. It's a common belief among dermatologists -- and because they are dermatologists, they have many, many myths that they have to deal with -- that acne in blacks is less aggressive, less important, less scarring. That's absolutely wrong. Halder and myself at Howard have shown that that's not the case.

And here's a good example in the case of a black person. If you take a regular photograph like that, well, you can count those papules and pustules. That's pretty good. But the fact is if you look at digital

photography, which faces all surface contours -- you don't see any micro-topography. All the surface texture is obliterated. So now you're looking beneath the surface. Then you can see that there are many more lesions than you saw before and that each of these lesions are a great deal more disseminant. They have spread well beyond what you see on the surface. This is just an example of what you can do with digital photography.

So my message is this. We really have a repertoire of drugs for the treatment of acne which is really superb. You know what you're doing. You have a tremendous choice of oral drugs and topical drugs. And we now have within our hands, if we just bring about the necessary resources, to take this pleomorphic disease, this disease with so many different expressions, and really establish criteria rigorously defined, all the things that we have been talking about, and to make assessments which are reliable and believable and which will allow regulatory agencies to make their approvals based on objective science.

Thank you.

(Applause.)

DR. STERN: I'd like to thank all four of the speakers for giving what I at least thought were extremely lucid, informative, and fun to listen to presentations.

We're now open for -- yes, I'm sorry. Could I ask the four speakers to come over to the side so it will be easier for us to ask them questions and for them to respond?

Dr. Kilpatrick.

DR. KILPATRICK: I don't really have a question as yet. I may come up with one as I think. But I wanted to inform you, sir, that statisticians have gone beyond the level of development in our subject that we can deal with categories, ordinal or otherwise, as well as counts or measures. So there are techniques and perhaps Dr. Tan and other statisticians here will come to this as we come to the quantitative aspects. Thank you.

DR. STERN: I want to address a question to Jim. My own biases are very much along yours in terms of the need for objective photographic assessment. In fact, that's done by people who weren't involved in the investigation who are blinded to both the temporal order of when the photographs were taken and also obviously what treatment group they were in.

One question I had, though, is you mentioned the use of photographic or digital images for doing dynamic assessment by investigators at the time. One of my observations has been that when I look at a photograph of an individual taken very recently where there couldn't have

been much change in their clinical status, they often look worse in the photograph. There's something more impressive about many clinical conditions on a photograph than if two days later you look at that person in vivo. And I was wondering if you could comment. Is that just my own bias or have you tried to look at it?

DR. LEYDEN: Well, I haven't looked that soon. The soonest I've looked at is a month. I mean, I think the criticism that was covered this morning about the former ways where people were judging how much better they got based on memory -- you know, you can't remember. You had to have some kind of interaction with the patient and kind of look at the case report form and see whether they got better or not. So it wasn't very distinct from what was already done. But I think now, as Alan pointed out, if your grandson can tell you that they're better, they're probably better.

I would just stress again I think clear or almost clear doesn't tell the whole story and greatly understates the value of drugs. It seems to me that what should be done is something should be done to see whether or not drugs are safe, number one and two. Are they safe? Are they safe? And number three is do they work. Not how much do they work. Are they better than what we already have or a big step forward or a little step forward? Those

are things to be decided by us in the clinic in combination with other drugs when you have a multi-factor disease, not as monotherapy. Monotherapy just establishes it has activity, and then we decide whether it's good enough for us to use sometimes, all the time, or never.

DR. STERN: Other questions?

DR. KATZ: Jim, I have a question. I wasn't aware of studies -- you probably know of some -- where the drug is evaluated as whether it works or is clear or almost clear.

DR. LEYDEN: They're not presented to us by the pharmaceutical companies in that way, but the approval process for the last whatever number of years -- you know, eight or so -- has been the global assessment. Whether there was statistical difference between the vehicle and the active was based on complete clearing or almost complete clearing. That's the way it's done, but that's not the way it's presented to you.

DR. KATZ: Most of the studies or all the studies that I see in the literature are 50 percent better or they're --

DR. LEYDEN: Yes. The last century.

DR. KATZ: Those studies that I remember that are presented in the literature that are only almost clear or clear or 0, there's a certain amount of improvement, how

many people get clear and how many people get 50 percent better.

DR. LEYDEN: Well, Jonathan can tell you that right now, as of so many years, the criterion for clinical success has been complete clearing, absence of disease, or almost complete clearing. And the qualifier to that would be such that further treatment would not be indicated. That is the current standard.

DR. STERN: Is that in fact the case, Dr. Wilkin?

DR. WILKIN: That's essentially correct, and as it turns out, in the acne studies often there aren't many subjects who fall into the win category, if you will, on global in either the active group or the inactive, the vehicle, group. What we ask for is it doesn't have to be a majority. It just simply has to be a statistically significant proportion of those who are in the active got better compared to the proportion of those who were in the vehicle who got better in terms of that dichotomous cutoff.

Now, I think actually it was Dr. Leyden that earlier made the point that that is an easy cutoff where one can look and see the difference between whether it's a 1 plus, 2 plus, or exactly what. It's a little bit more, if you will, objective than perhaps some of the other changes in grades through that kind of scale. I think that

that's basically part of why the agency began using that way of looking at it.

That's not to say that someone who doesn't make it all the way down to the almost clear or completely clear category isn't a success. I mean, someone may get something less than that and they may feel happy with it and they might need some other form of therapy.

But sponsors come in -- again, I can never remember a sponsor coming in and saying I want my product only for this one lesion type so that dermatologists can use it in sort of their polytherapy.

Having been in practice in Houston and Richmond and Columbus, Ohio, I can say I got an awful lot of patients who came after being seen by general practitioners, and I don't think in general they practice the way Dr. Bergfeld described at the beginning. I mean, I just have not seen general practitioners picking out lesion types and targeting that. I think we have the best experts in acne in the world here today, and they're describing to you not a bronze standard, not a silver standard, but -- and it's probably not even gold. It's probably the osmium. I think that's the most expensive element. It's probably the osmium standard for treating acne.

And ultimately one of the questions that the committee will need to think about tomorrow morning is what

kind of indication really fits for these kinds of products. Are we sending products out for this small subset of osmium-standard practice, or is it for really the bulk of the practitioners who are using these products out there who are not dermatologists? I think that's pretty clear.

DR. LEYDEN: Could I just comment on that? The other thing is that dermatologists figure how good drugs are or aren't. Those of us who have been around long enough know of several drugs that were out and are no longer on the market. They got approved, but they didn't make it. There are drugs that get out there that have a very small market and they never increase, they have a tiny use, and then there are other drugs that are used very commonly.

So I would just say again I think the aim should be to establish the safety and whether or not there is efficacy, not how much efficacy or how good it is. That's up to us to decide.

MS. KNUDSON: I'd just like to ask about inclusion criteria. We mentioned several times the population of patients that are included in trials. Do you make a distinction between naive patients, patients who've never been on any therapy, and patients who might have failed other therapies?

And then my second question is, how do you

control for all of the over-the-counter medications that are available for people to take? And certainly if someone is in a trial for a long time and they're not immediately getting better, I suspect they're also using over-the-counter remedies. So how do you control for those things in your outcome assessments?

DR. SHALITA: If I may respond at least to start. You're absolutely right. Compliance is an extraordinarily important issue and unfortunately we don't have a good way to measure compliance. The most popular measure is to have the volunteers bring back the empty tubes to see how much they've used. Well, they're not stupid and they know they're not going to get paid if they bring back a full tube. And they do what I refer to as the sink test.

Dr. Bergfeld showed a paper of actually mine or I was a co-author on it where two drugs were compared. One was shown to be more effective than another, which is not terribly important. And they were shown to be roughly equal in side effects in spite of the fact that one of those two drugs was promoted as much, much less irritating than the other. Well, the answer is they didn't use the irritating drug, but you couldn't tell that by measuring the empty tubes because they're not going to let you know.

In terms of what else they use over the

counter, they sign a consent that they're not going to use anything else, and you tell them. But there's no way to control that unless you have a captive population like we did with that zinc study in reform school. We actually put the medicine on. They have no access. And that's very difficult to do.

The final part of your question. We'd love to be able to use people who have not responded to prior therapy, but in real life it's very difficult to do that.

DR. KLIGMAN: Can I add something? There's ample evidence that dermatologists are much more effective in diagnosis and treatment than general practitioners. And I think, Jonathan, in the regulatory requirements, these kinds of studies should not be monitored by general practitioners. They just simply don't know enough, and they're very often affected by other things.

For example, they like drugs that are non-irritating. Most non-irritating drugs are less effective. In fact, there is some relationship between the amount of inflammation induced in the case of retinoids and of efficacy. If they are influenced by the notion this is a nice drug because they're not complaining of stinging and burning and redness and all those adverse effects, that shifts their bias toward drugs that really don't work. So I hope that having an M.D. doesn't qualify you to become an

acneologist.

DR. LEYDEN: There's one point I think that might be worth mentioning to the panel, and that is that I can tell you that at least in the last maybe 8 or 10 years, every company that I know of who has been involved in a clinical trial, when they've had investigator meetings, they have conducted sessions where they establish the reproducibility of counting lesions, both non-inflammatory and inflammatory. It's a big part of what they do. And the reason that they've had to do that is that in order to get enough patients into a study, they've had to expand the number of investigators and sites because all of us are having trouble getting patients. So if you're going to expand the number of sites and you can't have three or four or five centers doing all the studies, you have to make sure that people know how to count, and they are doing that.

DR. BERGFELD: I'd like to go back to a little bit about the FDA standards and these tests first. What I heard was that one of the endpoints was a 3-month treatment and no need for further treatment as being one of the targeted endpoints. Is that correct?

DR. WILKIN: Well, if you're talking about the global, the global has come in different ways. I would say some of the globals that have been used, the success

criteria included patients that probably still wanted more treatment. So, no.

One of our difficulties is we don't have one global that we recommend industry use. We are really coming to the committee to find out if there is a global that the committee that would recommend that we recommend to industry.

DR. BERGFELD: Well, I would like to, as a consultant, recommend that that not be used because if we truly believe that acne is basically familial and it's driven by androgens, which are high in the adolescent and in some of the women are high in their older years, that we have a continued hormone stimulation for this and that does not go away in 3 months.

DR. WILKIN: Let me be more responsive to your question then. I took it to mean would someone need any additional treatment, meaning in addition to what is being tested. I realize if someone discontinues a product that has got them under control, they're likely to have a flare. No, that's not what we're asking.

DR. BERGFELD: Well, the second part of that in your statement is you heard today from everyone who's a dermatologist that it's polypharmacy that we use that is most effective, and obviously after a study those patients will then resume the polypharmacy which includes the

topical agents and some systemic depending on the degree of acne. That's one thing.

The second question and sort of statement I'd like to direct to the experts. If you were to design an ideal study, it seems to me that what you've all said is that it should be simple. The second part is that there would be some lesion counting in some way and they would be differentiated between non-inflammatory and inflammatory. And, Dr. Leyden, you suggested there be two different judgments made, not that they be combined statistically, and that we use current technology that has been mentioned by all of you and that includes some of the new photography methods, digital photography.

DR. LEYDEN: And that you draw a more clinically relevant line in the sand of what constitutes success because I think you can have great success without being anywhere near almost clear, especially when you have monotherapy.

DR. BERGFELD: Would there be any additions to what I've outlined, other than Jim's?

DR. ABEL: I'm asking the members of the panel if they feel the sponsors might seek approval for different lesional types, comedonal versus noncomedonal, inflammatory --

DR. LEYDEN: I think that's likely to evolve

because we are now in the age of the development of non-corticosteroid anti-inflammatory agents, but until somebody discovers that some of these things can't or shouldn't or whatever be used on the face, so far that's one of the reasons why we use them is that they don't have the problems that steroids have. I'm hearing every place I've been how dermatologists are. As soon as there's a new drug, they try it on everything. Dermatologists have already decided those drugs work. Now, the manufacturers may just say the hell with it, why bother with all this stuff, let them do it.

But if they decide or if some of these other molecules that are not yet approved for any indication were to be used -- and I know one company who has several molecules that make a lot of sense to me. I can't imagine them having an effect on the non-inflammatory part of acne. If it happens, great. It will give us something to think about. But right now I can't just imagine that. So to say that they're going to have to do a study that shows effect on non-inflammatory lesions to me is ludicrous.

DR. KING: I guess when I thought about this conference, I came up with the thought, that if you're going to generate a new system or a consensus, then you're going to come up with the issue of innovator versus generic products. What kind of approach would you take or give

guidance to the FDA about if you're going to implement a new system, what standard would you hold for the innovator versus generic products to give guidance?

DR. LEYDEN: I have an easy answer which I know is not popular with the dermatology division. I have a great deal of difficulty thinking how a product that is absolutely identical is different clinically. I mean, it just doesn't make any sense.

So if I were doing it, which I'm not, what I would do is just show that this formulation has the same release characteristics, penetrates in skin, Franz chamber or some modification of that, to say that it's not fundamentally different because of some quirk in the manufacturing process, et cetera. That is the way it's done for solutions. As soon as minoxidil went off patent, there were generic formulations within a week because they didn't have to do anything.

Nobody can agree upon a surrogate method so far other than doing clinical studies which are laborious and difficult and expensive. How the same formula can be different I think just brings up all the issues of clinical trials, and whether it's tinea pedis or eczema or acne or whatever it is, it's just not easy to clinical trials.

DR. KING: A related question, but to follow up, then how do you decide when you're doing dose response?

We know about how enzymes respond and they have parallel curves. So I disagree with the concept of parallel curves don't mean statistics, but they do. But how do you deal with the dose response in even the same drug?

DR. LEYDEN: I think there you have to look for a non-effect or a low effect and a dose above which there's no increase.

DR. PLOTT: I have a question for Dr. Leyden. You suggested eliminating global assessment.

DR. LEYDEN: As it currently stands at least.

DR. PLOTT: As it stands. And replacing that with kind of a comparative pre-post --

DR. LEYDEN: Dynamic.

DR. PLOTT: Dynamic --

DR. LEYDEN: Or leave it out. One or the other.

DR. PLOTT: Well, assuming that you have it in there, is this scale of better or no change simply a different dichotomization to say --

DR. LEYDEN: Well, I'll tell you what we did in a study that Alan and I were involved in. We decided that a two-grade change was clearly something that -- I'll say it negatively -- nobody would disagree was not meaningful. And they all constituted people who had at least a definite or marked improvement. So you can do it a couple of ways.

DR. PLOTT: So you agreed upon a clinically meaningful change --

DR. LEYDEN: Yes. It was easy for us to do. I guess it's more difficult when you're in a regulatory position. You have to be careful because when you deny somebody approval, you have to be prepared to defend it. So it was easier for us to make that decision, I recognize, but that's what we did.

DR. STERN: Just a quick clarification. This is two grades out of your six grades?

DR. LEYDEN: Yes.

DR. STERN: Just because there have been so many scales --

DR. LEYDEN: Yes, of that. Yes, right, exactly. And looking at the photographs, we all said, yes, that person is better. We didn't have, well, maybe they're a little better. That person is better.

DR. STERN: I understood that part.

DR. LEYDEN: Yes, two grades.

DR. STERN: There have been ones from anywhere from 4 to 10 grades within the scale.

DR. KLIGMAN: Dr. Stern, another source of mischief -- and Dr. Kilpatrick can respond to this -- is to put the data not in absolute numbers but in percentage differences. I think that's really unacceptable. And

that's done very often because it's easier to make the drug look better than it is.

If you go from 4 pimples to 2 pimples, that's a 50 percent reduction. That's great. If you don't know what the actual starting condition was, the number of lesions, and the actual number of lesions at the end, you can end up recommending drugs which are damned near ineffective, and it's very often done. Instead of giving real numbers, you get percentage differences from the baseline.

DR. KILPATRICK: Dr. Kligman, may I answer that at length this afternoon?

DR. KLIGMAN: Yes.

DR. KILPATRICK: Because there's a lot going on here and I think some of the rest of us may want to get in on this. But repeatedly -- I'll just say this and then stop talking -- the thing that I keep hearing is the difference between clinical significance and statistical significance. I think that will affect what we come up with in terms of our recommendation.

DR. KLIGMAN: That's true.

DR. ABEL: Getting back to monotherapy versus combination therapy, most commonly dermatologists use combination therapy from the beginning and different types of therapy for the inflammatory component and different

types for the comedonal component. So I guess this is more of a question for the FDA. Would they consider using different standards for drugs which are not usually given as monotherapy?

DR. LEYDEN: It's a tough thing to do.

DR. WILKIN: Different standards? Well, in other words, you're saying if a sponsor comes in and says we would like monotherapy because we know a lot of docs are going to use only this product, would that have different standards than, say, another product might get if that sponsor comes in and says, well, we'd like this to be only for inflammatory lesions. Is that the --

DR. ABEL: No. I think it's more the disorder, the acne being the type of disorder it is, that most agents are used in combination with other agents. So would this affect your bottom line response criteria necessary for this drug to be approved? Could it be lower than, say, completely clear knowing that it is going to be used in combination therapy because there are different elements to acne?

DR. WILKIN: There is a word for that. I mean, the word is adjunct or adjunctive. In other words, if a patient is already on a particular product and then you look and see what adding a second product can do in addition, yes, I could see that as having some different

ways of looking at it. But it would be in the indications section of the labeling. It wouldn't be that nice, clear-cut, marketing-friendly, you know, treats all of acne kind of indication that sponsors are now seeking. It would be more limited. It would say adjunctive. The benefit is documented while using -- and then another product or class of products.

DR. ABEL: That might be more realistic.

DR. STERN: Dr. Raimer.

DR. RAIMER: I just wanted to ask our panel of experts, who have done a lot of studies, do you think it would be at all practical to count inflammatory lesions by size, like count the number up to, say, 3 millimeters and the number that were 3 to 6 millimeters or above? So if you started out with a patient that had 50 5 millimeter lesions and they went down to 50 1 millimeter lesions, that's definite improvement. Do you think that would be at all practical to do?

DR. LEYDEN: No.

(Laughter.)

DR. RAIMER: And why not?

DR. STERN: How about with digital photography, though?

DR. LEYDEN: You might be able to do it better with image analysis, yes. The volume of the lesion could

be determined. But it's hard enough to count them accurately on the hoof. The patient wants to get out of the room. They're embarrassed. They start looking down. They want to leave. They just want to get out. So if you're going to have to start sizing them, it will never happen other than on photographs or --

DR. KLIGMAN: And dermatologists have to make a living, you know. There's a matter of time.

(Laughter.)

DR. TAN: Yes. I just want to ask the panel. Dr. Kligman mentioned the inhibition of new lesions, emerging lesions is important. I just wonder how this is incorporated in current lesion counts.

DR. LEYDEN: It's done over time. You do it over typically a 3-month period except for oral contraceptives. So anything that comes in month 3 is new or month 2 is new. What he was saying is that you don't want to just do a count at the beginning and the end. You get a better overall view of the change by counting at multiple time points.

DR. TAN: But it will be hard to track individual lesions because some of those are hidden.

DR. LEYDEN: Some of them what?

DR. TAN: Are hidden. A few months ago you wouldn't see it. Right? It would be hard to track how the

individual lesions change.

DR. LEYDEN: When they're gone, they're gone. There may be a residual pigment or residual redness that gradually fades, but we don't count them, as was mentioned a couple of hours ago. Somebody brought it up.

DR. STERN: Dr. Katz.

DR. KATZ: Jonathan, a question. I just want to get something clear because it's not logical to me. You mean products are approved only if they get people clear or almost clear? There are so many things on the market that have been approved, except for Accutane, that don't make people clear up. I don't understand that.

DR. WILKIN: Yes. I was hoping to clarify that in response to Dr. Bergfeld's question. We have what are called end of phase II meetings with industry and different things get proposed to FDA, and ultimately what we convey back to industry is we agree with this. If you do these sorts of things, we will find efficacy in that. On the other hand, if they sometimes can fall a little bit short of that, they may still get approved. We can be really definitive on things that for sure are strong enough that we know that they're going to cross home plate right at waist level and right down the middle, and that's what we describe. But there are some things then that sometimes go to the edge of that. So there may be a product that

doesn't often lead to almost clear, but nonetheless when you compare that product with its vehicle or with its placebo, if it's an oral, it may have a statistically significantly greater proportion who fell into that success criterion than the inactive.

So I think it's just as Dr. Leyden has pointed out and as you've mentioned, that these are not like another therapy that you mentioned that can completely clear and perhaps keep things completely clear. Dr. Leyden mentioned that the marketplace is actually more Darwinian in what happens to the eventual success of these products than coming through FDA.

I think one of the things that we really want to hear from the committee is where are we with what we've been doing. That ought to affect where you think the compass ought to be set or the goal posts, how wide they ought to be tomorrow morning when you consider this. Do we have the goal posts too narrow? Or do you want some products that might even be somewhat less effective than what we currently have going through? Or do you want it tightened up a little bit?

I think the other part that especially the invited experts have articulated today is what if we could have monotherapy really sort of being moved into a polytherapy which fits with the practice of many

dermatologists. I think it would necessitate a different kind of labeling structure, you know, products that are for specific acne lesion types. I think that would be fine if the committee believes that that's the way American physicians in general, not just the osmium standard dermatologists who are here, but in general that's how it's going to be, or if you think there's something we can do in labeling that will help maybe bring non-dermatologists up to that standard of therapy.

DR. STERN: I'd like to give Jim a chance to make a closing comment, but I think we should close this part of the program. Preceding even his comment, I'd just like to thank the expert panel who took their time to come and educate us and were so helpful and clear and straightforward about their opinions and have been, at least to me, extremely helpful. It's also nice to get to see all of them.

DR. LEYDEN: You don't really think we're opinionated, do you?

(Laughter.)

DR. STERN: I don't think. I know.

(Laughter.)

DR. LEYDEN: I was just going to say I think trying to do studies where you take multiple classes of drugs for a new drug will be kind of a nightmare situation.

I think from my viewpoint anyway, it's more realistic to try to modify what's been going on to a more clinically relevant endpoint, perhaps using some of these newer ways of evaluating efficacy, rather than trying to design studies where you're going to have this new drug added to a certain other -- I mean, don't do that.

DR. STERN: Again, thank you all very much. We'll come back at 35 past the hour, 45 minutes from now, and resume after lunch. Thank you.

(Whereupon, at 12:52 p.m., the committee was recessed, to reconvene at 1:35 p.m., this same day.)

AFTERNOON SESSION

(1:44 p.m.)

DR. STERN: The first presentation of the afternoon will be by Dr. Alosh of the Food and Drug Administration, and he's going to speak in two parts. So we'll have his first presentation on statistical analyses of acne clinical trial data, questions about that. Then he'll give a second part presentation and questions about that to follow.

DR. ALOSH: Thank you. Good afternoon.

The stat presentation, as Dr. Stern pointed out, will be two parts. The first part, I'll be speaking about efficacy assessment, evaluation in acne clinical trials, where I'll be touching on some of the issues which were raised this morning concerning counts, change in lesion counts or percent change. I'll be touching also on the efficacy assessment by baseline category. I'll stop, take some questions. Then in the second part I'll be speaking about global evaluation and how it's related to lesion counts.

The first presentation is joint with my colleagues, Kathy Fritsch and Shiowjen Lee, from the team.

The outline of my presentation is as follows. I will revisit choice of the primary endpoints from a statistical point of view. I'll be discussing the

statistical analysis methods and data transformations, and I think this is very relevant because we had a lot of questions this morning about the appropriateness of using percent change. It was raised twice.

Then the other point, which Dr. Wilkin pointed out, whether we should take multiple assessments instead of just taking the final assessment. With that approach one could increase the power of the study. But there are issues which we need to address.

I'll be, as I said, talking about the effect of baseline severity, and this really came from questions raised by industry, and Dr. Leyden in particular, whether we should have people with a smaller number of lesion counts for enrollment in the study. So I'll be examining the efficacy results across categories by breaking people according to the baseline severity.

Then I will conclude with final comments about the statistical analysis.

The primary endpoints, as the discussion came this morning, we talked in terms of lesion counts in general or in terms of the investigator global assessment. When someone speaks in terms of lesion counts for the statistical analysis, we look for inflammatory, non-inflammatory, and total lesion counts. And the discussion came this morning whether one should analyze only

inflammatory or non-inflammatory without the need for total lesion counts.

I think Dr. Ten Have also questioned the rule to win in two out of three, whether there is a need for a multiplicity adjustment. I'd like to point out really the interpretation for two out of three is a nested hypothesis approach. So first you need to win on the total lesion counts. And now if you win on the total, you go to the subhypothesis to test whether you have a result for inflammatory or non-inflammatory. So with that nested approach, we don't need a multiplicity adjustment.

I think concerning the discussion this morning here, if someone wins on inflammatory and has a trend in non-inflammatory, you will be winning in the total lesion counts. So consequently, the drug will get the acne indication in general. Similarly, if you win on non-inflammatory and you have only a trend in inflammatory, you will be getting the general indication.

One of the issues, which I think the committee needs to think about, is whether in the study at the design stage you need to claim for the two types of lesions, for inflammatory or non-inflammatory, and if you don't win in one of them, how would you adjust for that. So those issues probably need to be discussed later.

So now once we have each type of lesion count,

whether inflammatory or non-inflammatory or total, you could analyze the final lesion counts by comparing the active versus the vehicle and look for a statistical difference.

I would like also to touch on the point of the discussion this morning that we should look for safety and if there is efficacy. If the vehicle itself has efficacy, then one needs to judge the magnitude of the difference between the active versus the vehicle. So the point I would like to bring here is that the vehicle itself will show efficacy.

Then the second one will be analyzing change from baseline and we could analyze percent change. There was a lot of discussion whether percent change is appropriate or not. I agree with Dr. Tan. A statistician would not prefer such a measure. I would agree that it does not have normal distribution which is what we look for in terms of statistical hypothesis testing, and I'll be touching on that. But really we were driven by the clinical request in a way. This is the preferable measure, but for a statistician I would agree that percent change is not the ideal measure to look at and I'll examine the data in a short while.

Then the other endpoint is the investigator global evaluation. In the first part of the presentation,

I'm not going to discuss the investigator global evaluation, but the second part will deal with that.

When you analyze percentage of change, there are pros and cons. Definitely change is easy to interpret and analyze, and the goal here is to attempt to remove the influence of baseline counts, how it will affect the final assessment. The cons of that, baseline may still have influence since change is negatively correlated with the final counts. The point which was made this morning, when you look for change or percent change scores, it may have highly skewed distribution. There will be a heavy tail distribution. With that, probably you don't need the .05. It might be not precise which we use for symmetric distribution for normal data.

Coming to present to you some data from acne clinical trials, as we have discussed this morning, there is a large variability in acne data. So it's difficult to choose one drug or one data set which will be representative for the acne data which we see in practice.

With that in mind, I tried to present here data sets from two drugs and will show you the range of what's the delta, the magnitude of the delta you need to reach statistical significance. Also, one of them has led to a very small p value, highly significant, but the other one is not. We'll see one of them at work on inflammatory

lesions, but the other one at work on non-inflammatory lesions. One of them, the study was for 12 weeks; the other one was a contraceptive drug for six cycles. So with that representation for the data from the two drugs, I think you should get some good idea about the range of variability in the data which we observe in real life.

Here the first drug we'll call drug X. We have a plot here. The study was for 12 weeks with about 400 subjects enrolled in the trials. There is an evaluation done at weeks 4, 8, and 12. So what I have here on the x axis is the week, and I have on the y axis the mean lesion counts. This is broken by inflammatory, which is the red line. The solid line is for the active, and the dotted line for the vehicle. So we have lesion counts over time for inflammatory for the active arm as well as for the vehicle.

If you could compare the lesion counts, you see a very small difference here. It's, I'd say, roughly about 2 lesion counts between the active and the vehicle. We'll see the impact of this in the p value.

The blue line represents the non-inflammatory lesions, and you start to see here separation. This is the magnitude of the difference. We are looking between the active and the vehicle, which we'll see about probably 6, 7 lesions. The total, which is the black line, which is the

magnitude of the difference, about 11 lesions. With that magnitude of difference, we see drug X resulted in a highly significant p value.

The point I want to make here is you could see subjects who are on the vehicle, as was indicated this morning, will achieve some kind of efficacy. So the point that we should look for efficacy, disregarding the magnitude, one needs to tell how much difference between the active and the vehicle because the vehicle itself, as you could see, has an effect there, as indicated in the morning.

So this is for drug X. I'll move next to drug Y.

As I have indicated, this was done in 400 subjects. It's for six cycles. It was a contraceptive drug. Again you have the red line for inflammatory, blue line for non-inflammatory, and total. And you can see the difference between the active and the vehicle here a little bit bigger, and you see this drug will make it even with about a 3 lesion count difference in inflammatory lesions. This is about 5 lesions. Here we have non-inflammatory, and the total about 8-9.

So if we're analyzing final lesion count, we'll be comparing, as I said, the active versus the vehicle at the final study endpoint. If we are analyzing the change,

we might take the baseline measurement minus the final assessment, which will give you the magnitude of change. And if you are analyzing the percent change, you'll take the change divided by the baseline.

Here we have plot for inflammatory counts by baseline which is on the x axis, and what we have on the y axis is the inflammatory lesion counts at week 12. This is here for the vehicle arm. I will have a similar plot for the active. What we have here, the 45 degree line. People below this line achieve reduction in terms of inflammatory lesion counts. People between the 45 degree line and the other line experience an increase in their lesion counts between 0 to 100 percent. Of course, the closer you are to the 45 degree line, there is no improvement. Here you could see people with an increase over 100 percent.

Just to make the point about percent change, let us take this dot here which represents a subject. You can see the subject at the baseline. They have about a 10 lesion count, but at the final assessment at week 12, they have roughly a 60 lesion count. So if you calculate the change, it would be about a 50 lesion count, and the percent change will be about 500 percent.

Now, we had the discussion you could have one subject like this subject to account for so many patients here in this group because the percent change here is very

small numbers, compared to one subject that would have 500 percent. And you might end up having a few patients driving the results. The impact of this -- you can see here a lot of scattered points in that plot. You would be increasing the standard deviation for your percent change, and we need that to calculate the statistical test for efficacy assessment. So in addition to the magnitude of change, we would like to look also to the scatter or the dispersion of those data, i.e., the standard deviation. So keep in mind how much variability scattered points here for the vehicle.

And the next plot, we'll see the same plot but for the active arm. You can see here for the active arm, again it's for inflammatory lesions, and you can see we don't have much variability for those lesion counts compared to the vehicle, and you can see much more improvement here in this section. We don't see people here with increasing their lesion counts over 100 percent. You see the scatter is less, so you expect the standard deviation to be less here.

Those will bring the point with those outlier observations whether one should analyze original data or some type of transformation of the data. And dealing with the transformation, we got a lot of ways from sponsors for what kind of transformation to be done. Sometimes we get

people have proposed to use log transformation or add constant to the log transformation. Sometimes we have ranks.

And I want to make the comment about using log transformation or adding constant to that log transformation. It's difficult to interpret when you have log transformation. I mean, there is no interpretation which I see reasonable to convey it to a non-statistician. I don't see its appeal.

Also adding a constant is subjective. Someone could add 10. Another one could add 20, and you would lose a lot if there is any constant which you could add.

The third point I want to make, this type of transformation can data dredging. In a way you have to wait until the study is completed, and now you'll go and see what transformation will bring this.

So the point, percent change needs to be used and if it does not meet the normality assumption, normally what we'll take, the rank transformation, and the way you order the data and by working with the ranks, you get rid of the magnitude of those outliers.

Here the point we are making, if you analyze percent change, you can see the trend over time from week 4 to week 12. And those quantities here represent the standard deviation, and you can see the magnitude of the

standard deviation is very large compared to that.

So to summarize, because of those outliers and percent change, we tend to analyze, in addition to the original data, transformation and, in particular, the ranks.

What I have here, people in acne trials, as has been discussed in the morning, experience a flare. In a way you could come at one time point and the subject have many lesion counts, and you could examine at another time point. Those lesion counts disappear. This again raises an issue in terms of how you analyze those data.

To make the point here, I'm taking data from study X for one investigator, and they have here about 8 subjects. Every line of those represents the time trajectory for a patient, total lesion count. So you can see here the blue line. You have the subject experienced a high lesion count at week 8. Then it dropped. Similarly the red line here, this subject at week 12 started to show a high increase in total lesion counts.

This brings the point whether we should take some kind of average repeated measurement toward the end of the study once the drug reaches its plateau, instead of dealing with the final assessment. The point here which needs to be discussed, once you decide on using a repeated measurement, you need to consider how many time points you

are going to take into account in the repeated measurement. Definitely you could increase the power by having several repeated measurements just because you reduce the standard deviation, but also I think a clinician would like to see clinical benefit not only reaching statistical significance by having so many repeated measurements.

So in terms of the statistical analysis, the analysis unit could be the original data. You examine the original data. You could analyze the transformed data, and we discussed you could use the ranks. We don't prefer to use the log or adding a constant to the log because of interpretation. And we talked about the pros and cons in terms of interpretation findings.

Now, in terms of the analysis method, if we are looking at the final assessment, i.e., week 12 or cycle 6, you could do a simple comparison between the active and the vehicle. You could do what the statisticians call an analysis of variance in which you could fit a model with the treatment centers and their interaction and look for the treatment effect. And you could do an analysis of covariance to include baseline as a covariate in the model. Remember change and percent change, we try to account for baseline severity in the model. What we are doing here in analysis of covariance, we are putting the baseline as a covariate in the model to account for that.

So we'll be comparing the efficacy results later for the two drugs which we have seen their plots.

The next bullet is about repeated measurement versus final assessment. When you talk about repeated measurement, as I have indicated, you might increase power for detecting a treatment effect. But the question was the number of time points to be included in the repeated measurement model. In terms of the statistical model or technique, we have multivariate analysis of variance. We have the generalized linear model or a mixed model. There is a battery of stat methodology which someone could use for the repeated measurement approach.

I'll be coming now to compare the efficacy results for the original data versus rank data for change, percent change, and I'll be taking a comparison also for the final assessment versus the repeated measurement.

Here this is for drug X which I want to remind you we did not see much activity going on for the inflammatory lesions, but we have seen something for non-inflammatory and total lesions. This table is for the counts and the way you analyze the final assessment. We'll be coming to analyze change and percent change.

I want to point out normally we don't compare this. We look for change and percent change, but I thought in terms of logical sequence, I'll present this quickly and

I'll move to the next one.

So this is week 12, which is the final assessment. Those two columns for inflammatory lesions, this column for the original data, and this is for the rank data. The next two columns for analysis of non-inflammatory lesions, which is again data and ranks. Here you have the total for the original data and ranks. We have the week 12 assessment here. You could see highly significant p values for total lesion count in the non-inflammatory.

I want to point out the delta which we are getting the highly significant p values. We are speaking about a delta of about 9 points roughly in non-inflammatory lesions, and about 12 lesions in terms of the total.

Now, that drug, we did not see separation in terms of inflammatory lesions, and you can see the difference is about 2 units. So it did not make it.

As you can see here, I have results for week 8 and week 4. They are not intended really to examine efficacy, but to make the point how do previous weeks, week 4 and week 8, impact the efficacy result of the repeated measurement. Again, you look here to the analysis of covariance. You have an almost significant p value here for inflammatory lesions because you are adjusting for the baseline covariate. And this is the multivariate analysis

of variance where we take repeated measurements, the last three values, generalized linear model, repeated measurement, and analysis of covariance. But you are diluting the treatment effect here because the previous measurements were not significant.

The reason I included them, if you analyze change or percent change, you start to see effect for the drug. So in that repeated measurement approach, I took week 4, week 8, and week 12.

I'll move to the next slide where we'll talk about analysis of change which normally we consider it secondary in addition to the percent change. So we'll be looking usually for percent change as well as change.

Again, you see here the result for change. Week 12, now inflammatory lesions make it when you analyze percent change. And you look here how much difference. We are talking about a 2.8 difference in terms of mean change, inflammatory lesions. Highly significant p values for non-inflammatory and total. I'd like to point out the non-inflammatory p value is close to those of the total, and the reason most of the total inflammatory lesion counts, they are coming from non-inflammatory. There is high correlation between them. So if you win on non-inflammatory, almost with certain probability you'll be winning in the total.

Again here we have the discussion. The analysis of covariance. The p value .03 which for a statistician is expected because week 12 -- when you analyze change, it's already you are accounting for baseline which is the same like analysis of covariance in which you take into account the baseline as a measure.

The multivariate analysis of variance which takes the repeated measurements has a bigger p value because you have the previous week, they are not significant.

So to summarize, highly significant p values for non-inflammatory and total lesions. And you can see really all what you need, as you indicated, is a small number of lesions between the active and the vehicle.

In this slide, we'll be looking at analysis of percent change, and this is the result for week 12. Again, it's highly significant, however you look at it, for inflammatory lesions, even though we have seen 2 lesions originally the difference. For non-inflammatory lesions, almost you make it however you look at it. You have a significant p value for analysis of covariance, multivariate analysis of variance. There's the repeated measurement. It starts to show close to the significant level here.

Now I'll move to drug Y. Before I go to drug

Y, let me just summarize the comments, which probably I listed most of them. The results for total lesion count are similar to those of non-inflammatory because of the strong correlation between non-inflammatory and total, most of the total coming from non-inflammatory lesions.

There is no general pattern for the p value for ranks versus the original data. I generally found the rank has a smaller but really there is no rule practically. It switched.

For inflammatory lesions percent change has a smaller p value than counts or their change.

For change and percent change, the analysis of covariance has similar results to week 12 analysis because in the two ways we are accounting for change from baseline.

The p values for repeated measurement in general are larger than those at the final study endpoint, and the reason for that, the results at the previous week, they were not significant.

Now here I'll be presenting the results for drug Y, and I want to remind you for this drug we have seen a small activity for inflammatory lesions. It's about less than 3 lesions roughly. And the drug shows separation early. So you expect the repeated measurement to result in a smaller p value compared to drug X where we did not see that separation early.

Now you can see here we analyze the count, which is the final assessment at cycle 6. The drug makes it for inflammatory lesions, even though the difference is like 2.8 lesions. But it does not make it for the non-inflammatory or the total lesions, which was opposite the drug X where we have seen the results coming from the total and non-inflammatory and we did not see much activity for inflammatory lesions. This is the intention to see drugs working differently by presenting two data sets.

In this study we looked at the results. We started to see some significant p values for change or percent change at cycle 4. So in the repeated measurement approach, we considered cycles 4, 5, and 6 to be included in the repeated measurement. Again, here you can see the analysis of covariance which takes into account the baseline. You have a significant p value. Once you take the baseline into account, you make the result also for non-inflammatory as well as for total lesion counts by just taking into account the baseline in the model.

For the multivariate analysis of variance, you see a .06 p value which is close to the significant level. The generalized linear model with repeated measurement, you have significant p values because you have observed a trend in non-inflammatory lesions, some separation early.

Again, as I indicated, this is the delta, which

generated those p values here at the bottom. You can see it. We are talking about 2.8, roughly about 6 non-inflammatory lesions, and about 8 to 9 total lesions. So this is the magnitude of the difference. The delta between the active would generate, as you will see, significant findings when you analyze change or percent change.

In this table, we analyze the change from baseline. And you can see now you have non-inflammatory lesions. They start to show significant results, as well as the total. And remember the delta was very small.

You analyze cycle 6. You have analysis of covariance. You make it and there an issue here. We have interaction, center-by-treatment interaction. So you have the analysis of covariance, significant p values, and the repeated measurement. You make it in the generalized linear model in which you have a treatment effect. The MANOVA will take into account other factors which could be time-by-treatment interaction.

On the next slide, I'll be talking about analysis of percent change from the baseline. Again, you can see the drug makes it for non-inflammatory and total lesions. However, things shifted for inflammatory lesions because of that high variability would generate larger standard deviation. At the bottom here, what I have is the mean percent change for those.

So, the results for the total and non-inflammatory lesions are, as in drug X, similar. But when you analyze the count, they are less significant because you have a small delta between the active and the vehicle.

Again, there is no general pattern for the p value when you analyze ranks versus the original data.

For inflammatory lesions, percent change has larger p values than the count or the change.

And for change and percent change the analysis of covariance gives similar results to cycle 6.

The p value for repeated measurement in general are smaller than the final assessment, and the reason for that, we have seen separation in the drug at an early period compared to drug X, between the vehicle, I mean, and the active.

Here we are looking at the efficacy results by baseline category. As I indicated, when discussing a phase III protocol with the sponsor, frequently a sponsor would like to enroll subjects with a smaller number of lesion counts to start with. So we tried to see if you include subjects with a smaller number of lesion counts, what impact does it have, if any, and the efficacy results.

So to address this issue, we divide the subjects according to their baseline category. We put them into groups. You could do any number of groups. Here I'm

going to consider four groups, i.e., quartiles. So I divide the subjects by the baseline category with almost an equal number of subjects in every group. I'll be comparing the efficacy results across baseline category. Of course, I'm not going to do formally statistical testing because you are reducing the sample size. The study is not done.

All that I'm going to do is look for the delta between the active and the vehicle in every group and see if there is some kind of a trend or pattern with the baseline category. I'll be doing this for inflammatory lesions, non-inflammatory lesions, total lesions, and I'll be looking also at investigator global assessment. This morning it came for people with a smaller number of lesions it might be easier to achieve success according to the investigator global evaluation. So we'll be addressing that.

Here I have a plot. This is week 12 lesion counts for drug X, which we discussed. We have seen this drug has very small p values, highly significant p values. What we see here at the bottom, this is people in category 1. We divide them inflammatory active, which is the dark one, and the inflammatory vehicle. Then we have the green one which is non-inflammatory for the active arm and the other one non-inflammatory for the vehicle. Then we have category 2 is the same thing. Category 3. So we break

down those people by the type of lesions they have. And this is the mean lesions again.

I'd like to bring the point here. You can see most of the difference among those categories coming from non-inflammatory lesions. You can see a number of inflammatory lesions across the four categories. There is an increase, but you can see there is much more difference in non-inflammatory lesions for category 4 versus category 1. So it sounds like most people who come with a high number of lesions at the baseline, mainly they are coming from non-inflammatory lesions. So this is for drug X.

I think we have another plot for drug Y, which is this efficacious. Again, you can see it here, the same phenomenon. You have people in group 1 which we have the smallest number of baseline lesion counts. Then people in the second category, they are classified. Again, you can see it's more pronounced here that the difference at baseline lesion count is coming mainly from non-inflammatory lesions.

In this table, we are comparing for drug X the delta, which is the difference between the vehicle and the active in each category to see if there is a trend across categories. In a way if it's easier to win if you have a smaller number of lesions at baseline, this will be reflected in the delta.

So first I'm taking the count. Those are the people in category 1. The first column is the active. The second column is the vehicle, and the third column is the difference. So people in the first category for inflammatory lesions have 13.3 at the final assessment for the active versus 13.2 for the vehicle, which gives you a delta of .1 if you are in category 1.

If you go to category 2, in the active you have 17 versus 20 in the vehicle. So there is a difference of minus 3 negative.

If you go to the active category 3, the difference is minus 3.5; the last one, minus 3.6.

So this is the magnitude of the delta. As you can see, we do not see a trend. You have in category 1 really .1, the other one minus. It's not much of a trend to speak about.

If you look to non-inflammatory lesions, again the same comparison. In category 1, you have a difference of minus 5.4; for category 2, minus 11. Then it goes back to minus 5.7, minus 25. So there is no pattern if you are looking to lesion counts.

If you look to the total, the same phenomenon. The difference, minus 5.3, minus 13.9, minus 9. So there is no clear pattern. Anyway the delta will increase as the baseline increases.

If you examine the change, again for the inflammatory in category 1, you have 1.9 versus 1.6 in the second category. So there is no linear trend or any type of trend in which you could examine -- you could see people with a small number of lesion counts at the baseline. They'll have a better chance of winning in terms of lesion counts.

If you analyze percent change, again you have for inflammatory the same phenomenon. So this is for drug X.

Let's see for drug Y. I'm sorry. What I'm doing here before I go to drug Y, I'm still examining the investigator global assessment to see the delta in terms of success across categories.

So for category 1, you have 35 percent of the subjects achieve success. I think the question in the morning was whether the drug achieved a clearance. We don't expect everyone in the active to achieve a clearance for the drug to win. All that you need to achieve is a significant difference. We see here the total overall for the active. For example, you have 18 percent versus 11 percent for the vehicle. So all that we are looking for, 7 percent, the delta. This is for the study overall to win in the investigator global. So we don't expect everyone in the active to achieve a clearance or almost a clearance.

So let me go back. So people in category 1 have the chance of achieving a clearance or almost a clearance. You have 35 percent which is higher than those in category 2, 21 percent, or category 3, 9 percent, or the other one.

But look what would happen. If you look to the vehicle and you are in the low category, you have also a higher success probability. You have 27 percent compared to people in the other arm.

So the point I want to make here is you would not look to the absolute number when talking about efficacy. We'll be looking at the delta, which is the difference between the active and the vehicle. This is really what's important. This is what drives the p values. So just to say that we'll achieve efficacy, we need to compare it to the vehicle.

So you take a higher chance of winning if you are in category 1, but this is again the same. So you end up with delta 8 percent if you are in category 1, 10 percent if you are in category 2. You have it reversed, minus 1 percent, 3 and 10 percent. And the overall difference is 7 percent. So again you don't see some kind of a trend in that probability to achieve success.

Next I'll go to drug Y which we have seen has lower efficacy than drug X. Again, we look at the results

by baseline category. We divided again into four groups, and the first part of table 1 for the count change and the last part for percent change. And I'll go quickly through it since it's the same discussion.

So you have the active, 9.6 versus 10. The difference, minus .6. In the second group, you have 2.3, minus 2.9. So really there is no general trend for inflammatory lesions.

If you take non-inflammatory lesions, it's the same phenomenon. The total is the same. There is no general trend there.

You look for change. You have .4, .4, 3.2, 5.8. Again, there is no clear pattern, if you are having a smaller number of lesions at baseline, that implies you'll have a better chance of winning in terms of lesion counts.

In the next one, I'm looking here to the investigator global evaluation and the success rate across the categories. You look for people in category 1. If you are in the active, you have a 65 percent chance to be in the win category compared to 49, 46 if you are in category 2 or 3. So here really the smaller the number of lesion counts at baseline, you have a higher chance of winning.

But again, it's the same phenomenon if you look to the vehicle. People who are not taking the active, if they are in category 1, they have a chance, 57 percent of

them, they end up in the win category. So you take the delta. You end up with 8 percent if you are in category 1. This is the delta between the active and the vehicle, and this is what we look for statistical testing.

You come to category 2, the delta, 9 percent, 20 percent, 8 percent, with an overall delta 10 percent.

I'd like to remind you for this drug, we have seen a small difference between the active and the vehicle. In particular, it was about less than 3 lesion counts for the inflammatory lesions, about 5 lesion counts for non-inflammatory, which translates to 8 or 9 lesions total. And we have a delta here of 10 percent for the investigator global, and the drug makes it in terms of statistical testing.

So a comment about the efficacy results by baseline category for the two drugs we considered, there is no general pattern for the results for lesion counts by type, their change, or percent change.

Similarly, for the two drugs, there is no general pattern for the investigator global evaluation.

For the range of lesion counts in these studies, efficacy results do not appear to vary by baseline severity.

And the following, I give general comments about the stat analysis overall.

Analysis of change from baseline or percent change and final counts with baseline as a covariate, all those approaches are an attempt to address or to take into account the baseline severity in the model.

Percent change data could have extreme outliers and could have heavy tail distribution when the baseline count is relatively small. We have see that by taking a plot for inflammatory lesions because I tried to make the point inflammatory lesions are the smallest of the three groups and we plot the data. So you end up with extreme outliers which have impact on the efficacy assessment.

A repeated measurements approach attempts to reduce the influence of outliers, the flares, by averaging over time, but the impact of repeated measurements on the p value depends on whether efficacy reached a plateau at the previous time points or not.

For the data sets we considered, treatment efficacy did not vary by baseline severity whether one considered analysis of lesion counts or the investigator global assessment.

I think this will end the first part of the stat presentation. I will stop here to take questions about this part. Then, as I said, the second part I think is exciting probably for statisticians, as well as clinicians. We'll investigate the relationship between a

global assessment and lesion count.

DR. STERN: I'll take the chair's prerogative and make a comment, which is really not very much statistical. From a clinical perspective, one reason that looking at multiple points is perhaps a pro and a con and could be counted in many ways is when I look at an agent for acne, what do patients want, they want consistency of effect and persistence of effect. So an agent that persistently removes 50 percent of lesions and keeps it that way may in some ways be more desirable than an agent that on two occasions reduces the lesion count by 80 percent but on another occasion, unpredictable, had no effect on the disease. I think you have to consider the clinical aspects of repeated measures and if in fact, in addition to reducing variance because of measurement error, something has to be put into our equation that from my clinical perspective that agents that are less persistent and consistent in their effect are, in fact, less clinically desirable than agents you know what they do and they keep on doing it.

Would you like to comment on that?

DR. ALOSH: Yes. I'm in complete agreement. I think the point which needs to be made, you could achieve statistical significance, as you pointed out correctly, by taking repeated measurements and averaging them and

reducing the standard deviation. But a clinical judgment needs to be made whether that significant p value is clinically meaningful or not.

So this will bring the design issue -- I mean, like in this trial we have assessment at weeks 4, 8, and 12. If we are going with the repeated measurements approach, how many repeated measurements are you going to take. We don't want to go too far by taking several repeated measurements, reduce the standard deviation, and get significant p values. We need to maintain, I think as Dr. Stern pointed out, whether the results are clinically meaningful or not.

DR. BERGFELD: I'm going to speak as a non-statistician, but when you displayed all this information regarding the activity of the vehicle, it brought to mind that perhaps there needed to be a third arm here of petrolatum because the vehicles are chosen not only to suspend the active, but because they offer some efficacy in themselves and patient acceptance. So we expect the vehicle to be active in some way. But you would have a greater delta if you use it against petrolatum.

DR. ALOSH: Well, I think it was proposed in the morning whether it's ethical to have people on the vehicle or not I thought. From the data set which we have, I think it showed efficacy. The vehicle itself, as you

pointed out correctly, has a large impact on the efficacy and the delta.

DR. KATZ: I'd hate for the positive effect of vehicle to enter the vernacular as being vehicle efficacy. That's an assumption. Vehicle positive effect could be investigator bias. In fact, the original reason for controlled studies was not because we had such a fantastic number of efficacious vehicles but the reason is to help us measure investigator bias which is -- I don't mean any pejorative sense, but it's something that exists. So just because there's a positive effect of vehicle, we shouldn't use that as vehicle efficacy.

DR. TAN: Yes. Dr. Alosh, you presented a lot of information here. I'm trying to digest it.

I think the percent change under the changing total lesions, they reflect two different aspects of the measurement of the clinical efficacy.

What does percent change mean? The patient's condition improved over the pretreatment condition. Right? So that could be anything. What you're talking about, those abnormalities you observed is natural by the definition of percent change. This is just relative to the patient's previous condition. So, therefore, you do need to the absolute change. That's the original data. So you need both aspects.

I think the statistical significance here is not -- I mean, this is not relevant because you have a designed study and in the protocol you should specify specifically what kind of change you're looking for. This would have to come to agreement from the clinical point of view, what kind of change, 10 percent change, is relevant or not. So this will be determined before you even start the trial.

DR. ALOSH: Well, a couple of points. As a statistician, I would not prefer percent change personally. And for the same reason which you have seen, you have extreme outliers, et cetera. I would agree with you in terms of interpretation. If you have someone who started with 10 lesions, a reduction of 5 lesions would be translated to 50 percent compared with another one who started with 200 lesions. I think it's a measure which to me a clinician prefers.

We do look for percent change as well as change, by the way. So we analyze both of them jointly, having said that.

In terms of the magnitude of the difference, I think in terms of a clinical trial, we came across several trials. I gave two examples of what is the range to achieve statistical significance. I think Dr. Wilkin could speak to that. With that range, it seems clinically it's

acceptable.

Now, concerning the point of it needs to be prespecified or not, definitely we have communication with the sponsor at phase II and phase III trials, and we agree on what endpoint needs to be analyzed, in particular percent change, and we'll be looking for change in addition to the investigator global assessment.

So I share the concern you have about analysis of percent change, but really, we look at it with other factors. Percent change would reflect what happened to the patient over time, whereas investigator global -- this is the co-primary endpoint. You are looking at the final assessment, the assessment at final study endpoint. So it's a co-primary. It's not the whole story behind winning because you still need to win to achieve clearance or almost clearance.

DR. STERN: But if you come to those two charts you showed of drug and placebo, the scatter diagrams, my interpretation of those results -- one interpretation would be we have an active agent that prevents people with a little bit of acne from flaring substantially, and otherwise the effects seem about the same. And the question gets to be, if all of the essentially significance comes from a difference in a few people on vehicle who started out with not much disease flaring, is that really

an effective agent for acne?

DR. ALOSH: Yes, I think this is a good point. As a matter of fact, the plot which I presented was for inflammatory lesions only. And that drug in particular what we have seen at week 12, there is a difference only of about 2.8, if I remember the number of lesions. So the drug with that scatter, in a way it showed you the drug controlled the flare because you have more scatter data in the vehicle arm compared to those on the active arm. So the drug has activity in reducing that variation. But when you come to analyze final lesion count, it did not make it.

But I think the point here, we have the baseline as the other measure. We need to take into account the baseline score. In the plot we tried to show the baseline by week 12 assessment. When we took the baseline as a covariate in the model, you make it whether you analyze the change or you analyze the final count and you take the baseline into account, which is what we call the analysis of covariance.

DR. STERN: Dr. Kilpatrick.

DR. TAN: Just one.

DR. STERN: Sorry. Dr. Tan.

DR. TAN: Does that mean your baseline analysis -- you have several slides showing that. Does that just confirm that you do need a randomized study because there

is no pattern in terms of the response? You have four categories there. Right?

DR. ALOSH: Right.

DR. TAN: But if you do randomization, you have a sufficient number of patients in the two groups. That should not make any difference.

DR. ALOSH: Well, let me clarify in case it wasn't clear. You have a randomized trial at the baseline. So, of course, people at the baseline you expect to be distributed randomly in every category. We are looking to the efficacy result at week 12 by baseline category. So anyway, if I divide the people according to the baseline severity, do people who have a lower number of lesion counts at the baseline achieve higher probability of success if you look to lesion count or the investigator global compared if they have -- let's take an example.

If I started with a subject with a 50 lesion count, what's the efficacy result for that subject compared to someone at enrollment that has a 200 lesion count? So you need to compare what's the delta for those people in the lower category of the baseline compared to the delta -- what I mean by delta is the active minus the vehicle -- at the high category.

The point here is if you have high efficacy results for people with a smaller number of lesion counts

at baseline, you might be better off to win if you enroll subjects with a smaller number of lesions. We are looking at here is most of the difference coming from non-inflammatory lesions, from those plots which we have seen, and the delta is similar. If you look to lesion counts, change, or percent change, we looked again to the investigator global, what we have seen in the investigator global, the people in the lower category have a higher probability of success, but the same thing holds for the vehicle. So you end up with a delta roughly the same.

Does that answer the question?

DR. TAN: Yes.

DR. KILPATRICK: Thank you.

I can get into this in a roundabout way or follow my own personality and be more direct. I've looked ahead, Dr. Alosh, into your next section in which I notice -- and again, I presume in this one, when you talked about IGE, the percent of success, you used a logistic regression, logistic regression I presume because the proportions are not normally distributed. Counts are not normally distributed. So my question is, are some of these phenomena that you're talking about explicable by the fact that you use a normal distribution in your analyses rather than the Poisson distribution?

DR. ALOSH: Dr. Kilpatrick, I think going to

the second presentation, which I'll come through it in some detail, what I'm modeling in the second part of the presentation --

DR. KILPATRICK: No, sir. I'm really asking about the modeling of counts when you say you're going to be use an ANOVA, a MANOVA, et cetera. Why not use log linear regression?

DR. ALOSH: Okay. This is another point. I think when you talk about logistic regression, logistic regression came in the second part. But I agree with you. If you are going to analyze counts which has a Poisson distribution, the number is small.

Yes, indeed, I use the normal approximation. We are talking about a trial with about 400 subjects. So if you take 400 subjects with number of lesions not small, we have seen the normal approximation for the data works.

But I agree fully with you. If I have a small number of lesions with a small number of patients, as you pointed out correctly, I'll use the Poisson regulation. But that type of lesion, as you know, the normal theory would work for that.

DR. KILPATRICK: This may be my only opportunity to say this in front of other statisticians from FDA. I don't see why we should continue to use the normal distribution when it is not appropriate, when there

are other models that we can use. I have really little feel for how much of what we've seen today is due to the non-normal distribution or how much of it is due to the true differences between small and large.

As regards the baseline, I agree with Ming that if it's randomized, you shouldn't have to use it, but then if you do use it, I agree with you that you should put it in the right-hand side as a covariate rather than dividing which assumes linearity, et cetera.

DR. ALOSH: Well, definitely it's a good comment. Personally I think I'll go back -- the normal approximation. I'll not say really the analysis here is not appropriate because you could take -- I mean, it's a technical point. I'll be happy to discuss it with you. As you know, n times lambda where lambda is the mean of the Poisson distribution, 10 to something, it will go to the normal.

It's a technical point. I don't expect personally the p value which I'll get from fitting a log linear model to be different than that. But definitely I could investigate it. We could discuss it. It's a technical point. There are other statisticians who might give their opinion as well.

DR. STERN: I actually think, though, it's more than a technical point. It's a bit of a conceptual point

going to this whole issue of how much does baseline status affect what happens with the data subsequently, if I understood you correctly, and your feeling about what is, in fact -- is this distribution of changes a normal one or not and how it's related.

DR. KILPATRICK: Well, I reiterate. I think both my feeling -- I'm perhaps more of an idealist than members of the FDA. Since I've taught these methods to my students, some of whom are now employed by CDER, I know they have the techniques. Why don't they use them? But I agree with Dr. Alosh that it may be unconventional, but it's certainly modern statistics.

The logistic model is much easier to explain than the log linear model, but I'm concerned not so much with p values as with error distributions and predicted values. Predicted values may be quite different under the normal assumption and the log linear.

Thank you.

DR. TEN HAVE: Yes. I'm the third statistician here. I guess I should probably make a comment.

But getting back to this issue of the normal distribution, there's another related issue and that's this variability issue which has come up a number of times in today's conversation, in addition to your consistency comment. And I have a couple of questions.

One is, you mentioned the difference in variability between the active arm and the vehicle arm, and that also has consequences obviously for your test statistics. And that's related to whether they're normally distributed or Poisson distributed or whatever.

But there's a second issue which is probably more difficult to consider and that's should variability itself be a measure of efficacy. You're looking at differences in mean scores or mean counts. Should you be considering differences in variability whether one is more consistently better than the other across patients but also across time within patients?

DR. ALOSH: That's definitely a good argument. I think in the morning we had an example in which a drug was approved for the indication and the other drug not approved. What we look for is collective evidence. We look for consistency of finding across centers. So, for example, one might get an application which barely makes it. We could go back. We don't take just this p value. We look for consistency across centers what you see. Definitely at one point in time, we were looking at the final assessment. We are going back here to look at the repeated measurement approach whether we see some kind of a plateau reached, whether there's a consistent finding or not.

I would agree both with you and Dr. Kilpatrick. There are many assumptions underlying the statistical test which I presented here about the generalized linear model or repeated measurement. What's the type of the H matrix you need, et cetera. So there is a lot behind those p values which are reported here.

But I want to make the point, definitely we look for consistency across centers. If there are outliers, we'll go and investigate back.

In the second presentation, I'll be fitting a model and I'll discuss exactly how far we go to see if there is an outlier and how we dealt with that.

DR. TAN: Yes. I just want to add just one point to the log linear model here. I noticed on your slides, you already mentioned the generalized linear model. I think nowadays all those models are falling into this generalized linear model. That includes the log linear model. And it's readily available. I agree with --

DR. KILPATRICK: I think the term "general linear model" --

DR. TAN: Generalized linear model.

DR. KILPATRICK: But to me generalized linear model involves the Nelder -- I call it Nelder generalization of the general linear model. Dr. Kligman, are you with us, sir? Okay.

DR. TAN: Yes. That would include what is called a log linear model into that.

But actually I have another question. You said for the repeated measures analysis -- I actually have a different view from what we talked about this morning. You talked about the inhibition of the new, emerging lesions. We all agreed in the morning that it is important to see the consistent improvement throughout the course of this treatment. There are certain defined periods. So, therefore, the success really should be defined as not just at one shot. It should be at maybe 8 weeks and 12 weeks. So instead of you increase your power, you actually have less power. You need to have more patients in this way. You should have the improvement both at 8 weeks -- maybe not 8 weeks -- maybe 6 weeks. At two points maybe.

Actually in the cancer research area, people have been using this because patients who have cancer respond to a new therapy and then come back again. So people now redefine responses. The tumor has to be shrunk by 50 percent at two time points. And this would capture that emerging new lesions.

DR. STERN: That was exactly the point I was trying to make, that rather than combine, it's probably more appropriate to do multiple, independent testing, and you've got to pass both tests as opposed to combining the

data to reduce variance across them so you can pass one test more easily.

DR. TAN: Yes.

DR. STERN: Dr. King.

DR. KING: Actually I have a lot of trouble with this, the concept of the washout and the whole area. I think clinically they say, quick use the drug because it quits working soon enough. So if you're going to start off with the baseline, should not all the patients start with a washout period that would stabilize it and then you actually measure the consistency or persistency of effect? Because the fact that you may be better at one time point that's being stressed here is, like cancer, you may have a recurrence. It seems to me that you not only have to start with everybody having a washout period, but you need multiple measurements at the end. Where like two points make a line or three points make an even better line, it seems to me that just having one point is going to lead to an erroneous result.

So would you comment on the washout period and then the multiple points showing a persistent effect? Because that's really what patients are after, persistent effect.

DR. ALOSH: Yes. Thank you for the question.

I don't think really I'm in the position to

comment on the washout. I think it's clinical. But I think the point here, if we are seeing people could experience a flare during the course of the trial, whether they take the active or not, we expect even if we observe people before enrollment in the trial, they could experience this flare as well due to some factors. As a clinician, probably you know it more.

So, consequently as a washout period -- do we put people on a certain drug and we are looking for improvement or just examine them? I don't understand much about the nature of the disease, whether we could control things here in terms of washout.

I think in terms of the repeated measurements, indeed it's a good point, because that flare, that high variability should come having outliers. We'd like to get rid of them, reach to a more reliable measure by taking probably two or three repeated measurements instead of one.

Now, the question we are addressing here, how many measurements you are going to take and we need to maintain a clinical relevance not only to reach a statistical significance.

DR. KING: My point is quite simple. With the washout period, it's been my experience and probably others' that once people start getting bad, it doesn't matter which therapy you give them. They just keep on

getting worse. You start off with a small bump and it just keeps going. The purpose of the washout period is to try to pick up those who are going to become outliers. As you showed, the outliers can really affect the outcome, and so you'd like to have a period where they end up truly being stable because when you start off with saying you can't have medicine or any other therapy for about 4 weeks, some of the delayed effects are such that once you stop their polypharmacy or the multiple drugs, somewhere around week 6 after stopping that, they start off getting a lot worse. So it seems to me you have to control for the outliers, and then you average the last three or four visits.

DR. ALOSH: Thank you.

DR. STERN: Thank you. This will be the final comment in this section.

DR. PLOTT: I have a question regarding the analysis of covariance. Are you in this analysis taking into account the different numbers, inflammatory and non-inflammatory lesions, and how that impacts the total lesion count? Because consistently in clinical trials, we've found about a quarter of the total lesions are inflammatory, maybe two-thirds or something like that, three-quarters are the non-inflammatory lesions, and in the analysis of covariance, how is that taken into account?

DR. ALOSH: This is a good point. In terms of

the analysis of covariance, if we're analyzing the total lesion count, what I'm taking into account in the model would be total lesion count at the baseline. And if I am putting a model for inflammatory lesions, I'll be putting in the analysis of covariance inflammatory lesions at baseline. So whatever the model I'm using there, whatever the final assessment I'm modeling, I'll put the corresponding value at the baseline.

I think you could ask the question, when I did the efficacy assessment by baseline category, there you could break it by number of inflammatory lesions at baseline or non-inflammatory lesions or total lesions. For the data I presented here, I break it down by total lesions. I felt this is more representative.

You could do the analysis for any one of them, but when we presented the data, most of the difference is really coming from non-inflammatory. There is a little bit of change in inflammatory lesions from one category to the next, but most of the difference between the different categories is in terms of non-inflammatory lesions.

DR. STERN: Thank you. I think we need to move on to the remainder of Dr. Alosh's presentation, and there will be questions after that as well.

DR. ALOSH: The second part of presentation -- I think we heard this morning a lot of discussion whether

investigator global evaluation is more rigorous, whether it's needed in addition to lesion count. And we have seen also a discussion on the other side that probably we should do only with lesion count without the investigator global assessment.

Most of the work come here really -- we don't do it in analyzing clinical trial data, but we get questions from the sponsor in many cases that they would like to power the study for change or percent change, but they found it more demanding to power the study for the success criteria according to the investigator global evaluation.

So in this presentation, I'm going to talk about assessing the relationship between the success on the investigator global evaluation and the acne lesion count. In the morning, Dr. Wilkin presented data in which you have some artist draw lesions. Here I'm going to take actual data distinguished between inflammatory and non-inflammatory lesions, fit the model, and see whether the investigator global evaluation expressed as a success is more rigorous for efficacy evaluation than analysis of change or percent change.

The outline of this part of the presentation. I'll be giving some background, why is this needed. I'll be modeling the investigator global evaluation, the success

criteria. We are reducing this to success/failure, even though we start with a 6-point scale or a 5-scale. I'll give an interpretation and assessment of the fit, and I'll conclude with some final comments.

Just to go back a little bit, we talked about the measure for efficacy evaluation in acne trials consists of two parts. In the first part, we are talking about a lesion count based measure. What I mean by that, change or percent change. And the other co-primary endpoint is the investigator global evaluation which is ordinal data on a 5- or 6-point scale.

Now, lesion counts is based on counting the data. It's more rigorous probably. The second one is based on visual evaluation or visual assessment. But we need to keep in mind we have the same subject. We are doing the efficacy on the same subject whether we are counting lesions or we are giving a score to the subject.

Then also the same investigator doing the assessment, one time counting the lesion count and then the second time giving a score.

For those two reasons, we expect the two measures, whether lesion count and investigator global assessment, to be related to each other.

The goal here is to investigate the relationship between the dichotomized investigator global

evaluation and the lesion count.

Specifically I'll be using empirical modeling to address the following issues. Was the impact of lesion count or their change on success according to the investigator global evaluation?

The second question I'm going to consider, whether a certain type of lesion has more impact on the investigator global evaluation success. We talked about inflammatory as well as non-inflammatory lesions, and I'd like to see whether one type of lesion has more impact than the other.

And then I'll be talking whether there is utility of adding the baseline count to the model.

What we have here, I'm going to use logistic regression model to model that relationship. The reason we use that, what we term it as a binary data which we express it as a success or failure. The p here represents the probability of success. So I'll be modeling the odds of success or failure. I'll be taking the log of that which is what's known as logistic regression. This is what we call a dependent variable. And I'm taking this as a function of the covariate here. The beta is what we call a set of parameters of the model. And the X's could be lesion counts by type, inflammatory, non-inflammatory, or whatever, but also to call it independent variables.

Now, the interpretation of the parameters of the model. For example, if you take what's the meaning of beta 1, if you increase X1 by one unit, beta 1 will give you the magnitude of change and the log odds of success on the investigator global assessment as a result of increasing X1 by 1 unit.

Now, as I said, X1 could be number of inflammatory lesions or non-inflammatory lesions or baseline. So this is just the generic form of the model, and when I'm going to the actual modeling, I will replace X1 by a certain type of lesion.

The data set I'm considering for this analysis is what I presented in the previous presentation, which is drug X. I have 400 subjects. The study was, as you have seen, for a 12-week duration with assessment done at weeks 4, 8, 12. We have the investigator global evaluation done on a 6-point scale from 0 to 6 where 0 means clear or no lesions to 5 which is very severe.

Now, success here is defined as to be in category 0 or 1. And 1 says "minimal," but there's a definition of what's meant by minimal. A certain number of inflammatory lesions and non-inflammatory. A clinician will judge that. Now, in the investigator global assessment, the success criteria is defined, as I said, as 0 or 1.

In this model, I'm taking the final lesion count. I'm modeling this. This is X1 and X2 which are the inflammatory lesions and non-inflammatory lesions at week 12. What we have here, as I said, is the probability of success according to the investigator global evaluation.

I would like to point out in this study I excluded one outlier from the model. And the reason for that, I fit the model in the beginning and I got barely the model make it in terms of interpreting the data. Going back, I found an extreme outlier. I looked to that outlier. One subject that was assessed as a success was given a score of 1, and this subject had 17 inflammatory lesions and 41 non-inflammatory lesions. This does not fit with the criteria of 1. I mean, that subject would not be defined as a success. So I ended up taking that subject from the study and refit the model because that subject definitely should not be classified a success.

Now, in terms of interpreting the parameters of the model here, what I want to point out, this is the beta and what we call the intercept. This is the coefficient for inflammatory lesions and this is non-inflammatory lesions. I would like to point out the coefficient for inflammatory lesions is about four times in terms of magnitude as non-inflammatory lesions. So inflammatory lesions have much more impact on the success criteria

compared to that of non-inflammatory lesions.

The second point I want to make is those coefficients are negative. So as the number of inflammatory lesions increases, your chance of winning decreases. And you could say it differently. As the number of lesions decreases, you have a higher chance or higher probability to achieve success.

The interpretation of the parameters. We could say a 1 unit increase in inflammatory lesions at week 12 would imply a decrease of e to the power minus 41 or .662 in the odds for success according to the investigator global evaluation.

The same thing for non-inflammatory lesions. A 1 unit increase in non-inflammatory lesions at week 12 implies a decrease in the odds for success. As I said, you could put it differently. You could say what's the impact of a 1 unit reduction in inflammatory lesions at week 12, how much it has an impact to increase your chance of winning.

I want to go back to this slide. What we see here is only the final lesion count, inflammatory and non-inflammatory, in the model. We don't have the baseline lesion count in the model. I think this is very logical. If you have the final assessment, you could judge whether the patient is clear or almost clear. You don't need to

know the baseline because you have the final count.

The coefficient of inflammatory lesions, as you have seen, is about four times that of non-inflammatory lesions. This might be due to appearance, color, size, or the surrounding halo of erythema of inflammatory lesions. When the final lesion counts are given, as we said, baseline values provide no additional information for explaining the investigator global success.

Here we fit the model, but we'd like to see how good the model fit the data. For a good model, we could predict the probability of success according to the investigator global evaluation from the number of lesions at week 12. A good model will give you the predicted value from the model similar to the observed successes in real life.

Now, this statistical test here, which is the Hosmer-Lemeshow test, breaks down the number of subjects in the trial presumably into 10 categories, but we have 8 here because you don't need to have a smaller number of categories. If there is a smaller number of categories, you need to lump them with the other categories. But here we have only 8 categories.

In every category, as you said, you calculate the probability of success and you could calculate the number of successes and compare it with the actual number

of successes in that category. Now, those categories are based on the predicted probability of success.

Of course, you could make the correct classification in every category. You might have in one category 20 people a success and 10 failures. You could classify 21 a success and 9 failures. The total will be the same, but once you make an error in one of them, it will be reflected to the next category.

So, for example, if we go to group 1 here, we have a total of 135 subjects. The observed number of successes in this group is 0. From the predicted model we got .01. Of course, we don't expect to get an integer value from the model, but the observed are going to be integers.

Now, if you take this, it means if we have observed success as 0, the observed number of failures is going to be 135, and you could see the expected from the model is 134.99. And the sum of those two should give you the total 135. The same here.

You go through this. You come to the second category. You compare it. You could see the observed successes is very close to the expected. And we come in terms of goodness of fit statistic. We give the chi-squared test .95 with 6 degrees of freedom, which gives us a p value of .98, indicating a very good fit for the data.

On the previous slide, we modeled the final lesion count. What I'm going to consider here is a model for change from baseline, because this is what we analyze in an actual clinical trial.

The same model we have here. On the left-hand side, we have the probability of success according to the investigator global evaluation divided by the failure, and we take the log. On the right-hand side, this is the intercept. X1 is change in inflammatory lesions; X2, the change in non-inflammatory lesions. Now we have two terms added in the model which are X3 and X4, and those are the baseline covariates. So X3 is the baseline for inflammatory lesions and X4 is the baseline for non-inflammatory lesions.

I want to point out in fitting the model, I used what's called the step-wise approach. You fit the simple model in the beginning and you include covariates in the model if they could explain some additional variation from the model. So the addition of those covariates to the model in the beginning, the intercept would find X1 which is change in inflammatory lesions more important. So we'll enter this one. Then non-inflammatory change will explain additional variation. So the model will take that. But still the baseline could explain the variation in the model.

Again, the point I want to make here is you could see the coefficient for inflammatory lesions, .412. It's still about four times of that of the change in non-inflammatory lesions. The same holds if you are looking at the baseline lesion count. You could see the coefficient for inflammatory lesion count at baseline, .43 compared to .089 for non-inflammatory. So again we could see the inflammatory lesion coefficient is about four times. It's a more important covariate than non-inflammatory lesions, probably for the same reason we discussed. It could be the color of inflammatory lesions, more red. It could be the halo of erythema, just different factors.

Again in this analysis, I'm excluding the one subject which showed success even though this subject has 17 inflammatory lesions and 41 non-inflammatory lesions. I'm excluding that subject from the analysis.

On the next slide I show the comment. Change in inflammatory lesions do not fully explain the investigator global evaluation. This is in contrast to the previous model. When I modeled final lesion count, I did not need the baseline lesion count. All you need is the final assessment. But here when you are talking about change, it's not sufficient to tell me that I have a reduction of 50 lesions. I would not know from where you started. So baseline is still an important covariate in

the model to explain that variability.

So we have seen larger reductions in inflammatory and non-inflammatory lesions increase the odds of investigator global success. So the more reduction you have in inflammatory or non-inflammatory, you have a higher probability of winning.

On the other hand, increases in baseline inflammatory or non-inflammatory lesions reduce the odds of investigator global assessment. So if you start with a higher baseline, you have a lower chance.

Inflammatory lesion again has about four times the impact as non-inflammatory lesion on the investigator global success.

Here again the same discussion about assessing the goodness of fit or using the Hosmer-Lemeshow test statistic in which by calculating the predicted probability of success for every subject we divide the subjects in the trial into groups, and here it's 8. In every category or in every group, you could see the number of successes observed and those expected from the model. Definitely the closer the two to each other, the better the fit is.

In terms of calculating the chi-squared goodness of fit, we have chi-squared of .83 with 6 degrees of freedom, giving again a very good fit for the data.

So to summarize, if you have final lesion

count, you don't need baseline assessment to tell success in the investigator global, but if you have the change, you need the baseline. So the success according to the investigator global assessment is more rigorous criteria for success than analyzing change in lesion count. I think this will bring the question now we understand why industry would like to power for change but not require more patients for the trial to power it for success according to the investigator global assessment.

I think the discussion came also this morning whether one should do an analysis of count without the investigator global or vice versa. The discussion came on two sides. We see really here is they're in a way complementary to each other. I see change in lesion count. You are looking to the time trajectory what happened over the course of the trial, whereas the investigator global assessment will give you the shot at one time point, what happened to that patient, whether he's clear or almost clear.

The final comments. Inflammatory lesions have more impact on the investigator global evaluation success than non-inflammatory lesions. Absolute change in lesion counts alone do not fully explain variability in the investigator global success because baseline is still an important covariate in the model. The fitted model is

useful for checking consistency of a study finding based on the investigator global.

And I'd like just to remind you about that outlier. Without fitting the model, we wouldn't be in a position to see that there's some observation. The data is not consistent in that observation.

I'll stop here. If there are further questions, I'll be happy to answer them.

DR. STERN: Dr. Kilpatrick.

DR. KILPATRICK: Thank you, Dr. Alosh. I want to congratulate you on introducing goodness of fit. That's the first time I've heard that in an FDA presentation. That is not a joke, sir.

I wanted to ask at what level would you consider the goodness of fit test failed. What p value would you use? This is something that really has to be discussed I think and put up because would you use the 5 percent? Are you going to be as stringent? And then again, the ramifications, as you well know, of how much leeway will you or the sponsor have in bringing in subjects or throwing out subjects, et cetera. There's a whole feeling there.

DR. ALOSH: Well, thank you first about the comment of goodness of fit. I'd like to point out indeed we do a lot of statistical methodology. We read papers.

We do extensive work in the background. Although I think for the purpose of a presentation such as this, we tend not to bring -- because, as you know, the background. So we'd like to communicate just the main findings.

The second point is addressing how good is good, the way I see it. It's a matter of judgment. You could see data. You get a p value, for example, for goodness of fit, 20 percent. At .2 we could say it's acceptable.

In this case, when I found I'm getting a small p value, I ran SAS, examine influence, and I find just extreme in terms of the percent chi-squared. One observation has 16.something. So with that, I said it cannot be. There's something wrong here. So I go back, examine the data, and just one subject has 17 inflammatory lesions and 41 non-inflammatory lesions, and this subject was classified a success. So I think both you and me and probably most of the audience here will agree that this subject should not be classified as a success in the first place. Now, you take that subject out, and practically we do a sensitivity analysis to see how much improvement in the fit. And by taking that subject, my p value went from .05 to .98.

I think this will give you an indication that really you are looking for consistency in findings. I

think the model itself has a good check on the data. If we go analyze the number of successes without looking deep, as I pointed out, and consistency across centers, looking for outliers -- and this I think brings why we do rank analysis because the point we made about outliers, we look to the data in different ways to reach to collective evidence about approval. So really there is a lot of work done behind the scene before we arrive at the final comments in our report.

DR. STERN: I may be completely off base here, but I've never seen a model fit so well, and I wonder whether it's appropriate to do it this way or one should have randomized half the data set and bootstrapped it and see how well it fitted on the other set. Maybe it's just me, but this is an extraordinary fit for a model of this kind in my very limited experience. And I'd ask the experts about that. I've never had any data I've worked with produce a model with this kind of fit.

DR. TAN: I just want to mention here that the purpose of doing this analysis was to see the probability of success based on the global assessment, how that success is related to other factors. I think that's legitimate just to use the whole data.

If you want to do a prediction, now in the future I'm just going to use this total lesion to predict

the global success score. Then you may need to validate the model and use the bootstrapping.

DR. STERN: Is this an unusually good predictive model?

DR. TAN: Not entirely. I have seen data fitted this well, yes.

DR. ALOSH: Well, let me give you my reply since I fitted the model, at least.

How good the model, I think it depends on how close the two variables are to each other. Now if you take into account -- as I said in the beginning, you have the same subject, the same investigator, one time doing the counting, counting lesion counts, and then the second time seeing if we have either success or failure. So if you are doing it, I would not expect you to give a patient a 50 lesion count and to classify him as a success.

On the other hand, if I'm doing, let us say, getting data on different phenomena in real life, especially epidemiological data or social science data, we reached a p value of .4. So I'm in full agreement, but I think we need to keep in mind here the theory behind it, the same investigator doing the two evaluations. And unless there's some error, I don't think you will be -- and you are dealing with intelligent people, I mean, with dermatologists. So it's not like someone who might do

something on the side or someone not educated. So for that type of data, I think it's reasonable.

I'm going to take your point and fit it to another data set because this is for drug X which we have seen a high efficacy result. This also will play a role in that data.

DR. KILPATRICK: May I ask a follow-up question? May I take it then that you did do -- did you do goodness of fit in the count data also? Were you looking at how well this model fitted in the earlier presentation?

DR. ALOSH: The earlier presentation, yes. We fit analysis of variance, generalized linear model, and the p value was very small -- I'm sorry.

DR. KILPATRICK: I'm asking about the goodness of fit. Did you test the model in the analysis of variance, MANOVA, et cetera?

DR. ALOSH: You look to that, what's the proportion of variance explained by the model. And that proportion is small compared to what we have here. You might end up to have a significant treatment effect, but how much variability in the model is explained.

DR. TEN HAVE: You mentioned that the companies are saying that they have a hard time powering their studies for the IGE as opposed to the lesion count outcomes.

DR. ALOSH: That's right.

DR. TEN HAVE: I was just wondering in your experience has it usually been that the lesion counts are where the statistically significant differences occur between the active treatments and the vehicles and it's not such the case in the IGE outcome based analysis?

DR. ALOSH: That's right. As a matter of fact, since we analyze the change, if you look to the second model in which you have the investigator global assessment as a dependent variable and we have change in inflammatory lesions and non-inflammatory as the independent variable, they did not explain the variability in the model. So you still need the baseline to interpret --

DR. TEN HAVE: Right, but I'm just thinking in general terms across studies. When the pharmaceutical companies submit their analyses and you look at the results based on the lesion counts using, say, analysis of covariance where you do adjust for baseline versus whatever analysis they use, logistic regression or Fisher's exact test, or a chi-squared test for the investigator evaluation, where do you usually see the treatment differences occurring? In both?

Is there consistency usually or is there usually significance for lesion counts but not the IGE outcome? Just in general terms. Is it harder to get

significance with the IGE than it is with the lesion counts?

DR. ALOSH: We see a result -- consistency in general. You will observe results, for example, in total lesions probably in one type of lesion, and you'll see it in the investigator global. But it's harder in the investigator global compared to the analysis of change or percent change from baseline. So analysis of success according to the investigator is more rigorous. I mean, you need really more number of patients to achieve it compared to analysis of change or percent change.

DR. STERN: I think we'll have to stop now, and for the remainder of the afternoon, I'm going to become much more stern with presenters and keep them to their time. I think if everyone would like to take literally a 5-minute break for those who need to, and then we're going to start in 5 minutes with the first presentation and go on through in a sterner manner.

(Recess.)

DR. STERN: For the next 15 minutes, Dr. Markham Luke is going to talk to us about combination topical products for the treatment of acne vulgaris.

DR. LUKE: Thank you, Chairman Stern, members of the committee, Dr. Wilkin, Dr. Bull. I'm going to address the combination topical products for the treatment

of acne vulgaris. I am not going to be speaking about adjunctive therapy or about co-packaging issues that you had raised. Those are issues for a different time.

The Code of Federal Regulations has in it a passage by which the agency addresses fixed combination drugs. Notice the term "fixed" combination. So there's a set ratio. These are drugs that have two actives mixed together. "Two or more drugs may be combined in a single dosage form when each component makes a contribution to the claimed effects and the dosage of each component (amount, frequency, and duration) is such that the combination is safe and effective for a significant patient population requiring such concurrent therapy as defined in the labeling for the drug." And I cite 21 C.F.R. 300.50(a).

For the situation of acne combination drugs, the combination topical products for the treatment of acne vulgaris require evidence for the contribution of each active component or components that are purported to provide for added efficacy.

To clarify a little bit more, in applying the combination drug policy for two drugs, component substances A and B having the same endpoint, in a three- or four-arm clinical trial, success is demonstrated by A plus B, the combination drug product, being better than either of the monads, A or B, and both of these monads being better than

the placebo.

For the acne combination drugs, we have currently marketed combination topical drug products that have the combined topical antibiotic either erythromycin or clindamycin -- and for our purposes they can equal A -- with benzoyl peroxide, which I have put on the slide as equaling B. The safety and efficacy of other combinations for the treatment of acne are also currently being investigated.

Studies to address the combination policy for acne drugs have shown that the most difficult superiority to demonstrate is the contribution of the antibiotic, or A, to the efficacy already achievable with benzoyl peroxide alone, or B. And so demonstrating A plus B better than B is something that needs to be strived for.

In conclusion, each component of a fixed combination drug for the treatment of acne must demonstrate a contribution to the claimed effects of the drug product. This may be difficult if the contribution of one of the actives, for example, the topical antibiotic, is minimal and hard to discern when combined with another active, for example, benzoyl peroxide.

DR. STERN: Thank you.

We'll now have our next talk by Dr. Porres who will talk labeling for efficacy, and then there will be

questions for both at the same time. So Dr. Luke can come back up.

DR. PORRES: Hi. I'm Joseph Porres, medical officer, Division of Dermatologic and Dental Drug Products.

This will be a very brief presentation on what is usually included in the clinical studies section of the labeling for products approved for the indication acne.

As has been touched upon before, efficacy is measured by looking at endpoints such as acne lesion counts and the investigator global evaluation. So I won't delve into this in any greater detail.

In this section of labeling, the clinical studies section, we include a description of the types of studies that led to approval, the phase III pivotal studies, describing what kind of studies they were, how long they lasted, the number of patients who received the drug treatment or who received the placebo, if it was an oral medication, or the vehicle, if it was a topical medication, the mean age at enrollment for each one of the two arms, and whether a statistically significant difference was observed and for which endpoints. Also, we include information about the types of patients which were included or excluded in the studies. It may be important for the clinician to know whether maybe patients who had severe acne were not included or whether pregnant women

were excluded or perhaps whether certain age groups were not included in the studies.

In this slide I'm going to show an example of the kind of text that we include in labeling to denote the information that I just referred to. Here we have a paragraph describing that product P was evaluated for acne vulgaris in two randomized, double-blind, placebo-controlled, multicenter phase III studies which lasted for six cycles of 28 days each.

Here we have another sentence indicating that there were 295 patients who received the active while there were 296 who received placebo, and the mean age at enrollment in both arms was about the same, 24 years old. The study lasted six cycles, and at the end of the studies, in both of them a statistically significant difference was observed between the drug product and the placebo both for mean change from baseline in lesion counts, which we will show later in a table and a figure, and also for the investigator global evaluation.

We also noticed that in this particular set of studies, patients who were deemed to have severe androgen excess were excluded from the design.

Now, we also used, besides text, tables and figures. That way we convey different types of information, trying to facilitate to the clinician to have

a bird's eye view or a glimpse of what the data from the pivotal studies showed.

Here we have an example of a table and there are several pieces of information. First of all, we tell that this is a study done for acne. Normally we evaluate each study separately and there must be a win on both to win approval, but here for the sake of simplicity, I'm presenting to you the pooled data.

So there were two studies, P1 and P2, and both of them lasted six cycles. We showed the types of lesions that were studied, inflammatory, non-inflammatory, and total, and for each one of them we showed what the baseline mean count was and the count at the end of the six months or cycles.

We also show in these columns the actual counts for both the active and the placebo. For instance, for inflammatory lesions, we started with 29 lesions for the active arm, and we ended up with 14, which translates in a 52 percent reduction in lesions. However, for the placebo, we started with 29 and ended up with 17, so that means a 41 percent reduction in the counts. Here we have similar numbers for non-inflammatory and for the total.

On the last column we show the treatment effect which is the difference between what was observed with the drug product and the placebo. And as you can see, in this

case for inflammatory lesions a difference of barely 3 plus/minus 2 lesions was enough to reach approval. I'd like to stress this because sometimes I hear that people have the impression that it's very hard to approve things at FDA, and as you can see, a difference of just 3 lesions can sometimes make it statistically.

Again, for non-inflammatory, the difference was a little larger, 5 plus/minus 3.5, and for total lesions, 7 plus/minus 5.

Now, sometimes there are differences in between the two arms, the active and the placebo arm in which case we may want to add a sentence or a paragraph denoting the differences. For instance, in this particular case, drug product users who started with about 74 acne lesions had about 42 after 6 months of treatment. The placebo users started with about 72 and ended up with 49 lesions after the same duration of treatment.

Now figures can also help to provide important information at a glance especially because you can get a time relationship of the effect. Now, again, in this case we're just showing a graph for the mean total lesion count where we use against cycles what happened to the mean percent reduction. And this slide is the one for placebo, and this one is for the active.

Although we apply statistics only to the

prespecified evaluation time, in this case 6 months, I'd like to show you that in this case some differences were noticeable even at the second cycle. However, they don't reach statistical significance until cycle 6.

In summary, presenting information as text, tables, and figures offers prescribers a comprehensive summary of the efficacy data observed in phase III trials. The three formats complement each other since each one is helpful in conveying a particular aspect of the data.

Thank you.

DR. STERN: Thank you very much. This section is now open for questions. Dr. Katz.

DR. KATZ: Dr. Porres, I assume that was two topical trials. Is that correct?

DR. PORRES: No. The information that was conveyed here was for an oral medication.

DR. KATZ: Did you list the difference in side effects between the placebo and the oral medication? Was there a significant difference there? You didn't show it.

DR. PORRES: Yes. We didn't show that here because we wanted to concentrate on the efficacy aspect, but of course that information is reflected and it's in the package insert and it's in the labeling of the drug product. It is there. So it's not like we didn't look at it.

DR. KATZ: My point is that many -- many -- double-blind studies -- that's used as some godlike quality, double-blind studies, and it gets repeated in the literature that they were double-blind studies -- start as a double-blind study, but they don't end up as a double-blind study, and nobody ever mentions that, not in the first study and then not in any literature that follows, especially with topical medications. So a double-blind study that shows perhaps an 11 percent advantage to the drug, but if you look at the side effects, 70 percent of the patients in the drug -- I won't mention drugs, recent topical drugs for acne -- 70 percent have irritation versus 10 percent with the vehicle.

Well, somebody should mention that those did not end up being double-blind. They were controlled, but the blind was broken and nobody mentions that. That's why even with an oral medication it's important to know is there a significant difference in the side effects because that breaks the blind. I think that's very important. And that's not mentioned in any studies in any of these borderline effective drugs that come out.

DR. PORRES: The point is well taken. In fact, that information is collected at the time of approval, and it may even have a bearing as to whether or not the drug is approved if the side effect profile turns out to be

horrendous. But that information is collected and it goes into labeling, and most of it is probably reflected in the PDR.

DR. KATZ: No. But my point is that it's not that the side effects might be horrendous. The side effects might be very minimal. After all, when we treat patients in the office, a very high percentage have some dryness with, let's say, topical retinoids. That's an acceptable side effect. But it does bias the investigator. It breaks the blind in the study.

DR. PORRES: Well, oftentimes in these studies, the blind is actually not broken until the end if the side effect is not severe enough to break the study or to interrupt the continuation of such patients within the study. You may not actually find out whether the adverse effect was related to drug or to the vehicle until the study is completed.

DR. KATZ: But the investigator would be biased.

DR. STERN: Right. I think Dr. Katz is bringing up a point that's always a problem with products that have irritancy, which is unblinding of the investigator. In fact, there was a huge discussion about this with retinoids and the treatment of photo-aging where the effects were perhaps even more subtle than they are in

the treatment of acne. That's always a methodologic problem. Are you really unbiased and blinded as you go on? And how does the agency deal with that, Dr. Wilkin?

DR. WILKIN: Well, after Dr. Katz' comment, we'll be thinking about it just a little bit differently in the future because I think the question that he's asking is should we not craft into the clinical studies section of labeling, where we're talking about outcomes, whether there actually was such a difference in local adverse events as to disclose which was the active and the inactive arm. Dr. Porres is correct. One can move further into the package insert and find that information in the adverse reaction section of labeling, but I think the point that Dr. Katz is making is should we not also put that contextual piece in right there where we're talking about the efficacy.

DR. STERN: It's not either a formal inclusion or exclusion criteria, but it's some other parameter that lets you look at these data and say what are possible things that make them either more or less believable given the limitations. Is that the point you were trying to make?

DR. KATZ: That's correct.

DR. STERN: Let me ask a question. You've shown here an oral agent versus placebo, and Dr. Luke talked about combination agents. When you present data for

a newly approved combination agent, do you then present A plus B versus A versus B versus placebo to show the differences in efficacy versus all of your choices so in one summary you can say this is how much I gain or this is how they played out within this trial?

DR. LUKE: In general, combination studies have multiple arms, and you would have an arm with A plus B in new vehicle and A and B arms in the same vehicle, and then the vehicle arm.

DR. STERN: I understood that in terms of the trial, but I didn't know whether you would report that in the manner that Dr. Porres had where you'd give the results of all four arms.

DR. LUKE: Not all the arms are reported in labeling in the past.

DR. STERN: And which ones are generally reported?

DR. LUKE: Actually I'd like to ask the committee here. Do you think it would be helpful for us to put all of the arms in labeling?

DR. STERN: I certainly think it's extremely useful, if it's a combination agent, to compare it against the single agent, as well placebo, that came closest because really what you're asking is if I give this combination agent, how much better am I doing than either

of the alternatives. Now, the reason not to give all four is it's kind of confusing, but I'd want to know that what you implied, that if BP has results almost comparable, how much did the combination beat BP by?

DR. LUKE: I can see your point and I also see your point regarding the labeling can be very cumbersome if you were to put a lot of data in there and it would confuse the issue. I think we've addressed that in some labeling by indicating in writing, rather than in the table itself, that one of the arms may be less efficacious or they haven't proven efficacy for that arm.

DR. STERN: And I guess if it was a combination against an established therapy, as a clinical decision maker, although I want a placebo arm in the trial, the four arms you described, I guess my own opinion would be what would be most useful for me as a clinical decision maker is how much better is it than either agent alone and having BP were the stronger agent with the single agent that did better comparing the combination versus BP would be the most meaningful in terms of clinical decision making, not either placebo or not versus --

DR. LUKE: That may be difficult to discern from the data from a given study because keep in mind that the monads are in the same vehicle as the combination A plus B. And therefore, with the new vehicle, you throw in

a different twist to the product. They're not the approved benzoyl peroxide alone product that's on the market. This would be a monad with the new vehicle that is being studied that has been developed for the combination, and that vehicle often, one would think, would help enhance the stability or do something to improve the efficacy of the combination.

DR. PLOTT: I'd like to ask from your presentation are you suggesting that combination drugs could be studied with one of the ingredients that is thought to be the most difficult to show superiority? And jumping off the last question, maybe that most difficult product would be an approved product versus the product in its vehicle.

DR. LUKE: I'm not suggesting that. I think we are governed to some extent by the rules. The Code of Federal Regulations does state that we have to demonstrate a contribution of each of the actives in the combined product. So comparing it to an active in another vehicle probably would not provide any regulatory utility for a 505(b)(1) application.

DR. KING: I guess I have a conceptual problem. I thought the purpose of having combination drugs was to make it for convenience. That is, it seems to me the appropriate trial would have been if you're taking drug A

in the morning and drug B in the night, which is how most dermatologist prescribe things, the purpose of having combination drugs is assuming that the nighttime and the morning are efficacious in synergy, that the combination drug would provide just convenience. So I guess I'm lost here.

DR. LUKE: Dr. King, that's a very good issue. I think the concept that you're visualizing is a combined product or a co-packaged product perhaps where you have --

DR. KING: I'm just saying what the standard practice is now. You give one in the morning and one at night. And why you put them together is you noticed there's a synergism between A plus B in the morning and night, and giving the combination one time a day, in this fast-paced world, is likely to get done by the kids as they run to the school or classes.

DR. LUKE: Right. I think the regulation addresses the fixed combination drug. You are combining two actives in one product. What you're saying is when you take one product in the morning and one product in the evening and the two products are given together, you're either co-prescribing, which is the practice of medicine, or if a drug company wants to market the two together, that's co-packaging. And that's a different issue.

DR. WILKIN: Coming back to the point of which

arm is the most rigorous, there's nothing in the CFR or any of the stat guidance documents that says that all of the arms have to be equal-sized in the studies. So I would say that's one of the take-home messages. If you know one particular comparison that is the most difficult, you may want to increase those arms to get more information.

The second part, which is Dr. King's comment on let's say you have product A that you take in the morning, product B that you take in the evening. One is an antibiotic. One is benzoyl peroxide because that's the sort of standard sort of thing. If those are products that are already on the market, even if they have the active in the same concentration, they're going to have different vehicles than the vehicle in the combination product to be marketed. One of the things that we've found over the years is there is an enormous difference in performances of products when you change the vehicle, even if you keep the active constant. It becomes one of the hurdles to getting generics approved if they're topical semi-solids because it's not the same thing as the -- I think Dr. Leyden is gone, but he talked about how simple it is for the solutions and wished it might be that for the semi-solids. But there are multiple phasic structures. They can affect the stratum corneum, some of the inactive ingredients. So to interpret 300.50 in the CFR in the combination products,

it really needs to be in the same vehicle. So that's what makes it different from just comparing two products that are already on the market.

DR. STERN: Thank you. Our next speaker will be Dr. Lehmann from Johns Hopkins and he will be speaking on his methodologic review of acne therapy.

DR. LEHMANN: Good afternoon. Thank you very much for this opportunity. I'm very honored to be speaking here. I'll be speaking about the work that we did over a couple of years for the Agency for Health Care Research and Quality. The full report is two volumes, and I brought a number of spare copies in a box near the slide projector if anybody would like a free copy. My mother has enough copies. She'd be happy to share them with you.

That was the joke.

(Laughter.)

DR. LEHMANN: So the Agency for Health Care Research and Quality has kind of a mission to document evidence for controversial or concerning clinical issues. They get nominations for different topics every year, and one year both the Academy of Pediatrics and the American Academy of Dermatology nominated acne therapy as a question that needed a synthesis of evidence. So we put that together.

So the process of the Education Policy

Committee is to recruit technical experts. In fact, a number of dermatologists, including Dr. Shalita, were involved in reviewing what we did, although we take all the blame for any of our results.

We identify the patient population, formulate, refine specific questions, perform a comprehensive literature search.

Also, before this point, besides recruiting technical experts, we also recruited a kind of committee of people who would be interested. We went to the pharmaceutical industry, to a number of the lobbying organizations and research organizations who declined involvement, but we did get involvement by a number of professional societies, such as ACOG and others.

So perform a comprehensive literature search, summarize the state of the literature, construct evidence tables, and submit a report for peer review.

So the objective was to evaluate types and quality of evidence available to support decision making, clinical decision making, after what Dr. Stern was just talking about, in the treatment of acne vulgaris. So we're taking a little bit of a step back from the approval process and saying now the medication is approved, what should or what do clinicians do with them.

So our perspective was that of the practicing

generalist. I'm sorry that the dermatologists aren't here to argue, but I think it's clear that generalists have to take care of acne at some level. We were hoping to find out what the evidence basis was for the phase at which you refer for a dermatologist to take care of acne. So these are the type of generalists we had in mind.

Now let me go through this diagram a little bit. This is a causal diagram. The idea is what is the nature of clinical decision making. What should the nature of clinical decision making be, and then can you define the type of evidence that you would need to support that model, that decision making.

So for instance, all -- I'll say kids, but all patients are assumed to have some level of self-care. And so one immediate question is what do we know about the patient's care of their own acne. They may come into the physician, and at that point the physician makes an assignment, knows what the baseline characteristics of the patient are, not so much for determining the efficacy of the treatment, but in terms of actually making a decision of what needs to be done. So at the point of making the decision, which is in the box, they've made an assessment of the baseline characteristics. They've made an assessment of what the acne is like, and they've made some assessment about how likely this patient is to comply with

therapy. And then they prescribe therapy, and then the patient comes back. And then if the patient "fails" therapy, then something else is done.

We were hoping that at some point we could see, again, as I say, that one of the things to do is to refer to dermatologists.

At each point along the way, in talking to clinicians and thinking about this, we figured there were at least four major axes or major dimensions that weigh on a clinician's mind. What will be the result of the acne long term? What will be the patient's current quality of life? What is the cost, and what are other morbidities, depression and so forth?

So ideally we would like to see data that says given certain baseline characteristics, what do patients do? Given baseline characteristics, what should be prescribed? Given certain prescriptions, what are the long-term results? What's the quality or life? What's the cost and what's the morbidity? So that would be the ideal literature on acne.

We searched through the Cochrane Collaboration, their hand-assembled database of randomized clinical trials, the Medline, OldMedline, PsycInfo, the nursing literature, and reference lists from key articles.

By the way, we did not include the European

literature and this became important, for instance, in isotretinoin where some of the best work was done in Germany, but we didn't have enough money basically to pay for translations.

In the review process, all abstracts were screened by two independent reviewers. All the articles were read. They were read serially by two or more abstracters and then me and one other senior methodologist. And then, as I said, we tried to include dermatologists on the reading staff and other reviewers.

Articles that were excluded were those that did not address the management of acne, so articles talking about resistance to medication were not included, evidence that was not directly on humans, articles that addressed non-acne vulgaris, review articles or letters to the editor, and again as I said not in English.

We started out with about 4,800 citations. We ended up with 237 controlled trials. I should say we ended up with 275 studies which were 298 trials because some articles contained more than one study within the article, and then we had to exclude some. So we ended up with 237 controlled trials.

Just to give you a sense of over time, going back to 1951 -- I think those were Dr. Kligman's articles -- and a lot of the people you saw here today and then a

lot of the work done in the '80s and the '90s.

So just in terms of the results of our review, if you had the ideal literature, you'd be able to know how generalizable the results were. The studies should have been performed well. The treatments should be well defined. A small set of comparisons so you know what to say, a consistent set of outcomes, stratified outcomes, and free of commercial influence. I don't think I have to tell you what the punch line is, but we're going to go piece by piece.

So in terms of geography, it is worldwide, continental, United Kingdom, USA, Asia, Middle East, Oceania, and Africa. Obviously, most of it is in the Anglo-Saxon world.

Enrollment. There were only 42 studies that actually used word "recruited," otherwise it really was unclear how patients got into the study. Now, recall clinicians want to say given certain baseline characteristics, given a certain history of therapy, what's the next best thing to do. If the literature doesn't record these data, then the working clinician has no idea really when to use a certain therapy at what point in time.

In terms of comparability of the arms, most of the arms were comparable. Only four studies had arms that were clearly not comparable.

Study quality. Dr. Kligman I think referred to this study before in his talk. It's a little bit hard to see in this graph. There was no clear-cut assessment tool for saying whether you have a bad study or a good study. We used a very qualitative judgment. If the paper said it was double-blinded and it said how it was randomized, we said that that was a good thing. If they told you nothing about who the patients were in the study, we said that was a bad thing. We simply said a study could be good, good and nothing, good and bad, or nothing/nothing. So just looking at studies that had only high quality elements or only had low quality elements, you can see that they're mixed throughout time. Unfortunately, quality does not go up in time.

In terms of treatment administration, 90 were systemic. The rest were topical. Just in reference to the question that came up several hours ago about whether treatments that are effective in other parts of the body, this represents the total amount of controlled trial data that we have in the published literature to answer that question.

These are the therapies. I think these are all without repeats. Vitamin A and vitamin A palmitate. So in terms of a small set of therapies, we were kind of in the hole here, and these are about 150 different treatments,

including tea tree oil and some other therapies, as well as FDA approvable medications.

In terms of characteristics, these are the number of trials simply providing data. Tanner stage is referred to as pubertal stage. It was mentioned before as being a very key element. No studies referred to pubertal stage. Age, 74 percent of the studies reported on the age of the patients; 73 on the sex of the patients; 8 percent on race; 2 percent on the skin type, and decreasing there.

In terms of where the patients were being cared for, about 80 percent -- you could tell whether this was a generalist study or a dermatologist study.

Again, in terms of the clinician or people trying to figure out whether or not the study applies to their patients, this is an unfortunate state of affairs.

Let me go through what this graph is showing, which is a little complicated. We divided the therapies in terms of classes of therapy, which is not radical, anti-androgens, antibacterial, combinations, antibacterial and other keratolytics and retinoids. So this is the comparator arm and this is the target arm. So this study represents an antibacterial versus an anti-androgen study. It's a little bit hard to see on this because it's cut off.

So this gives you a map of what the comparisons are that have been done.

The size of the box gives you a sense of the sample size. So you can see, for instance, the whole keratolytic and others are relatively small sample size studies, whereas the antibacterials have a fair number of large. And the anti-keratolytics should include -- the retinoids are up here, if I'm not mistaken. And then the little star indicates high quality. So you can see these tend to be high quality. These tend to not too have much high quality. Irritation is mild, moderate, severe. We'll say a little bit more about that in a minute. But this gives you a quick map of the entire world of acne therapy.

This recaps what we've been talking about basically all day. These are some of the scales that were used in the studies, and we basically said, okay, we're going to call all these mild in our synthesis. These are all moderates and these are all severes. This is the 6-plus stage that Dr. Leyden was talking about.

This simply points out how many studies used different types of outcomes. Most of the measures are in terms of either overall change, physician change, either in terms of the patient or the physician, the integrated global assessment that we've been talking about, or then counts, percent change, delta percent, and delta counts and so forth.

If we are concerned about outcomes other than

just counts, we would imagine that there should be a study or two that actually assesses quality of life. This is not in your handout. A couple of slides I put together sitting in the back during the session to show you the power of PowerPoint. There was one study I think that had a quality of life scale separate from the overall assessment. This is in distinction to a number of clinical trials in other areas where either SF-36's or other quality of life scales are used.

In terms of stratified outcomes, when we started the review, a lot of the dermatologists said it's really important to stratify patients. You can't say anything helpful unless you know whether a patient is mild, moderate, or severe or categorized by age and sex. Only eight studies stratified their results sections by these factors that were deemed to be really crucial in terms of evaluating efficacy of treatment.

In terms of funding, Dr. Kligman mentioned this before. There were seven NIH-funded studies. Eight were miscellaneous and 100 were drug sponsored. Of those, 12 were first author, 38 with a co-author. 13 provided the funding to the authors, and 35 simply provided medication or analytic support. And then the rest were basically, I suppose, hobbyists.

So that says something about the state of the

literature and the problems, and it's easy to say that there were a lot of problems with the literature.

One analysis that I did after we published the report was to say if in fact the literature is a mess, then in fact there should be inconsistencies in the literature. We just heard that if you're going to look at combination therapies, if the combination is better than A or B and better than nothing, that's a good thing. If you have treatment A is better than B, and treatment B is better than C, and then treatment C is better than A, you have an inconsistency in the literature. So there was a question. Can we find an inconsistency?

To do that, we said let's divide our studies up by the studies that seem to be mostly mild patients, mild, moderate, and severe. So let's see what the results were.

So the iconography here is that an open arrow or an open bar means level B evidence, that is, only one clinical trial that had good data. A dark bar or arrow meant two or more clinical trials that gave pretty good evidence. So, for instance, we can be pretty certain here that doxycycline is better than placebo. Thank goodness.

So here's a little island that salicylate and vitamin A are better than placebo. Doxycycline is better than placebo, but doxycycline seems to be as good as fusidic acid from the literature. Here tetracycline

topical seems to be as good as tetracycline oral from the literature. I understand that dermatologists could say that this is not true, but I just want to say this is just straight from the literature. So here we have a nice little island, another island here, and here a little island compared with benzoyl peroxide.

Mild and moderate. Now we have two smaller islands, a little bit more certain data. This is weird that tetracycline was as good as placebo in the moderates, but that's what the data seem to say. These are separate studies. We have clindamycin, erythromycin, isotretinoin, and tretinoin. This is the combination.

Moderate. Basically no solid evidence but we have this notion that these guys are above these guys.

And moderate to severe, again a bit more complicated.

And in severe, not much that we have to say. These were two different doses of isotretinoin.

The only thing I can say is although people might argue about the specifics of the comparisons, I was surprised to see that there were no inconsistencies in the literature, which suggests that way we divided mild, moderate, and severe made sense and may have some clinical import.

Then this is where we couldn't assign a

severity from this paper and this is just a mess.

Now, one point I did want to make is that while we were doing this review, we were thinking does it make sense to use something other than placebo as the control, and this little analysis that shows that these islands are not inconsistent suggest that maybe placebo could be used as an anchor point in the treatment of most of these different severities of acne without biasing the results. In other words, I think there's some evidence from this review that benzoyl peroxide could be at least that active arm if you're not going to use a placebo.

Since we were talking about placebos, I drew this out of the database while we were sitting. These are studies divided by mild, moderate, severe, just the placebo arms of the studies, just looking at their percent change. And I apologize for percent change. This is 0 percent. A minus is good; positive is bad. So here you see in the mild it's kind of mixed. Mild/moderate, it's still mixed in terms of placebo response. The studies that reported placebo, they were almost evenly divided, and as you get towards surprisingly even some of the severes, the placebo still did pretty well.

This is just at 12 weeks. We did record 4 weeks, 6 weeks, and 12 weeks or whatever data we could get our hands on.

So in summary, it's difficult to generalize from the studies because the studies don't say who is in them. The studies were mixed, performed well. In terms of a well-defined set of treatments, it's difficult to say, and the bottom line is that clinicians are not left with a clear road map on how to treat acne even given approval. So too many comparisons, an inconsistent set of outcomes. The outcomes are not stratified, and it's not clear how much the commercial influence is.

So there's a limited basis for comparison of acne treatment from the controlled trials, even though we have to do it. Using available comparisons does not lead to internal contradiction. So that's a good thing.

On the other hand, only industry-sponsored research is available to help clinicians make clinical decisions, which means as a clinician, my thinking is as you ponder what outcome measures to use in sponsor studies -- I don't know how much you're allowed to say, but since no other studies are going to be done because, as we heard, there's no research in this outside of getting these drugs approved -- clinicians desperately need usable outcomes to help them make clinical decisions.

I'll stop with that.

DR. STERN: Thank you. May I start with a question?

It seems to me you are in part implying -- if you look at publication, there's both publication bias in all the ways we know, and in fact what is going to be published is written by people who are either employed by or under the sponsorship of industry trying to put forward their argument in a way to advance a product. It seems to me that some of what we're hearing today is we may have an opportunity to have data presented in a way that is neutral or judged by the same third party, that is, the FDA, across all products.

We know that some authors are much more successful at getting data -- the inference is presented in one way than others are, even with the same data set, or at least that's my experience.

So perhaps one of the lessons here is one can't rely on the current kind of data that is published in the somewhat variable peer-reviewed literature and that what we need is some objective uniform set of referees. I guess that goes to one specific question.

Did you look at the quality of papers -- and I recall that you did according to where they were published -- and what the impact factor was?

DR. LEHMANN: We did not do it in terms of impact factor. At the time we discussed this, we thought we would have to subjectively rate the journals, and we

were not ready to subject ourselves to that level of abuse.

(Laughter.)

DR. LEHMANN: But impact factor is an excellent thought. Thank you.

DR. STERN: Other questions.

DR. WILKIN: There's another source also. I don't know if you explored FOIA, the Freedom of Information Office. You can obtain the reviews on products that have been approved, and then you can go on and compare those reviews with how it's portrayed in the literature. It's not that the data are changed, but often the emphasis is somewhat different.

DR. LEHMANN: That's an excellent suggestion. I don't know if the HRQ talks about that tactic with their EPCs. That should be a tool that we use in our systematic reviews and we just don't. I suspect one reason is that we have a narrow time frame, and that's a lot of effort. But it's an excellent suggestion.

DR. KILPATRICK: Thank you.

I think Dr. Lehmann's presentation has brought us firmly back to Dr. Katz' point. I mean, that may be obvious, but maybe we should come back to that tomorrow and see how we can try to eliminate that type of bias that he was describing.

DR. PLOTT: Just a comment. I think many

investigators take a lot of pride in the work that they do in the unbiased evaluations. It may be unfair to suggest that an industry-sponsored trial has that bias. While I admit that there are many that undergo a lot of data dredging, probably the substantial trials that you see published in the literature have gone through the FDA reviews, not just by the Dermatology Division, but also by the advertising group, and are quite thoroughly scrutinized. So while I acknowledge that there is bias, there's probably an equal number of substantial articles that have been reviewed.

DR. STERN: Having at times been industry sponsored, I would certainly hope that some of the published research was good. But in fact in my local medical journal in the last week or so was a series of a sounding board, an editorial, and a paper that looked at the difficulties in maintaining objectivity and in fact putting out results in academia when you're under the sponsorship of industry. As I say, that's in the last two or three weeks in the New England Journal. So I think there are issues and it doesn't mean that everyone is good or everyone is bad, but there are certainly issues that seem to be out there in this area.

DR. TAN: Yes. I just wanted to ask Dr. Lehmann, for the industry-sponsored trials, how many of

them are investigator initiated? How many are, do you know, just for NDA purposes?

DR. LEHMANN: All the information we had was in the article, and it said at the bottom "sponsored by" or whatever.

DR. TAN: Yes, I would actually differentiate that if the investigator initiated this trial and then find a sponsor versus the trials that the industry want to do an NDA for. There is a crucial difference I think.

DR. LEHMANN: And that distinction is not made in the literature.

I do want to stress that my stress about industry versus non-industry is not so much bias as much as once the drug is approved, there's no energy, funding or otherwise, to evaluate the effectiveness in practice of these medications. So the approval process is the only shot the clinicians get to see what works, and that's a different perspective than FDA has, I understand.

DR. KING: I guess in terms of what the committee is deliberating, what suggestion are you making to this group or to the FDA that would have the highest impact on providing the information and high quality studies? It's your forum.

DR. LEHMANN: Thank you. So, first of all, the work that you're doing here is terrific, and just saying

maybe we need to have one outcome measure, that would be terrific because then you can start measuring across studies.

Number two is it sounds like from both the dermatologists and Dr. Alosh's presentations to have at least two measures, one the global and -- let me backtrack.

An acne outcome is a multi-axial, a multi-dimensional outcome. There's what the skin looks like. There's the lesions. There's how the person feels. It's multi-dimensional.

The drug companies and the FDA are kind of being forced into a situation where they have to take a multi-dimensional problem and squash it down to one dimension. That's always a problem.

Now, there are a number of ways of doing that. Most of them are subjective, utility measures and stuff like that. At the minimum, you can have a measure that is two or more dimensions, the global assessment, some sort of lesion counting. I don't know if you want to throw in a quality of life measure to give some sense of what's going on. On the outcome measure side, those would seem to be the recommendations.

On the incoming side, more explicit mention of who is in the studies, who the patients are in the studies in terms of where they've been before they got into the

trial, what their age, sex, and race breakdown is. Pubertal status I'm not ready to say at this point. But some notions that when I see the study, I have a lot more to say. As a clinician, I have more to make my decisions on.

Now, it's interesting that there's a project called Trial Bank going on from UCSF where details of trials that are really specific can be stored separately from what the output of what the article is, which means that a reader can actually see more details of the trial, not necessarily the raw data but more details than the space of an article allows for. So a project like that that uses these new informatics tools, in addition to new statistical tools, might be a way to go.

DR. PORRES: Sometimes we see drugs that come for approval and don't make it and yet we see publications coming from academia or some groups where the drug appears to be wonderful. I'm wondering if you have a suggestion as to how to obtain this kind of data so that you could analyze it.

DR. LEHMANN: You mean the data on the stuff that's not submitted to you.

DR. PORRES: Well, we cannot divulge information about the drugs that don't make it, unfortunately. That wouldn't go into the Freedom of

Information aspect of it. But you would need the kind of information or you would need to be able to assess whether the results that are being published by a certain group match the results, say, for the drug that we approved that in our hands seemed to be barely making it, and yet when you look at group X, they claim the drug is super wonderful.

DR. LEHMANN: I can only report on what I see to one degree, number one. Number two, that's one of the reasons why we made that map of the islands of care. I don't know if it really will work.

DR. STERN: Other questions.

(No response.)

DR. STERN: Thank you very much, Dr. Lehmann.

We have about 50 minutes for committee discussion in general, and I think what might be useful is to use the questions we've been presented with and rather than trying to answer any of them now, since we have at least some of the resources, in terms of guests -- I hope Dr. Lehmann doesn't leave -- try to think about any other points that we may have heard some information or want clarification to answer these questions so we can think where we are going forward. Does that seem like a reasonable way to proceed for the remaining time?

So the first question is -- again, this is not

to answer the question but further information. Should the current success criteria using the co-primary endpoints be retained? I guess I would say the idea of co-primary endpoints as opposed to necessarily the current two that we have or the current multiple ones that we have.

DR. KILPATRICK: Since there may be some experts here, I'd like to hear more about incidence. This came newly to me. The concept of identifying new comedones and pustules, et cetera and following them is rather different from this counting facility that I've heard and even the IGE. But how would you effect that is the problem. Is it feasible is what I'm asking.

DR. STERN: Well, I think that's probably not feasible short of frequent visits and computer mapping. I think what you're doing is you know that certainly in an 8-week time frame that with the exception of large nodular lesions, a single comedonal or inflammatory lesions, most will have resolved spontaneously, certainly inflammatory lesions. So what you're doing is comparing prevalence to time points and you're assuming that if there are fewer prevalent lesions at the latter time point that the incidence in those 8 weeks was lower or particularly the incidence in the couple, 3 weeks before that was lower than it was in the 2 or 3 weeks before your entry to the study. I think those are the assumptions.

But when I brought up the concept of incidence, I wanted to make it clear that -- which is a common misconception among patients. A lot of patients think that when you put them on a drug, you're clearing the pimples that are on their face on the day they start the drug. Rather, what you're hoping to do is reduce the incidence over time so that the prevalence, because of self-healing, will be lower sometime in the future.

DR. PLOTT: Dr. Stern, if I may, just to address this question. The difficulties I think were echoed in some of the presentations today, some of the clinical and statistical presentations, and maybe more clearly by the statistical presentation, that doing lesion counts where inflammatory lesions are at a minority in the total number of lesions that are being considered, a product that is acting solely on inflammatory lesions is biased against in that situation where they're only able to affect a small number, a minority of the total lesion count. And a win in that count requires winning both in inflammatory lesions and totals. So a product that just purely affects inflammatory lesions is biased against.

On the other hand, with a global evaluation, we've heard that a change in inflammatory lesions has four times the impact in global than a non-inflammatory lesion. Here the inflammatory lesion has the advantage. A drug

that's hitting just inflammatory lesions is at great advantage.

So you could see the difficulties in putting these two together being as co-primaries and why there is some frustration in requiring that we win in all of these.

Now, the resolution for that may be to allow a product that is only effective at inflammatory lesions to have simply an inflammatory lesion claim and handle that problem in labeling as opposed to a product that's not able to hit this great goal of having an indication for acne vulgaris as a whole.

DR. STERN: Since you speak for industry, I guess my question would be does a company, on the basis of phase II studies -- if we're going to have such a thing, would you be willing to say, and we will tell you in advance whether this product is for inflammatory acne and judge it according to the inflammatory lesion count and the global count? We won't use the comedones unless they're worse and they count against efficacy. We won't use the comedones or the total count before you do the phase III study. Because again, it's the whole problem of anytime you go back and you dredge through the data, you can figure out a way of cutting it and make small differences significant and sometimes even chance significant if you're a very good statistician or a poor one as the case may be.

(Laughter.)

DR. PLOTT: Of course, every firm must make a decision for themselves, but I could imagine a product that was purely effective at an inflammatory mechanism and how you would not expect to have effect in a comedone. And doing drug development in the proper way, you might find in a phase II trial where there was really no efficacy against comedones and that you had a dose response and you picked the appropriate dose. And moving into phase III trials, I think that there could be a situation where a product had just anti-inflammatory activity. You've heard of possibly some of them here today.

DR. STERN: Why don't we go on to clarifications for the second question, which is really the point that Dr. Plott brought up. How should lesion counts be analyzed?

I guess here I would like to put forward one question for the agency. Some of what we've heard from the experts is one way to reduce variance is to, in fact, use modern measurement techniques that rely on types of photography that are more standardized that also allow you to look at people truly side by side over the course of their treatment rather than trying to remember how they were, use observers who were not involved in the care who were perhaps less likely to bias. And in fact, with

digital imagery, one can even take out the background irritation and just concentrate on the lesions. When you see a patient, you know whether they're kind of rough and pink. With digitalization, there are probably ways of taking out the roughness and pinkness and just leaving the blackheads, whiteheads, and inflammatory lesions.

Is part of this that we can recommend not only what you should count but how you should count it in order to make these studies more scientifically valid?

DR. WILKIN: I'd like to speak to sort of the technological imperative aspect of this. It's possible to have sort of NASA-level technology that would detect lesions that could be adequately treated that the patient didn't even know they had. So I would hope that there would be some correlation of what was found with these high tech apparatus, how it related to actual clinically apparent lesions.

But having said that, that's sort of a validation stage. Assuming that validation stage can be made, then it seems like it's very objective. Once you buy the machinery, then it probably is cost effective to do lots of studies. It seems like it's a rational approach, yes.

DR. STERN: I guess what I was trying to imply was for once I saw the cup half full rather than half

empty. In fact, some of these methods would allow you to look at not only lesion counts but lesion volumes, for example. One of our guests talked about a real success is taking 50 large inflammatory lesions on day 1 and 8 weeks later turning it into 50 much smaller inflammatory lesions. That was the kind of thing I was talking about, not using ways of elevating what's not important, but rather in fact measuring the things that we all agreed and the experts agreed are very difficult to measure over time as an individual investigator because we're all human.

DR. WILKIN: I think certainly a sophisticated equation that would take those sorts of things into it -- but I did hear from the experts and from members of DODAC and Dr. Bergfeld, before she left. Her first word was "simplicity." There's this great appreciation for elegance of simplicity when one is looking at something that's supposed to be clinically meaningful. So I would come back to that.

DR. KILPATRICK: I'm very much attracted to the concept of using modern technological screening and measuring techniques and picked up on the suggestions that perhaps it even may be feasible now or nearly in the near future to do what you're saying, Jon, but not only number but size, density, color. And we have all of those things.

My problem then is, given these three, four,

five different parameters, how do you combine them. My feeling is that the physician, the dermatologist, is the best person to do that, and in fact that's what he's doing in the IGE. He or she.

DR. STERN: Other comments on question 2?

(No response.)

DR. STERN: Question 3 then, which is, what investigators' global scale should be used? At what level should it be dichotomized into success and non-success?

DR. KING: I've always had trouble with the concept that it's totally clear. I don't think I've ever seen any acne therapy except perhaps acne treated with Accutane where you get totally clear. So I guess a study set up so that your only measure of success is that a topical therapy is going to get totally rid of everything seems to me to be unrealistic. So I always wanted that scale in there 0 and 1 where, I think as Leyden said, should the Pope declare this sainthood, I'd like to see some weight given to nearly clear or cosmetically acceptable because it is true that we recognize our mother in a crowd because she looks like that, but we all have different mothers and we all have different variations of success in a simple kind of thing.

So I would like for the agency to take something to the effect that success, as far as the

physician and the patient, is different, and it's unrealistic I think to demand total clearance. Perhaps you can totally clear inflammatory but not comedones.

DR. STERN: I guess along that line, to me success depends on where you start. If you start with a larger problem in terms of the disease and make it into a smaller problem, that's successful. If you start with not much of a problem and only make it somewhat better, was it worth the trouble? So I think that's an issue in how we guide that.

DR. WILKIN: I'd like to say that I believe the FDA dermatology group is very much on the same page as Dr. King on his comment of having a good grade that is not completely clear but something that is close to that well defined. I think that would be incredibly helpful for us to hear from the committee what that mild category might be that would be regarded as appropriate for a win.

DR. STERN: Another sort of procedural question. In our business, especially in things like acne, things are often visual. So one set of criteria often used for many kinds of things is a set of standard photographs, that when a person looks like -- and you obviously have to have some differences because there will be two inflammatory papules and very few comedones or a small number of comedones, no inflammatory -- if you make it to

A, B, C, or D, if your patient looks like this, this we regard as good as you have to get to consider it a success. And is there a possibility of developing, in fact, standardized photographs for this or photographic standards?

DR. WILKIN: Well, yes is the answer. But along with that, it might be nice to have something in writing which would say this photograph allows post-inflammatory hyperpigmentation, allows X number of comedones, and sort of gives a description and has a photograph so you've got two ways of thinking about it.

DR. STERN: In fact, they may be, for example, gender because people look at -- at least I look at men's and women's faces differently. They may be gender-specific and they may be skin type-specific for some of the reasons that you spoke about as well.

DR. KING: Just as a commentary, having been in on the Accutane brouhaha, it seems to me that this may be something that the American Academy of Dermatology in some subcommittee should help generate this so that it would not be viewed as coming from the FDA down, but it would be an evolutionary process. And you've got to get a community to buy into change if you're going to effect change. So I'd rather see the FDA charge the academy and other interested folks to develop that and then go for agreement.

DR. STERN: Dr. Ten Have.

DR. TEN HAVE: I may have missed this, but didn't Dr. Leyden earlier today talk about standardized pictures? Is that what you're referring to?

DR. STERN: That's exactly what I -- he was talking about standardized pictures within individuals under investigation. Extending that concept, if that's not going to be required, one question gets to be, for judging success, can you give investigators a set of photographs that say this is what people who are successful by our criteria look like at the end of therapy which is a less technological way. You can just give people a bunch of 5 by 8's.

DR. PLOTT: I would second the motion for photographs. I think that we use that in alopecia. That's been a helpful measure. That might be useful.

Also that the global evaluation that may have been proposed -- I have some concerns about the biases toward certain types of lesions, whether inflammatory or non-inflammatory, and difficulty with inflammatory lesions moving from one category into another.

DR. STERN: Yes, Dr. Katz.

DR. KATZ: A question for information. Now, is question 3 for final approval? Or why can't success be evaluated comparing lesion counts?

DR. STERN: I think that goes back to question 1, and I guess question 3 presupposes that we're going to say that you need to make it by criteria in addition to lesion counts. However, we recommend whether that's total, separate for inflammatory and non-inflammatory. So that question presupposes that we come down that in addition to making it in terms of some way of someone quantifying disease, that there be some measure of success that is a qualitative one. And I think the question is, well, what are good qualitative measures of when you're successful, and there are all sorts of combinations there.

DR. TAN: My question is very appropriate, 3.5, in between 3 and 4. I think when I've seen an analysis, I think presented this afternoon, the problem is really with the quantification of non-inflammatory lesions. I think the immediate improvement for all of this is probably a refined measurement of this non-inflammatory lesion, either using the digital photo technology or some more refined procedure by comparing the pictures, even by physicians, investigators. Of course, it will have some subjectivity, but it still would be more refined and would immediately improve the process.

DR. STERN: This is strictly a clinical bias statement, and I'd be interested in the other dermatologists' on the panel feeling about this. When I

see people with mild to moderate acne, including my two teenage daughters, it's the inflammatory lesions that prompt them to have care and how much they care about the comedones, unless they're on their nose and want to use Biore strips on them, is decidedly less of a problem. That's my experience with only two children plus a few thousand patients who are other people's children. I'd be interested to know if I have a deviant experience.

DR. RAIMER: I was just agreeing, shaking my head.

DR. KATZ: Being a practitioner and doing this every day, I take care of both. And there are people, as Dr. Pochi pointed out, who have a massive amount of comedones and no inflammatory lesions. It also points to what you're saying. And there are people with horrendous cystic acne needing Accutane who have very few comedones. So I think it's very important to separate these as far as the appropriate proposed medications being indicated for one or the other.

I don't think that it's much different for the FDA to have criteria on whether a drug works relative to what we do in the office really every day, which is trying to evaluate people from month to month or 6 weeks and to decide whether that patient has improved on that therapy because there's all these very effective therapies that

don't work for everybody. We all know that. Tetracycline might work in 80-90 percent of patients. Well, we try to discriminate those where it doesn't work, and we don't remember. I can't remember 6 weeks later what that patient looked like. So I count lesions and the comedones. I don't count every comedone obviously, but are they numerous, are they a few, are they massive? And we can judge, and I don't see why the FDA can't use the same criteria.

DR. WILKIN: Actually I think it's almost like Dr. Katz has been in some of our internal meetings at FDA.

(Laughter.)

DR. WILKIN: It's just eery.

I think what you described is to get this dynamic sense, what is happening over time. You're actually doing quantification. Is that what I'm hearing?

DR. KATZ: In a loose way.

DR. WILKIN: In a loose way, but you're doing that sort of thing.

I think if you come back to question 1 and our earlier discussion of how we have framed these points, the co-primaries in the past is we see lesion counts as sort of a baseline and then what folks look like at the end, often 12 weeks. So we have sort of a dynamic piece to that.

The global we've sort of thought of as an

incredibly imprecise tool, but it comes closest perhaps to the clinical answer of what people may actually look like in terms of do they need more treatment or not. It's kind of a one-time snapshot because, as Dr. Leyden said, it's hard to go back and remember what folks actually looked like at baseline.

So I think that's the history of how we got there.

I should say that the folks -- and they're all over here. No one at FDA is wedded to a particular way of doing this. We really want to do exactly what Dr. Katz said. We would like somehow, if we can, to make it simple and to have the efficacy determination for approval based on a similar kind of measuring stick that clinicians use when they make their decisions with the patient. That really is why we're bringing the whole thing to the committee.

Having said that, Dr. Plott I think gave an articulate summary of some of the advantages that we may not be tapping into just yet by thinking about indications for other than acne vulgaris, the indication of perhaps inflammatory lesion. You never know, when you write up the questions a month-and-a-half in advance, how the discussion is going to evolve. But of course, if I could go back and redo this, I would make question number 4 number 1 because

I think question number 4 is really -- if you think that inflammatory lesions and non-inflammatory lesions by themselves would stand as indications and then also acne vulgaris would be an indication that would be separate, you may want to go down and suggest different efficacy endpoints for the different indications.

DR. STERN: It seems to me it may not be unreasonable to change the order of the questions tomorrow because, as you've pointed out, that kind of decision making about should there be separate approvability for an agent only for inflammatory acne and what would be the criteria for doing that could in some ways drive a lot of the rest of the conversation in terms of all these other things. So I think that's a very reasonable thing to do and perhaps we'll change the order tomorrow.

Shall we go on to question 4 which we've been really talking about? I'm sorry.

DR. SAWADA: Before you go on, I just wanted to address Dr. King's comment about bringing the American Academy involved in this so it didn't seem like the Accutane debacle. I wasn't present for that. And I knew that Jonathan had kind of a feeling for that, and I was wondering what his thoughts were with regard to this with the American Academy so it didn't seem like it was a one-way street.

DR. WILKIN: Well, I mentally jotted down Dr. King's excellent suggestion. Actually I like having the clinical group think about what the clinical endpoint ought to be. That makes a lot of sense to me.

DR. STERN: On to question 5. Should lesion counts be assessed at multiple time points late in the study and averaged to increase power?

I think the discussion perhaps should be two separate questions. One is how important it is to assess the outcomes at multiple time points when you expect the therapy to work, and then the second is how does one handle those in terms of what's the appropriate analysis.

Dr. Kilpatrick.

DR. KILPATRICK: On the matter of order, can we also bring in the IGE in terms of evaluating at different time points? That may not be feasible but maybe given photographs. Does this presuppose we're going counts rather than IGE? That's your decision, sir.

DR. STERN: I think it's our decision.

DR. KING: Actually it approaches an interesting to me which is that oftentimes we talk about giving therapy and it's evolutionary and we have history and all those things going on, but it seems to me that when the patient comes back at visit 2, 3, or 4, you're actually already doing that globally. When you're not doing a

study, you're trying to decide, well, is this patient going to go on toward Accutane. So oftentimes you tell them the bumps and lumps you've got for the next 6 weeks are yours. After that time, they're mine and then the drug's. So you do these kind of outcomes saying, okay, this looks like it's an explosive episode. It's just going to get worse and worse and worse and go toward scarring. And I'm willing to put up with all the hassle of Accutane and prequalification.

So in these kind of multiple time points, we're doing that already. We may not be doing it in a study, but you're actually seeing them at visit 3, 4, 5, and you're averaging and saying, well, I think the response is working pretty well. Hang in there. Keep taking the medicine. Check on diets and so forth. So I think we're actually doing that in real practice.

I don't know statistically about the power. That's why I was interested in this conversation because I think dermatologists do it routinely. We are measuring whether or not you're on the slope going up or down or you're plateaued, and if you don't get better in a certain time frame, you're already looking for other therapies for two reasons: one, you want altruistically to get them better; and two, you don't want to lose them as a patient.

DR. STERN: Dr. Tan, you had talked about this.

DR. TAN: I have a lot of related questions for the FDA and Dr. Wilkin here. Has the agency ever considered an endpoint using time to dramatic or satisfactory improvement as an endpoint? Maybe for Dr. Alosh as well. Using the time to great improvement, satisfactory improvement.

DR. ALOSH: I'm sorry. Could you repeat the question again?

DR. TAN: It's a time to event analysis instead of repeated measure.

DR. ALOSH: Time to event until you achieve success?

DR. TAN: Yes. How long does it take for the patients to reach a certain good clinical endpoint?

DR. ALOSH: Well, I think we need to agree what's a good clinical because, I mean, if you have well-defined evidence such as death or some well-known defined evidence, then we could talk about time to achieve that evidence.

Now, in terms of the investigator global assessment, we could have someone clear or almost clear. So now this is a clinically acceptable endpoint, and then I think we need to see what's the purpose of that. Are we looking in terms of a duration? What's the duration of the study to achieve that clinical endpoint? So this is one

point of two endpoints, count versus investigator global.

DR. TAN: Yes. Something like from the time you give the therapy to maybe 25 percent of the inflammatory lesions were resolved or gone.

DR. ALOSH: Yes, we could have this. In some application it could be a secondary endpoint, not necessarily for acne. But some sponsor might claim their product could achieve faster success in terms of time than other products, and this could be a secondary endpoint. We have not seen it in terms of acne yet.

DR. STERN: My question was a little bit different. When you look at acne and you have two products, one of which at 8 -- and I understand there will be variance around each observation, but one of which just in the ideal was a 50 percent reduction at 8, 12, and 16 weeks, or 8, 10, and 12 weeks, and you have another product that was 75 percent reduction at 8 and 12, but at week 10, that intermediate point, it was 10 percent worse, which is the better product?

If you average them, those products will be, if I did the math right in my head, identical in terms of the average percent reduction. It would be 75/75 and 10 to the worse. It would give you the same percent reduction as the 50 long. But yet, in fact, as a clinical experience, they would be very different products from a patient's point of

view. I don't know which would be better or worse, but they'd certainly be different in terms of persistence of effect or consistency of effect. And I think that's one of the things you have to talk about once you do multiple times.

I think the problem here, although I can see a sponsor doing that if they have something that acts more quickly than the usual 6 to 8 weeks minimum, you got to remember things can act too quickly because unless they have something that also is anti-inflammatory and reduces prevalent lesions at entry to the study, what we're really depending on for healing and improvement in acne is a natural course of healing. So they'd have to have more than an anti-acne effect. They'd actually have to be working on existing lesions, and then they'd have a big advantage.

The other thing is, of course, with these studies, they're not under daily or weekly observation. That would add a huge burden to the investigator, and you get to the problem of timing. The curves were very nice in that you saw the degree of separation just increased a little bit as time went out and probably the statistical testing, I would guess, for a life table analysis and for these differences in counts would not be that different. If anything, it would be my guess that meeting that

criteria in a life table might be a little bit more stringent.

DR. TAN: Yes, that could be.

But I think here the question is we do want to see how the lesion counts compare between the two groups during a defined period of time. We don't want to average them.

DR. PLOTT: One of the concerns with repeated measures is possibly an interaction between the treatment and time. As we've seen, acne may wax and wane, but during a clinical trial invariably, because it seems to work that way, the patients on placebo tend to get better. If we were to extrapolate that, eventually they may even clear if we waited long enough. What type of consideration is given to this interaction between the treatment and time?

DR. ALOSH: Yes, I agree. I think if you are dealing with repeated measurements, the issue of time by treatment interaction will arise, and you need to test for it. Those analyses which I put, one of them multivariate analysis of variance and the other one generalized linear model, the distinction really, one of them would take the treatment effect for that repeated measurement. The other one you could measure treatment by time interaction.

Now, all of this, I want to reemphasize what Dr. Tan and the discussion here going toward the repeated

measurement approach, really we haven't done it in the past. It was mainly the final assessment which could be week 11 or week 12 or cycle 6 in those contraceptives.

But there is a host of issues when considering repeated measurements. Among them how many time points you are going to consider, and I think this would be related to your question for treatment by time interaction and how close those measurements will be to each other. And if you are reaching week 12 and taking measurements at week 11 and week 12, it would have a different impact than if you analyze at week 8, 9, 10, 11. So there is an issue in terms of design I think, how many time points you want to assess, how close to each other.

Again, I think it's a clin stat issue. So there is more to be done, I agree with you, in that area.

DR. TEN HAVE: A follow-up to your question, Dr. Platt. I thought most of the narrowing occurred early on actually during the washout period and less narrowing occurred later on in the follow-up periods, that most of the placebo effect was that first couple of weeks.

DR. PLOTT: I think what we've seen in most of the graphs, there is a dramatic effect initially. Usually that next visit is at week 2 or 4, and there is quite a dramatic -- but still there's some improvement, maybe even a flattening, but just in the course of the disease, you

might expect that acne gets better or worse or, as individuals grow older, if you stretch that line out to some number in the 20's, much of it will improve dramatically.

DR. KATZ: I don't think that's a big problem.

DR. PLOTT: No, not for clinical trials.

DR. KATZ: No, because in 3 months, the natural history of acne doesn't get better. Now obviously a certain percentage, a small percentage would get better by itself. But in 3 months it's not rapidly, spontaneously clearing the problem like you would say over 3 years perhaps.

The other thing is that that's taken care of by placebo control. The fact is when you have a 60 percent placebo response, like Wilma pointed out in one of her studies with the Ortho Tri-Cyclen, 60 percent of those people -- I mean, talking about that saying, oh, 60 percent of the placebo patients get better. They're not getting better. They're getting recorded as getting better. But we know that 60 percent of people don't get better with nothing over a period of 4, 8, 12 weeks. So they're getting recorded. It's investigator bias which I don't use as a pejorative term for investigators. It's a natural bias. That's the original reason why controlled studies were done way back decades and decades ago.

DR. TEN HAVE: Could they be using something else on the side?

DR. KATZ: Well, the something else is that there are 200 things in the drugstore that don't help very much anyway unless it's a little benzoyl peroxide and that's borderline effectiveness.

DR. STERN: Question 6, how should the efficacy outcomes of clinical trials be portrayed in labeling to be maximally useful to clinicians and patients? What graphics and tables should be provided?

I think we had a rather nice presentation of at least one way that it's being done currently. I guess one question I have, for this very consumer oriented product, since we are certainly unlikely to be increasing life span in our society by treating mild to moderate acne, should there be different information or a different portrayal of information in fact for the learned intermediaries, the prescribing doctors, and for patients? Is this the perfect time to have patient inserts that are, if you'll pardon my use of the words, generic for acne?

MS. KNUDSON: Dr. Stern, I'd like to say as a consumer representative, if you will, unless I misunderstood earlier the discussion about patient satisfaction surveys, they were discounted in the consideration of a drug. I would like to suggest that

perhaps a decent patient satisfaction survey or quality of life survey should be demanded for every study and that part of the patient insert material should be what the reaction of patients has been to the various drugs.

DR. STERN: I think there is at least a group of us in dermatology who would love to see that happen, but so far, if you asked me for a validated acne instrument, I'd have a hard time coming up with one that I would believe gave one robust and interpretable results.

MS. KNUDSON: Does that mean it's just not possible to ever have one?

DR. STERN: Absolutely not. We've heard about where all the funding -- I assume all those NIH-funded trials were all the ones that were for isotretinoin and that was by happenstance because the drug was being investigated for keratinization at the NIH and Gary Peck made the observation that this stuff was dynamite for people who had a disorder of keratinization as well as acne. But to my knowledge, the NIH and government agencies, with the exception of the funding you have, have been particularly silent on this disease, and I don't think industry has seen it as being an avenue likely to be in their benefit.

MS. KNUDSON: Are other kinds of investigators? Psychologists might be willing to do this. There are

people who construct surveys for a living who could, with some input from the appropriate persons, develop a scale.

DR. STERN: I think we have the talent within dermatology. It's the important thing you said, who do it "for a living." And the question is where will the funding come from. That was my point.

DR. LEHMANN: I want to add one thing. We haven't been talking about side effects. As you start talking about how to balance efficacy and what to tell patients, you want to start saying, okay, is the side effect and the degree of side effects worth even the efficacy that has actually been demonstrated.

DR. STERN: I think that's clearly the key point in any clinical decision making, and I think we've been asked to focus particularly on the efficacy side. But I always assume that the agency will pay good attention to side effects and think about ways to portray them. I think as has been said over here, the best way of balancing it is if you had a good measure for patients to express their opinions about how much better on balance did this therapeutic experience make them feel.

DR. WILKIN: That was the clarification that I was seeking. I wanted to know that this wasn't just quality of life based solely on efficacy but based on everything related to using the product. It's helpful to

have that clarified in the transcripts because we'll be pouring over these transcripts for months.

DR. STERN: Any other comments?

DR. SAWADA: Well, in terms of all the modern technology and all, as a practicing dermatologist who looks at the package inserts and tries to glean pertinent information in between patients, if they get too complicated, it's way beyond me. The fine print is getting harder and harder every year to see.

I do not know, but does the FDA have a web site, since so many more of us are becoming computer savvy, where these studies can be consolidated for individual interest for docs who want to do some more exploration in the subject or have some sort of clinical research interest rather than trying to fit it all on the piece of paper?

DR. WILKIN: Well, we do have a web site, and certain drug products get labeling, and special warning discussions and public health advisories and these sorts of things show up on the web site. Independent of that, we're looking to a future some day of electronic labeling where you may still have your PDR and it will be a paper version and if that's what you like, you can -- what I always did, a new product came out and I would actually walk around and in my white coat, I'd have a couple of the new labels so that whenever I wanted to prescribe, I could go over things

and sort of learn about them in the clinic.

But in the future, you'll be able to -- it will be updated in real time, and it will be a lot easier system. So if you're computer literate -- but that's in the future. We don't have that just today.

DR. KING: I guess to come back to one of my issues, which is "yes but" in terms of labeling, it seems to me that once a product, regardless of its original indication, is labeled as effective for inflammatory acne or non-inflammatory acne, most people are just going to prescribe it. And if I were cynical and in industry, I would just try for one indication of inflammatory acne realizing that once it's out there, people are going to use it anyway.

So sometimes I worry about the labeling because when I saw the data that said the difference between placebo was only 7 lesions, if I were a computer game, jean jock kid, I'd say you mean I'm going to go through all this hassle for 7 bumps that are better? I don't think so. So I think we have to be careful with this. I think that sometimes it's better just to talk about efficacy and especially side effects.

DR. WILKIN: Yes, these products are approved with that level, but you have to remember there's a certain artificiality in a phase III study. In your office, you

never ever give someone a prescription and say, this may work for you, or half of the people that get this prescription, they're not going to get anything active and the other half are.

There was an abstract that was presented at the ASCPT meeting. It must have been about 5, 6, 7 years ago now. They looked at the efficacy for a product when it was compared against an active control and showed that it was a much higher impression of efficacy than when that same product would be compared with its vehicle or placebo. So I think there are enormous differences between what happens in phase III and what happens in the clinical setting. So you might actually get more. You do more for your patients than just give them a prescription. You give them all sorts of other things to do.

So I feel that our approval of products that may only change a couple of lesions at the end of the day is consistent with what we've heard from clinicians in the past in terms of something that they find useful and meaningful. And as Dr. Leyden said, not all those products make it on the market. The market can be more Darwinian than the FDA. Nonetheless, I think it's a level of efficacy that we should feel comfortable with. That's my impression.

DR. STERN: It's now 5:30 and I'd like to hear

a motion to adjourn the meeting, and we'll begin again at 8:00 tomorrow morning.

DR. KING: So moved.

DR. RAIMER: Second.

DR. STERN: Thank you.

(Whereupon, at 5:30 p.m., the committee was recessed, to reconvene at 8:00 a.m., Tuesday, November 11, 2002.)