at

AT

DEPARTMENT OF HEALTH AND HUMAN SERVICES

FOOD AND DRUG ADMINISTRATION

CENTER FOR DRUG EVALUATION AND RESEARCH

ORALLY INHALED AND NASAL DRUG PRODUCTS SUBCOMMITTEE OF THE ADVISORY COMMITTEE FOR PHARMACEUTICAL SCIENCE

Tuesday, April 26, 2000

8:30 a.m.

CDER Advisory Committee Conference Room
5630 Fishers Lane
Rockville, Maryland

MILLER REPORTING COMPANY, INC.
507 C Street, N.E.
Washington, D.C. 20002
(202) 546-6666

at

PARTICIPANTS

Vincent H.L. Lee, Ph.D., Chairperson
Nancy Chamberlin, Executive Secretary

MEMBERS

Richard Ahrens, M.D.
James Li, M.D., Ph.D.
Stanley J. Szefler, M.D.

CONSUMER REPRESENTATIVE

Gloria L. Anderson, Ph.D.

INVITED GUESTS

Michael Baaske, Ph.D.
Charan Behl, Ph.D.
Richard N. Dalby, Ph.D.
Harmut Derendorf, Ph.D.
William Gore, Ph.D.
Lester I. Harrison, Ph.D.
Walter W. Hauck, Ph.D.
Sylvie Laganiere, Ph.D.
Thomas R. MacGregor, Ph.D.
Nikhil J. Parekh
Sam C.K. Shum, Ph.D.

FDA

Wallace P. Adams, Ph.D.
Guirag Poochikian, Ph.D.
Eric Sheinin, Ph.D.

C O N T E N T S

C O N T E N T S (Continued)

C O N T E N T S (Continued)

P R O C E E D I N G S

Conflict of Interest Statement

MS. CHAMBERLIN: Good morning. We are ready to start. Bear with us with the sound system. It is very questionable if it is working. We are testing this room and it is testing our limits.

It turns out the room is full. We have set up a few chairs in the overflow room, but your seats are like gold. There will be a change in a few speakers so please bear with us. Yi Tsong will be speaking before Dr. Adams today. There is a change this afternoon. We will have Dr. Derendorf speaking after Dr. Roman.

I am going to go ahead and read the purpose statement for the conflict of interest and then, after that, Dr. Lee will have introductions and open the meeting.

The following announcement addresses conflict of interest with regard to this meeting and is made a part of the record to preclude even the appearance of such at this meeting.

In accordance with 18 USC 208, general-matters waivers have been granted to all committee participants who have interests in companies or organizations which could be affected by the committee's discussion of specific scientific issues where the additional expertise of the subcommittee is sought to aid the agency in refining draft

guidance for orally inhaled and nasal drug products in certain areas of chemistry, manufacturer and controls, and in vitro and in vivo bioavailability/bioequivalence.

A copy of these waiver statements may be obtained by submitting a written request to the agency Freedom of Information Office, Room 12A-30, Parklawn Building. In the event that the discussions involve any other products or firms not already on the agenda for which an FDA participant has financial interest, the participants are aware of the need to exclude themselves from such involvement and their exclusion will be noted for the record.

With respect to all participants, we ask, in the interest of fairness, that they address any current or previous financial involvement with any firm whose products they may wish to comment upon.

I just want to explain in a nutshell what we were trying to say with our legal words, and that is we are not discussing specific products at this committee. This is a subcommittee that is made up of industry. Because we are not discussing specific products, we have given the committee general matters.

This is a subcommittee that will not vote. It will have discussions on issues as the FDA presents them.

Call to Order

DR. LEE: Thank you, Nancy. I think, in the

interest of time, I will just introduce myself. I am Vincent Lee, the Chair of this subcommittee. I do have a full-time job which is I am Chairman and Professor at the University of Southern California.

I think it might be useful to go around the table and everyone introduce himself or herself, where they are from, and we will go from there.

DR. MacGREGOR: I'm Tom MacGregor. I am a Highly Distinguished Scientist at Boehringer-Ingelheim. That is the title. [Laughter.]

DR. ANDERSON: I am Gloria Anderson. I am Callaway Professor of Chemistry and Chair at Morris Brown College in Atlanta, Georgia.

DR. BAASKE: I am Michael Baaske. I am with Alpharma USPD.

DR. LAGANIERE: Good morning. I am Sylvie Laganiere, Director of Pharmacokinetics at Phoenix International now, under new merger, MDS.

DR. DALBY: I am Richard Dalby. I am Vice Chair of Pharmaceutical Sciences at the University of Maryland, completely undistinguished. [Laughter.]

DR. GORE: My name is Bill Gore. I am Director of Analytical Sciences at Boehringer-Ingelheim Pharmaceuticals.

MR. PAREKH: I am Nikhil Parekh. I am Director of Analytical Development at Whitehall-Robins and representing

CHPA.

DR. ADAMS: Wallace Adams, Food and Drug Administration and Office of Pharmaceutical Science.

DR. POOCHIKIAN: Guirag Poochikian, Chemistry Team Leader in Pulmonary and Allergy Drug Products in FDA.

DR. HAUCK: I am Walter Hauck. I am Professor and Head of Biostatistics at Thomas Jefferson University in Philadelphia.

DR. HARRISON: Les Harrison, Division Scientist, 3-M Pharmaceuticals. I am representing the IPAC Bioequivalence Component.

DR. DERENDORF: I am Helmut Derendorf, Professor and Chairman of the Department of Pharmaceutics, University of Florida.

DR. SHEININ: Eric Sheinin, Deputy Director of the Office of Pharmaceutical Science, CDER in FDA.

DR. SZEFLER: Stan Szefler at the National Jewish Medical and Research Center and also a member of the Pulmonary and Allergy Drug Advisory Panel.

DR. BEHL: Charan Behl, EVP and R&D of Nastech Pharmaceutical Company, also representing Nasal Drug Relief Focus Group from the AFBS.

DR. LI: I am James Lee. I am an allergist and internist at the Mayo Clinic, formerly of the Pulmonary and Allergy Drug Advisory Committee.

DR. SHUM: My name is Sam Shum, Director of Analytical Chemistry for Aerosols of KOS Pharmaceuticals.

DR. LEE: Thank you very much. I do want to remind everybody that these proceedings are being taped. Let me open the meeting by inviting Dr. Eric Sheinin from the FDA to give an introduction and state the objectives of the meeting.

Introduction and Objectives

DR. SHEININ: Good morning. I have to say I am a little overwhelmed by the size of the audience. I don't know that we expected quite this many people here. It is very rewarding and encouraging to see the tremendous amount of interest in this area.

[Slide.]

What I would like to do is kind of set the stage for the discussions that we are going to have throughout the day today.

[Slide.]

The responsibilities for the subcommittee are mainly three. One is to, certainly, address and discuss the questions that have been raised and presented to the subcommittee that are related to the issue of content uniformity for both orally inhaled and nasal drug products. There are, I believe, two questions that need to be addressed in this area.

Once we are finished with that, we will ask the subcommittee to go on and address and discuss the questions that have been developed and submitted to them that are related to both in vitro and in vivo bioavailability and bioequivalence for these types of drug products.

As Dr. Lee mentioned, there will not be a vote taken at this subcommittee meeting. What we want from the subcommittee is a thorough scientific discussion of the questions and the issues. There will be transcripts of what takes place today and this information will be used by the agency as we continue to go forward with the development of guidances and policy related to these types of drug products.

Finally, there will be a presentation by the subcommittee. As to who will actually make the presentation, it may be the Chair. It may be somebody else that is designated by the subcommittee to represent them at a formal meeting of the Advisory Committee for Pharmaceutical Science.

The next meeting has tentatively been scheduled for August of this year but there is a possibility that we may postpone this until September or October because there are other issues that we would like to take to that advisory committee meeting and we may not be quite ready by August to discuss those at the advisory committee.

at

[Slide.]

To date, we have issues three draft guidances in the area of orally inhaled and nasal drug products. Two are related to chemistry and manufacturing and controls and one is related to bioavailability and bioequivalence. We are planning to issue another guidance on bioavailability and bioequivalence for the nasal spray and inhalation, solution, suspension and spray drug products. That will be issued for public comment at some point in the future.

[Slide.]

Just to quickly go over what the questions are. I presume everybody has these questions now. There are the content-uniformity questions. These are the CMC or chemistry, manufacturing and controls questions.

Should there be a single content uniformity standard for all orally inhaled and nasal drug products or, conversely, one could look at it, should there be different content uniformity standards depending on each individual product or type of product. And should the FDA continue to develop a proposed statistical approach to evaluate the data that are obtained when content-uniformity testing is performed?

So those are the two CMC questions that need be addressed by the subcommittee this morning.

[Slide.]

Then we will go on to the bioavailability and bioequivalence. The first set of questions are related to in vitro bioavailability and bioequivalence testing. The first set will deal with profile analysis. Should all the stages of the cascade impactor be considered when looking at the data and evaluating, comparing the reference product and the product under development or discussion, or the product that is the subject of an application submitted to the agency.

Should there be a statistical approach as opposed to a qualitative comparison for these profiles. If the answer is yes, then is the chi-square comparative-profile appropriate or should there be some other approach? Is chi square, by itself, sufficient or should we go on from there?

[Slide.]

Also, under in vitro testing, for dry-powder inhalers, the comparability of them. But, prior to doing in vivo studies to establish the equivalence of these products, a firm would need to design its product to have the best likelihood of being found equivalent in these in vivo studies.

There is a type on this slide and I would ask everybody--the second line from the bottom where it says, "What comparative in vivo tests should be conducted?" that should be in vitro tests. If everybody would please correct

that. There is a significant difference there.

Also, what design features of the device and formulation and what parameters should be considered by the firm developing the product in trying to determine pharmaceutical equivalence of these products.

[Slide.]

So those are the in vitro bioavailability and bioequivalence testing questions that we would like the subcommittee to address. Once that portion of the discussion is completed, we will go on to in vivo testing questions.

The first set deal with clinical studies that would be designed for local delivery of nasal aerosols and sprays. In the draft guidance, three study designs have been proposed for drugs that are intended to have local action. These are traditional treatment study, days-in-the-park study and environmental-exposure unit study.

All these designs are based on seasonal allergic rhinitis. The first question is, is it feasible to demonstrate a dose response for these locally acting nasal drugs. If it is not, what other approaches would the subcommittee recommend? What else could we and the industry rely on to establish that these are equivalent.

[Slide.]

The next question is can bioequivalence be

established based on seasonal allergic rhinitis. Can this assure bioequivalence for other indications? I think that is a very significant question and we would like the subcommittee to address that today.

[Slide.]

In terms of clinical studies for local delivery of orally inhaled corticosteroids, again, a number of approaches have been proposed to assess bioequivalence of these products; as examples, clinical trials, bronchoprovocation tests, steroid-reduction model, trials with surrogate measures such as exhaled nitric oxide.

We would like the subcommittee to address these in terms of are any of these study designs proven to offer better discrimination in terms of dose-response sensitivity.

[Slide.]

Continuing with the clinical studies for orally inhaled corticosteroids, are there any other in vivo approaches. Again, there are some examples given, surrogate markers that might be sufficiently sensitive and validated to establish in vivo bioequivalence and bioavailability for these inhaled corticosteroids.

We would very interested in any advice and counsel that the subcommittee can present today during their discussions.

[Slide.]

Finally, in terms of the area of PK or PD studies for systemic exposure of locally acting drugs, are there any situations where in vitro data plus systemic PK and PD data can be relied on to assure local drug delivery for either nasal or inhaled products.

These are the questions that we would like to have addressed. It is a very, very full agenda. It certainly is all of our sincere hope that the subcommittee will be able to get through all of this today. Again, as Dr. Lee said, in the interest of time, I think we should proceed to the first topic and, hopefully, we will get through everything by sometime late this afternoon.

If you notice there is no adjournment time given. I assume people have flights to catch, so we will do our best to stay on schedule.

Thank you.

DR. LEE: Thank you very much, Dr. Sheinin, for this very nice introduction.

I would like now to move to the first section of this meeting which is the CMC on content uniformity. Dr. Guirag Poochikian is going to provide us with the current FDA practices for NDAs.

Chemistry, Manufacturing and Controls: Content Uniformity

Current FDA Practices for NDAs

DR. POOCHIKIAN: Good morning.

[Slide.]

First I would like to give you a brief background information concerning the genesis of the guidance.

[Slide.]

Because of the Montreal protocol and the various proposed phase-out programs of the CFC-containing products, the non-CFC-containing drug products such as MDIs, DPIs and other inhalation drug products have received a great deal of attention such as alterative formulations and container-closure systems to deliver the required dose to the biological target, appropriate regions of the lungs or the nasal airways.

Due to these activities, the agency took the initiative in drafting two of these guidances. They are called metered-dose inhalers and the IN DPI drug products CMC documentation, and the second one is nasal-spray inhalation, suspension and spray-drug products, again from a CMC perspective. They are cited on the website. The address is there.

The purpose of this is to cover essentially most of the inhalation drug products which are currently available or are under investigation. As all of us are aware, these drug products are complex units with many challenges. They do have unique features compared to other more conventional drug products with respect to formulation

components and their suitability for inhalation use with regard to container-closure systems, delivery systems and with regard to the controls of each of these components as well as the drug product, itself.

We, at the agency, are interested in publishing, both scientifically and regulatorywise, sound guidances, always having in mind, of course, the public-health interest.

We, at the agency, are of the opinion that such guidances will help drug-development efforts for these unique drug products, facilitate submission and review of these applications, expedite the approval of these important drugs and make them available in high quality to the public.

The content of these guidances are based upon experiences, issues that have been dealt with and challenges that have been faced during the development and review of numerous and different types of drug applications, particularly in the last decade.

Essentially, these two guidances summarize and organize the information acquired in the last decade in a user-friendly manner to be easily and equally accessible to the interested parties. In a nutshell, these guidance delineate the current practices for NDAs.

[Slide.]

The scope of these guidances are outlined on this

slide. As you see, there are two sets. The first set covers the NDIs and DPIs, non-aqueous based and the second set covers the aqueous-based preparations.

[Slide.]

I would like to say a few words about the guidance philosophy because it is important to our discussion. As any other guideline, these guidances also set forth approaches which are acceptable to the agency for submission of the CMC information. Also it presents the agency's current thinking on the CMC documentation for inhalation drug products. Also it indicates that alternative approaches may be used.

Also, in conjunction with that, it encourages discussion with the agency review division for significant departures. Like any other guidance, also there is a statement saying that it does not create or confer any right on any person and does not operate by FDA or the public.

[Slide.]

What are the activities since the publication of these guidances? The first NDI/DPI was published in late October and the public comment period was closed in early March, 1999. A workshop sponsored by AAPS/FDA/USP was held in early June and, similarly, the public comments for the second guidance which was issued on June 2 of 1999 was closed in early September of 1999. There was a preliminary

subcommittee OINDP meeting in early November.

[Slide.]

As to the dose-content uniformity, to insure the drug-product quality in terms of dose consistency, the dose-content uniformity issues need to be addressed from three different perspectives. First is unit-to-unit dose-content uniformity within a batch--that is inter-unit or inter-container or intra-batch dose-to-dose uniformity.

The second is dose-to-dose content uniformity within a unit, within a container, intra-unit from the beginning to the end of a unit. The third one is batch-to-batch dose-content uniformity which is inter-batch which is not the topic of discussion today. That is usually handled through stability studies.

[Slide.]

What are the acceptance criteria currently being used for NDAs at FDA? First, with regard to inter-container dose-content uniformity. It consists of two tiers. In the first tier, there are ten containers or ten units and one determination from each unit.

That particular batch would be considered acceptable if not more than 10 percent, outside 18 to 120 percent of the target-emitted dose and none outside plus-or-minus 25 percent of the labeled claim provided the mean of the ten determinations are within plus-or-minus

15 percent of the labeled claim.

If 20 or 30 percent of those determinations are outside plus-or-minus 20 percent of the labeled claim and none outside plus-or-minus 25 percent of the labeled claim, and provided the mean still is within plus-or-minus 15 percent of the labeled claim, then the second tier may be utilized by doubling the sample size to twenty.

So, in total, there will be thirty determinations. Out of those thirty determinations, an oral batch will be considered adequate and acceptable if not more than 10 percent is outside plus-or-minus 20 percent of the labeled claim and none outside plus-or-minus 25 percent. Again, the mean shall be plus-or-minus 15 percent of the labeled claim.

[Slide.]

With regard to intra-container dose-content uniformity from the beginning to the end of a unit, again, it consists of two tiers. The first tier uses three samples and taking samples from the beginning, middle and end so there will be three determinations per unit. In total, there will be nine determinations.

The batch will be considered acceptable for intra-container dose-content uniformity if not more than one out of nine shall be outside plus-or-minus 20 percent of the labeled claim and none outside plus-or-minus 25 percent

provided each of the means, beginning, middle and end separately are within plus-or-minus 15 percent of the target-emitted dose.

If two or three of those determinations are outside 80 to 120, provided those are not outside plus-or-minus 25 percent and provided each of the means are within plus-or-minus 15 percent, then a second tier may be utilized, again by doubling the sample size.

So, in total, there will be 27 determinations considering the initial tier 1. In that case, the overall batch will be acceptable if not more than three are outside plus-or-minus 20 percent of the target-emitted dose and none outside plus-or-minus 25 percent of the emitted dose.

The mean for each at beginning, middle and end shall be plus-or-minus 15 percent of the target-emitted dose.

[Slide.]

I would like to say a couple of words about the testing conditions. When these acceptance criteria were specified, we had in mind certain assumptions. First, the samples are stored under specified storage conditions and orientations because it is well-known that some of these products will have significant variability, high variability, if this is not done. So that particular variability has been removed from these test conditions.

Second, trained personnel are used to follow certain standard operating procedures for testing each of these units because, again, it is known that high variability will be obtained if these procedures are not followed in terms of uniform shaking, how long they shake it, how frequently the mouthpiece or the actuator is cleaned or new actuators are used, what is the depression force and the actuation force, what is the store plant and so forth.

All of these conditions will impact negatively if they are not controlled. Again, these factors have been eliminated from these test results.

More importantly, all these units are fully primed before testing and all of us know what is the significance of priming because unprimed units will give totally aberrant results, also.

Next, the test results are obtained under specified testing conditions; predefined flow rates and predefined duration. Of course, this applies mostly to DPI situations. So we need to consider all these factors in our deliberations.

[Slide.]

As to the public comments, I would like to summarize the various categories concerning dose-content-uniformity specifications. One category of comments, actual specifications for DCU, should not be incorporated into the

guidance. That category, they didn't want these specs in the guidance. One explanation, which is in quotations, says, "Note that each drug is unique with respect to the capabilities and reproducibility of the manufacturing process, device components and analytical methodology."

Of course, that is a disturbing comment, but that is what the explanation is. "And that these parameters should be considered in establishing appropriate specifications."

[Slide.]

Another category says to establish a process by which DCU specs may be determined on a case-by-case basis. However, that category of comments did not provide what process they had in mind.

The next category recommends to retain the guidance specifications, however, to widen the individual dose-acceptance criteria. Here, they are referring to the inner and outer limits of plus-or-minus 20 percent and plus-or-minus 25 percent, respectively.

The next category of comments indicates to recommend, retain guidance specifications in the draft, but, However, to delete the mean criterion for the first tier. As you heard, in our proposal, there is a mean criterion at each tier, tier 1 and tier 2. This particular one recommends to delete from tier 1 and applied only to tier 2.

The last category I will comment on was the recommendation to provide a process for setting dose-containing-uniformity specifications using statistical procedures. That, Dr. Hauck is going to discuss.

Thank you.

DR. LEE: Thank you.

Dr. Hauck:

Alternative Statistical Approaches

DR. HAUCK: Good morning.

[Slide.]

We are feeling a little bit on the wrong side of the technological divide this morning. Transparencies, it is. I was asked by the agency, actually some time ago, to do an evaluation from a statistical perspective of the dose-content-uniformity criteria that were in the draft guidances.

What I will be presenting this morning is sort of the state of what that evaluation is.

[Slide.]

The usual disclaimer applies here. I am speaking that the work is supported by the FDA through a contract to Jefferson but the opinions I will be expressing are solely those of myself and should not be construed to represent the agency's opinion.

[Slide.]

I will do this, kind of first comment from the statistical perspective on the content-uniformity standard that Guirag just presented to you and then outlining alternatively how a statistician might go about doing this.

[Slide.]

So this is the within-batch between-canister dose-content-uniformity standard that Guirag just presented to you. There are the two tiers and the variety of the requirements. Largely, in terms of what I will be talking about, I will actually be focussing primarily on this part of the requirement in terms of the 80 to 120 piece.

This part, the 75/125, I tend to think of as a safety net and it really needs to be thought of separately.

The first thing that is important when looking at this criterion, or really very similar criteria from the USP or the CPNP European guidance, is what is the unit, what is being looked at as the unit of analysis here. So the dose-content uniformity within batch that Guirag was talking about is one dose per container.

So we are talking about ten or thirty containers or canisters from a single batch. This is important because one of the things you sometimes hear is why doesn't the FDA adopt the USP requirement. The first thing you have to see is that the unit, really, is very different than the USP because the USP is doing one or three canisters with up to

ten doses per canister.

So the first statistical comment is to realize that the FDA is intending to draw inference or trying to say something about the batch that the USP requirement is not.

[Slide.]

So this is what I am just mentioning here. That is really the first thing when you look at the different requirements. So the USP is actually different than any of the other criteria that have been proposed by both the FDA, the JP, the CPMP and PhRMA.

The next thing, when I look at this, what I am saying is a statistical-hypothesis test. It is not being labeled as such, but there is a standard to be set, there is a decision to be made, data collected, somebody to evaluate the data, the data meets a certain criterion and you say you pass. If it doesn't, you say, no pass.

While it has to form the statistical-hypothesis test, there is something very crucial missing; that is, what are the hypotheses. So I think, really, the primary takeaway message that I would like to leave you is that maybe the focus should be not on whether it is one tier and ten canisters but on what the hypothesis should be that the dose-content uniformity is intending to address.

While I am focussing today on the FDA's criterion because that is the topic of the day, I guess I would like

to be clear that most of what I am going to say really applies to all the other criteria as well, so this is not singling out the FDA in that regard.

The other issue that is sort of statistically evident from the structure, again, at both the FDA, USP and CPN--actually all these proposals--is that two tiers has looked at what, in the clinical-trials literature, we call an interim analysis. You collect some data. If the data is good enough, you say pass and you are happy. If the data is not good enough, you go on and do some more.

That is relevant statistically because that is two opportunities to make the decision and that needs to be thought about in the statistical methodology.

[Slide.]

To put this in a statistical perspective or an alternative statistical approach, I now need to be kind of sure we are in agreement on some of the language, so I am going to ask you to use my language a little bit for the next few minutes. We need to have two different error rates, or error probabilities, that we talk about in this field.

A more general term would be a false-positive rate first. This is sometimes referred to as the consumer risk. In standard statistics books you usually see this referred to as alpha or the type-1 error rate. What we are referring

to there, in false positives, it means a batch that, in some sense, is unacceptable. I put that in quotes because the unacceptable part of it is really not my part of it. That is more Guirag's field.

Some batch that is unacceptable, then, is passed by this criterion putting the consumer at risk. The flip side of that is the false-negative, most typically referred to in statistics books as a type-2 error or beta and sometimes referred to as producer risk. This is the chance that a batch is absolutely fine.

In a clinical-trial context, what we typically would have set up is the false-positive rate would be set usually at 0.05, sometimes less, and the sponsor of the study determines what producer risk, or false-positive rate, they are willing to accept.

In the context of dose-content uniformity, you need to have two more things I need to talk about. One is the target interval. That target interval corresponds to the FDA's criterion, the 80, 120 percent, the idea that most of the batch had fallen in some interval. In that sense, it is a target. And then a target-coverage probability; how much of the batch had fallen in that interval.

Again, as soon as we start talking statistics, 100 percent is not going to be the number there. It is going to be some target probability that we would like to

achieve.

[Slide.]

I put this in here to help you convert back and forth between the notion of a target interval and target-coverage probability to mean and standard deviation in the batch. I had trouble trying to come up with versions that print well in black and white, so the main thing in the handout for the committee is the lowest curve is the narrowest so it is the 85/115. They just go monotonically up to 65/135, the top one. The widest is 65/135.

The idea in this graph is that, in the batch, the batch has some average, here expressed as a percent of labeled content, some standard deviation. All combinations of mean and standard deviation that are on or below the curve satisfy 90 percent coverage--that is the target probability--on that target interval.

So, again, we have the target coverage, the target interval. So everything on or below this curve corresponds to at least 90 percent of the batch falling within 85 and 115 percent of labeled content.

That is to help you translate back and forth between the two.

[Slide.]

If I were starting this from the beginning, and once you say it is a hypothesis test and recognize that it

is really in the form of hypothesis test, how might we go about doing it? I would say the first thing is the regulatory agency should, in fact, specify the hypothesis, what is the claim that they are seeking to have demonstrated.

In this context, what that might look like is a statement in the form here. The numbers in here are ones that I put in here just to be specific. Again, I am not trying to advocate particular numbers, just a notion or concept of how to go about it. But the agency could say, "demonstrates an alpha of no more than 5 percent, a consumer risk of no more than 5 percent, that at least 90 percent of the batch falls within 120 percent of labeled claim."

What they would not do is specify number of tiers. They would not specify the number of canisters per tier. That would fall to the sponsor. Both of those issues are producer-risk issues and if you are going with this sort of approach, the sponsor should choose what producer risk is acceptable to them.

That would leave the sponsor to say, do they want to do thirty or forty or fifty, whatever risk they would like to take or accept for themselves.

The last comment here is just sort of the statistical side of things, to remember that if you go with more than one tier, and there is certainly nothing special

about two, either, that, when thinking about what the false-positive rate is for the process, that all those tiers have to be taken into account.

[Slide.]

There are two types of language I think that actually comes out of a paper that was published by the Japanese Pharmacopeia. You have tests for attributes and tests for values. The FDA, the USP and the CPMP are tests for attributes. They only look at whether or not the particular sample falls within the target interval and they are not using the actual value.

So a test that is right at the limit of the target interval and a test that is labeled claim are treated identically the same in this sort of approach.

Item 2 is, again, more in my language because when I look at what the CPMP and USP and FDA are doing, I say, "Well, I recognize this. I do this every day in my job. This is the standard test to use in designing phase II oncology trials." It is a very standard test and there is a good literature for it.

[Slide.]

So I am able to go to what is pretty much the standard reference. I have given it here, a paper by Richard Simon in 1989. It tells you how to design these sorts of trial.

Let me take you through this so you can see the sort of approach that this would be. This table is for designs where the agency has had 5 percent consumer risk and the sponsor is choosing to design at 10 percent producer risk.

The first column is the target. The agency has specified this. I am giving you a kind of a range of numbers here. The second column is what the sponsor thinks their batch would actually satisfy. In the first row, the agency has said to demonstrate that at least 80 percent of the batch falls in the target interval. The sponsor says, "Well, you know, I have got a really good manufacturing process. I think 95 percent of my batch falls in the interval, so I can do this study by doing the first tier of 23. If no more than one are outside the target interval, I am okay. I stop. Otherwise, I add another 28 for a total of 51. If no more than four are outside, I pass."

That would be a two-tier test for attributes that has the specified statistical properties, the 5 percent alpha and 10 percent user risk.

I have given you a couple of other examples here. The greater than 110 is really just a software limitation, the particular version of Simon's program I am using doesn't go higher than that. I sort of figured anything higher than that, you weren't really interested in anyway. That is

high.

The last row is an attempt to reverse-engineer the current FDA criterion--that is, to go back and say, if we take the two tiers with 10 and 20, what would the hypothesis be in order for that rule to correspond to a 5 percent alpha and a 10 percent producer risk.

To get the alpha down to 5 percent, I have to specify the target-coverage probability as 60. So, in effect, the current FDA, or the proposed FDA, content uniformity criterion is really just seeking to show that at least 60 percent of the batch falls in the target interval.

Then, to get the producer risk down to 10, you would have to be saying that the sponsor would have to saying, the sponsor would have to be saying, that at 91 percent of the batch is actually in the interval. You can see, I can't quite match up exactly, but I get similar properties with 22 instead of 30.

[Slide.]

That is the test by attributes. I mentioned, there is an alternative which is test by value, to actually use the values of the results, not just the dichotomy of whether they fall in the target interval. The proposals for doing that--there has actually been some work on this by the JP and PhRMA working group.

These are what are called tolerance intervals.

Tolerance intervals--first of all, they are not confidence intervals, so don't confuse the two. They are the intervals that you calculate the data and for which you can make a statement like the following; they have some level of confidence, and I picked a number, 95; that the interval--that is, the interval calculated from the data--covers at least some proportion of the batch, and here I picked 90.

The way you could work this in the content-uniformity situation is, again, we have some numbers here so that the regulatory agency could say, "95; yes," and 90, and specify target interval. When you calculate the tolerance interval, if that tolerance interval falls in the target interval, you pass. If it doesn't, you don't pass, and you have got a kind of very simple--it would be very analogous to what is done for oral products for bioequivalence except using tolerance intervals.

[Slide.]

It happens that tolerance intervals come in parametric and non-parametric forms. The non-parametric one, actually, looks a lot like the FDA criterion. Then, for the parametric, you assume a normal distribution which is, certainly, a testable assumption but probably reasonable here. I have given you a reference for one of the standard papers on the topic.

I am mentioning this today because, although I

can't give you the numbers on it yet, and that is why I said it is a work in progress, it would seem that, by making the parametric approach and going to the tolerance intervals rather than test on attributes, we would be making better use of the data and that should translate into smaller sample sizes for a given level of producer risk. That is certainly a desirable end.

[Slide.]

So, in summary, a statistical perspective on this would say that the agency should start by specifying their criterion and not the acceptance rules. The advantages to this seems, to me at least, that the agency would be concentrating on really working on what is an acceptable batch of product. It, in turn, would give the company more control over the design of their studies and, explicitly, then, over their producer risk.

It is fair to say there is a price to be paid for this which is that it certainly appears that the current standards are sufficiently loose that going this approach would tend to lead to larger sample sizes. That does seem to be the bane of the statistician, always giving you larger sample sizes than you want.

As I said, until we finish the work on the parametric tolerance intervals, we can't really tell you exactly how much larger that might be, if at all.

Thank you.

DR. LEE: Thank you very much.

Now, we have heard the background for the discussion. We do have about thirty minutes for discussion. I am not sure whether or not the audience is aware that there are two subsets of participants around this table. There are four members of the subcommittee. These are Dr. Szefler, Dr. James Li, Dr. Gloria Anderson and myself.

What I would like to do is to invite the members to express their opinion and then we will opinion the discussion around the table. Perhaps, I think it would be appropriate to devote about fifteen minutes each to the two questions.

I will read both questions to you. The first question, concerning content uniformity, is, "Should there be a single content-uniformity standard for orally inhaled and nasal drug products?" The second question is, "Should the FDA continue development of the proposed statistical approach to evaluating content uniformity?"

I would like to open the discussion of the first question, should there be a single content-uniformity standard for OINDPs?

Subcommittee Discussion

DR. SZEFLER: Very nice presentations in terms of organization but I wondered if some of the speakers could

reflect on the present status of the available products in terms of are these goals that were identified as achievable with our present products and do they lead to, then, kind of mass changes in the products that we have available.

DR. POOCHIKIAN: I would say that most of the products which have been approved in the last decade have been approved under those criteria that I presented.

DR. LEE: Walter, anything to add?

DR. HAUCK: No.

DR. LEE: Other questions from the subcommittee?

DR. GORE: Can I ask more of a procedural question? We will hear more information this afternoon in the 1:30 to 3:30 slot. Is it our intent to hold this discussion at this time or come back to it later today?

DR. LEE: I think at various times, we will come back to it but this is only the time for this section of the meeting.

DR. LI: I would like to ask a question having to do with the comparison of the dose-uniformity inhaled products compared with orally available products. If we just shift for a moment to orally delivered medications, what is the range of dose uniformity typically that is expected in that area?

DR. POOCHIKIAN: For solid oral-dosage forms, for example, we apply the USP specifications. If I remember

correctly, the concept is the same. However, the ranges are slightly different in the sense that the first tier has plus-or-minus 15 percent if I am not mistaken. Somebody can correct me. The second tier allows plus-or-minus 25 percent.

DR. LI: It is a little tighter but, really, not all that much tighter.

DR. POOCHIKIAN: And there is a standard deviation, again, going from memory, 6.something the first tier and 7.8, I think, is the second tier. There is a standard-deviation criteria added which we don't have it here.

DR. GORE: I had a question for Dr. Hauck. At the end, you indicated when you complete the project you started. Would you comment a bit more on what you think needs to be done to further this process of developing an approach to drug-uniformity classification?

DR. HAUCK: I think a simple answer would be that I need to be able to show you a table for a parametric approach that corresponds to the table I showed you today using the Simon approach. So right now, today, I can't tell you what a sample size would be. That is sort of the bottom-line number, I think, for a lot of people is what size studies would we be talking about for whatever level criterion and levels of risk acceptable.

So I can't give you that today, and that is something that I think you should have before making a final decision on it.

DR. GORE: Would it be helpful to actually bring forward some data that would reflect the performance of products that have in the market today? Would it be helpful to look at the model in the context of data from, let's say, currently marketed products?

As we look forward here, we have an array of products on the marketplace today. But shouldn't we also look forward to the future and a whole new generation of new products, new dosage forms, new delivery systems?

DR. HAUCK: I think your question actually sort of falls right here on the table. In some sense, yes; I would like to see--data is very helpful. There is one question about this. And it would serve two purposes here. One is to go forward on the parametric side, there will be some assumptions that are not currently there and we need to be comfortable that that is reasonable.

The second part of it is to get better experience with how the procedure would behave with real products and real data and that would be desirable as well. The reason i said it sort of falls between Guirag and me is that that also, then, feeds back into what is a sensible target interval. A sensible target interval is not my problem and

I am not qualified to give you what that interval should be, just to help you do the statistics on it.

MR. PAREKH: The first question I have is with regard to the intra-container document on uniformity. We have specifications in the first year which states that the means for each of the beginning, middle and end should be within 85 to 115 percent.

I think this question may be posed both to the statistician we have and also to Dr. Poochikian. I am not very clear. When you start with the inter-batch criteria where you have ten samples that you start out with and you have an overall mean of 85 to 115 percent, it somewhat makes sense to me.

But when you start going and creating a mean for beginning, middle and end, now you drop your N to only 3. That mean just doesn't make sense to me. I think, in my opinion, the mean in the case of intra should be only restricted to the overall which is after you complete tier 2, not at tier 1, because the N is just too small to make any meaningful data and related to the quality of the product.

So that is one question, or the comment I have. If you would like to comment on that one.

The other thing I would like to comment on overall with respect to what is the position of the OTC industry,

which your non-prescription portion of the organization is in the pharmaceutical arena, and this relates also to the question which has been posed here. One of the questions that has been posed here is that should the FDA consider the development of a statistical approach.

I think it makes sense in some cases but, from the consumer side of the business, these products are in the making for many, many years, in some instances, maybe even thirty years or more. To apply the criteria now, of course, forces the industry to go and look at the product but, ultimately, shouldn't the product quality also be looked at from the safety perspective?

Is there a need for restricting the industry to a place where industry ceases to do the business? It is ultimately affecting the consumer.

So those are the two comments.

DR. LEE: Are you expecting a response?

DR. POOCHIKIAN: With regard to the first question concerning the mean beginning, middle and end, there is a good reason for that. You can always increase the sample size. I don't think the agency will object to that. So I have to make that one point clear.

Second, the reason we did that is because we want to avoid sigmoidal curves of units which starts, for example, very high and, by the time of the 200 actuation, it

looses 40 percent. I am talking here, on the average. The average of 30 doses, for example, loses 40 percent. So if you take the individual dose, I don't know what we are measuring.

So, in order to avoid that approach, we wanted to establish that the patient's individual needs and that the patient is getting the prescribed dose rather than 80 percent lower than what is the LCEs on an individual basis, because this is the mean of 30 doses I am talking about. It lost 37 percent.

So we want to avoid those situations like that. As to the second comment, with regard to the oral products, as I said earlier, most of the NDAs which have been approved in the last decade fall into the category that I just presented.

There might be a couple of those in the case of CFC products which will be phased out anyway within the next several years. So it is not an issue. It is possible, also, that the old one can be grandfathered as long as they are on the market. That can be handled at the agency level with a policy.

But we are talking about where we proceed scientific from now on.

DR. DALBY: Let me apologize for my lack of statistical prowess. Neither one of your approaches seems

to pay any attention to the size of a batch, so if a large batch is approved compared to a small batch, does the producer or the consumer incur any extra risk if there is any, in both approaches?

DR. HAUCK: No. Really, both approaches are thinking of the batch as being just much, much larger than the sample size. Once you are out there, whether it is a thousand times larger or ten-thousand times larger, really, doesn't enter into it.

If you were doing batches of size 40 and sampling 10 to 30, then it would be an issue for the producer, clearly. The small end of it would change things but not the large end of it.

DR. DALBY: What about from your perspective?

DR. POOCHIKIAN: No. We do not, at least as it is being practiced now, take into consideration the batch size and the sample size relationship. But that is a very valid point because the batch sizes for some of these products varies significantly from product to product.

DR. DALBY: I guess I have a follow-up question for Walter which is, is there any disadvantage to the patient of the producer deciding that they will accept an enormous risk?

DR. HAUCK: As long as you set it up so the agency is specifying, this is what is deemed an acceptable batch,

and then the agency specifies the allowable consumer risk; that is, the chance that something that doesn't actually satisfy those criteria gets out the door.

Then, what the producer takes as their risk doesn't really affect the consumer in the safety sense. It might be affecting them in some cost-of-product sense if not enough batches are getting out the door. Producer risk has that impact but, in terms of safety, it shouldn't.

DR. DALBY: So there would really be no need, in this approach, for the agency to set guidelines.

DR. HAUCK: Just think--I keep coming back to the clinical-trial context which, in a sense--not that it is directly applicable but the sort of structure applies, which is companies designing a pivotal trial, they decide the power they want. If they don't pass, they don't file. It is their problem in that sense.

DR. DALBY: Thank you.

DR. LEE: Any questions from this side of the table?

DR. HARRISON: I also agree that the FDA should continue to develop a statistical approach. It makes a lot of sense to me, to let the producer determine its own risk, pick the numbers. I like that so I would certainly like to see that developed.

I would like to some real datasets evaluated,

though, to really look at the specificity and get a better feeling for what can be done.

DR. LEE: Anyone else?

DR. LAGANIERE: This is a question for Dr. Hauck. You are suggesting an acceptance range from 80 to 120 percent.

DR. HAUCK: No; I am not suggesting any particular number, there.

DR. LAGANIERE: I wonder if you can expand a little bit more about this number. In the context of oral-drug administration, concentration is kind of related to effect. What does it mean in the context of in vivo?

DR. HAUCK: The best answer that I can give is that it is probably not a question for me. As to what is an acceptable target interval, I can probably help you design a study to think about that, maybe.. But we are talking about data that may or may not currently exist in the literature, I guess. So we are dealing with surrogate endpoints, and all of this is a surrogate in some sense for eventual clinical application.

So you and Guirag would need to sit down and have a dialogue as to what that interval should be so as to properly protect the patient without causing undue burden on the company.

I tried to be clear in the presentation that I was

putting numbers in there to be specific. But I do not choose to defend any particular number as either too large or too small. I can't do that. That is the best I can do, I think. Does that answer the question?

DR. AHRENS: Richard Ahrens, University of Iowa. Dr. Hauck, Iwanted to clarify I think what I heard a bit ago and expand a bit. I think your approach assumes normality of the distribution of content uniformity, does it not?

DR. HAUCK: The parametric-tolerance interval approach, what I mentioned at the very end, would assume normality in the canister-to-canister values. The approach that I gave you, the table, is based on the method assigned and makes not assumption of normality. It is just, they are in the interval, they are outside the interval.

So there are the two choices there. If normality turns out to be an untenable assumption in this context, then you can fall back to the test by attributes.

DR. AHRENS: Is there any evidence as to whether uniformity tends to be normally distributed or not?

DR. HAUCK: I haven't seen enough--that is one of the things I would like to see or suggest more data. I would like to see more data to address that. From a theoretical perspective, it strikes me as exactly a situation where normality ought to be sensible. It is a thing called the central-limit theorem that says that if

your error is built up of lots of little components, what you end up at the end is something that is going to look approximately normal.

So, from that perspective, that seems to fit a manufacturing situation that would make sense and normality would be plausible. What I hear from the agency is, when I looked at the data, it seems plausible. I think the committee can ask to see some results on that at some point if you want.

DR. LEE: To show you how flexible the schedule is, I am advised we should take a break at this point because the sound system does not appear to be functioning very well. Let us do that. We will convene here about 10 o'clock.

[Break.]

DR. LEE: We are going to start with Dr. Yi Tsong's presentation and then we will come back to discussion. If there is a statistical question, we will catch him before he runs off to the airport.

In Vitro BA and BE Testing

DR. TSONG: Good morning.

[Slide.]

I want to apologize because I have an engagement in Lubbock, Texas this afternoon so after this presentation, I have to run and I cannot stay here for the discussion.

But, Dr. Walter Hauck can probably help the committee out on the questions.

[Slide.]

What I will try to discuss today is the comparative measures for the in vitro profile comparison of the profile measurements. The goal of this talk is that I will try to present the work-in-progress for equivalence approach for the profile measure from the in vitro studies and apply them as the work group came out to some simulated data.

Essentially, the approach we are looking into is pretty straightforward in the concept of how to measure the difference between two profiles. But it is also difficult in the sense that there is no statistical distribution available to be able to, for example, use a t-test, normal distribution, existing distributions to apply to this problem.

That leads to the problem as to, also, how we determine the cutoff point for the equivalence limit. That is why we have to do a lot of simulation to see where we make the cutoff point to make us feel comfortable with where they are supposed to be.

At this stage, the work is still in progress and we would like to have the committee take a look and think about it and give us feedback to see how we can modify it or

whether this is something you think we should keep on working on.

[Slide.]

First, I want to describe the profile. The profile we are talking about here is the particle-size distribution in cascade-impactor equipment. As this equipment measure, actually we separate what the distribution is in this cascade-impactor equipment. We have unit dose and we have distribution of the unit dose among the different states of this cascade impactor.

The unit dose, we put this in as the non-profile comparison. The distribution part, we put in as the profile comparison. So, in this approach here, in the profile-comparison approach, we don't worry about whether the unit dose is the same or not. Everything is standardized to 100 percent as a total. And then we just see how they distribute, whether the attached product and the reference product is the same. Unit dose is a different test to be satisfied.

Equivalence means that we needed to have the profile to be the similar one and also we needed to take into consideration there, because we have variability in this kind of data, versus the variability between the different life stages. Also, we have variability within the lot which is between the canister of the same lot. Then we

have variability between lots.

At the current stage, we put all this variability together. We didn't specifically separate that. There is some difficulty separating them because of the limitation of sample size we have. So, at the current stage, we put all this variability all together.

Number three is that we want to consider the profile measurement in the comparative sense. That means we wanted to look at the test-to-reference profile distance compared to the reference-to-reference profile distance. Sometimes people call that reference-reference variation. Here, it is also coming from the distance measurement of the two profiles.

[Slide.]

I think I have pretty much described this part, some of the non-profile observations and the profile observations. But there are some characteristics of these two types of observation.

Non-profile observation means that each canister gives only one observation which is, in this case, a unit dose. Profile observation means each canister gives one observation of particle-size distribution through the various stages of the cascading factor.

It turns out to be that the non-profile becomes only a univariate but a profile distribution, if you have

ten stages, that means you have ten category variables, ten numbers to deal with. This becomes a multivariate situation. Then we have a non-profile situation which is the univariate, it is easy to talk about mean and variability. In the profile situation, the mean and the variability is difficult to obtain explicitly.

So we are going to talk about average profile or standard-deviation profile, individual profile to the mean profile; we have some difficulty with that issue.

[Slide.]

Also we can see the approach in the sense of aggregated criteria. That means, we put the profile difference and the variability into one simple criteria. We don't want to separate them. We have to do an additional test, in that sense.

Profile distance at canister level, that is what we look at first. Suppose I have one canister from the test and maybe one canister from the reference or maybe two canisters from the reference. We look at how we look the distance between the canister profile at this level. Then we look at the ratio of the profile distance at the canister level. That means, we measure the distance between the test reference to the ratio between the reference at each canister.

Then we are talking about now we have reality. We

have more than one canister. Then we look at how we came from the individual canisters to be able to put together into mean of those ratios and how to calculate the confidence interval of that mean ratio.

From there, we try to set criteria to set how to satisfy the requirement.

[Slide.]

The ratio of profile distance at canister level is really, as I describe, the test and the reference canister distance to the profile distance between two reference canisters. The mean is just the expected value of those ratios of many of those canisters, in that sense.

Here, we don't really have to stick with this notation, but it is probably just easier to see later because we are going to refer to those notations. First, at each stage, we have the proportion of the particle distributed in that stage. So we have the proportion which is pT, standing for the test product. pR stands for the reference product; s means this stage, and the d_s of the tested reference and the d_s of the reference-reference. This means the difference between the test and the reference at this particular stage. We are looking to the distance between the reference-reference at each of the stages.

Then we have cd_s. I use this here for the test-reference and reference-reference. That means the

cumulative of the distance between the test and the reference up to that s stage. So it is just the difference at each of the stages together up to the s stage. Then I use the notation cd_s, which means cumultative.

e_s, for tested reference, is simply just the average of the two test and the reference at this particular stage of those distribution proportions.

[Slide.]

If we just imagine that the cascade distribution comes out with the test stage, we have test and reference product and each one of has a proportion at each of the stages, then, how do we measure the distance between the two profiles?

First, intuitively, you can see that the first thing we want to look at is the difference between the test and the reference at each of the stages. That will give us what I describe as the d of the test reference at the s stage. Then we try to put all of them together to indicate that this is the measurement for the profile.

There are many different ways we can put them together. One is just to take the difference, take the absolute value, add them together, average them. That is what we call the mean absolute distance of the two profiles. And we divide by the total number of stages.

The other one is just to add up all these

distances together into one. If we want to emphasize--if there is a large difference, I don't want to just add up. I want to enhance the difference. So we make the difference between the two at this particular stage and square it, and a large difference becomes larger, enlarged into a composite index of this distance.

So that is when we come out with the mean squared distance. That means to take the distance and square it together. Then you add them together, divided by how many stages. That will give us the mean square of the distance.

As you are familiar with dissolution tests, there is the f2 factor. The f2 factor is really a transformation of the mean score of the distance into a particular formula so we have standardized the f2 between 0 and 100. I think, in dissolution, you know that we use 50 as the cutoff point for f2 factors passing or failure.

Chi square is really taking us a weighted mean-square distance because, as I mentioned, you take the difference between the test and reference at each individual stage, square it. Instead of just squaring it, we also divide it by the average of the test and the reference, so this is a weight which is 1 over this average of the two.

What this indicates is that, for those stages which have a large proportion deposit to, a small difference probably isn't going to be of much importance. But those

which are not supposed to have a large proportion, if there is a large difference, you want to enhance it by using the weight.

So this comes out as the chi-square distance. We can also put additional weight to it if I am more interested in a particular stage than some other stage that you are less interested in. At this stage, we didn't put any particular weight in that sense.

So, in a sense, chi square is a weighted mean of the distance, but statistically it has--but probably, even though it doesn't necessarily apply, those probably do not necessarily apply to our data.

Also, we have been thinking about, in the beginning, an intuitive way is just to, at each stage, you do a t-test. Why don't we just do that. Then we have to do ten different tests and to study all ten tests to satisfy the equivalence.

The second problem is that, because the proportion--supposedly, if the proportion supposedly adds up to 100 percent, so there is a count trend of each one of them to be a total that adds up to 100 percent. An individual comparison doesn't have this feature and also we have this individual stage value as also correlated to each other. It is not totally independent.

[Slide.]

Here I give you an example of how we calculate this. Here is the example of supposing I have only one canister on the test and I have two reference canisters there. Those are the proportions distributed into all the different stages including the standard and the throat. I graded future stages because I think later we will jump to the future.

So, actually, when we do the work, I show the example is really up to stage No. 7. It really doesn't matter that much.

[Slide.]

Graphically, what I want to show is that this is the difference between the test and the reference. As you see, the first bar is the test canister and the next two are the two reference. I think one of the questions we have been asking about is should we separate standard and throat from the other stages because maybe there is not--we are interested in that, too, but we don't have a solution to that as yet.

[Slide.]

Here is an example of the calculation. As you see, the first is the table you have seen already. The second block is that we first took the two references to average them together. I have averaged the proportion of all the stages. Then I have the distance between the test

and the reference at individual stages which is really the difference between the test and the average of the two references.

Then I have the E which is the mean of the two references, and then we calculate the difference square divided by the average which comes out for each individual stage that we have these components. Chi square is just adding up all these components into one value. It becomes 7.25. This is the chi square which extends between the test and the reference.

As I mentioned, this one we don't put a particular weight except the weight by using the reverse average of the two references. So we can do the same for the two reference products and then we have a chi square for the distance of the two references.

Then we calculate the ratio, which 18.83, for the chi-square ratio. As you can see, the dominator of these two products are very much the same, but the numerator can go to as small as zero. So we know that the chi-square ratio should be the smaller the better, in that sense. But it could be as large as possible when they are very different.

[Slide.]

I also give an example to calculate the mean absolute difference and f2 and those ratios. Here, the same

calculation but instead of using--oh; here, we are using the mean absolute difference, just those differences added together so we get 7.288. If I use f2, I use a transformation formula to get to this f2 value for the test and the reference distance.

The same way, I can recalculate those for the reference and f2 for the reference-reference. Then we come out with the MAD ratio which is 6.605. We have and f2 ratio of 0.590.

MAD, which we know is just adding up all the differences. So, the smaller the better. For this one, we know there is a maximum amount which cannot be more than 100 because the maximum difference, theoretically, is 100. f2, 50 percent is no more applicable to this ratio because it is not f2. It is the ratio of the 2f2.

So, here, we know that f2, itself, the larger the better. We know that. So the f2 ratio, we still need to come out the larger the better. The value can be between 0 and it could be as big as infinity. But, mostly, it is going to be between 0 and 1.

[Slide.]

What I have shown you is just a one-canister situation. Now, suppose I have a bunch of canisters of test and reference. To be able to calculate those differences at the canister stage, what we need to do is try to match them

together into a triplet of one test and two references.

So we have to do a random matching to be able to create those and calculate those ratios. So let's take a look at this. Supposed we look at three lots with ten canisters per lot. Why do we come up with 30? I think we had some discussion and it seems this is something workable. It doesn't have to be necessarily the final answer.

So that is why you probably Dr. Walter Hauck uses 30. I use 30. We are consistent, at least in that sense. If we look at this triplet combination of the test and reference-reference, here we need two references to be different. Otherwise, we may have the denominator to be zero.

In that sense, we have a combination of 33 N--that is the total number of canisters of the test--factor times the combination of the three chosen out of the 3N, all this combination, distinct triplets, combination. This number could be very large.

We don't necessarily need to have all of them. So, if the triplet is very large, we take a random sample of this continuous triplet to calculate those ratios. If it is very large, the total combination number is very large, we just took a random sample without replacement.

If it is small, we probably want to take it with replacement to be able to do this work.

[Slide.]

So we have this sample of the combinations. For each of those samples, at the current stage, we are sort of satisfied to use 30 because the original sample size is 30 triplets to make one random-sample sample.

Then, after we have each one of them, each canister, we can calculate the ratio of the chi square or the ratio of the f2 or the ratio of the MAD. Then we calculate the mean out of the 30 triplets we sampled from and have the average of those which is the sample mean of the 30 triplets.

So we have a sample mean. To be able to have the confidence interval, we know we don't have distribution. We don't have parametric assumption and we don't take the asymptotic distribution. So what we should do is going to repeat the steps in this 30 triplet sampling for N times and come out with the distribution of those means, and take the lower-upper 95 percent or 5 percent to be the confidence limit of those means. That is when we come out with the confidence interval of the mean in that sense because, totally, we don't use any distribution to deal with the parameters here.

So, what comparison we need to do for bioequivalence is that we look at the upper limit. For chi square, we know that the smaller, the better. So we want to

control against the largest mean difference you may have. So we use the upper limit of those confidence limits compared to whatever we prespecify the limit.

If it is smaller than the limit, we are going to have equivalence. Now, here, I am puting that the limit is 7.66.

[Slide.]

When we come to the next one, we tried to determine the limit based on the simulation. Here we have one try and then we try different combinations, and 7.66 comes out to be a reasonable cutoff point. Here, I want to show you how we do that.

First, we wanted to simulate 1000 per product, we come out with ten lots at 100 canisters per lot. Then we do the simulation using the real mean and the percentage of CV defined in the simulation.

Here is an example of this. For the no-variation one, the CV is 20 percent and 10 percent as low in stage 1, up to 20 percent. Actually, I think I should have reversed these two. The upper one is high variability. The lower one here is the low variability.

So we simulated this out and used this one to figure out where the cutoff point maybe looks like.

[Slide.]

So, in order to be comparable with what we

proposed, we randomly select three lots from the ten lots and we simulate it out and randomly select ten canisters per lot from the 100 canisters we propose. So we have those combinations of a distinct triplet. And then we repeated this. We sampled 3/30, as we mentioned before, and we repeated this 100 times.

So we calculate the sample mean as we did before and we will be able to calculate the confidence interval by the percentile from this one. That is what we have on the next page, we have the simulation.

[Slide.]

The first one is that we have the tested reference to have the simulated--and we have no difference. But with the test product has high variation between and within, low variation, versus the reference has high variation and low variation. We also did a simulation with 10 percent difference, the simulation between the test and reference and we have a test product high variation, low variation, versus the reference product, high and low variation.

It comes out that we look at the 90th percentile, which comes out with the value 6.66. As you see, the simulation comes out here. This happens when the test product has large variation and the reference product has small variation. Also, we have a difference given at each of the stages.

So this is the point. Otherwise, those two ratios are quite similar. So we have this one can kind of stand out from the rest. That is sort of satisfactory to us.

[Slide.]

We also tried that with a similar one with the f2 ratio which comes out in the case here. We had difficulty to separate them one from the other. So this comes out as smaller, but there is really not that much difference, to tell the difference, in the f2 ratio.

[Slide.]

Then we also simulated for the mean absolute difference which comes out similar. This one has a little bit larger number but still not as clearly as we show in the chi square. There is a large distinction between the large variation versus the small variation of the reference with the mean difference.

So that is what we propose, tentatively propose, to use that point value as the equivalence limit.

[Slide.]

So I can summarize the points that I presented here. We have briefly summarized an equivalence criterion proposed for in vitro profile measures and we propose a criterion for paired test and reference canisters. We propose criterion which considers distribution variations as well as distribution differences.

We propose criterion which penalize increased distribution variability and rewards reduced variability. This work is still in progress with some further considerations we have taken.

I think I can stop here.

DR. LEE: Thank you very much.

DR. TSONG: If there are any questions, you can leave them to the FDA members. They certainly will ask me and you can ask Dr. Walter Hauck. He works closely with me. Lots of questions he will be probably able to answer.

Thank you very much.

DR. LEE: Okay; you can go now.

I understand that the audience was not able to hear what was said in the beginning. Can you hear better now? Good. But we don't have time to go back to the beginning.

I would like to go back to the agenda.

Subcommittee Discussion (Continued)

DR. LEE: We were addressing a very important question on content uniformity. I would like to pose the questions to everyone around the table and your feeling about whether this should be a single content-uniformity standard for all OINDPs.

I would like to go in order. Let's start with the highly distinguished colleague on my side here.

DR. MacGREGOR: As far as the single content-uniformity, I have seen a lot of simulations and a lot of statistical descriptions. However, I haven't seen any data. What has been promised is that there is a possibility to gain some data for this panel to evaluate.

So I would prefer, rather than answer this question now, to see the data. It is my understanding, from looking at all the documents we have been given--we have got a pile about a foot high, here--that there will be data forthcoming both today and over the next couple of months.

So, in my evaluation, I do not see how we can answer this question today. I think we need to see data because a single content uniformity for all the products that are out there sounds like a very idealistic point of view.

Now, if all the data comes in and it does point to a single content-uniformity guideline, then that would be the greatest thing in the world. We would all be on the same page. However, it is my gut feeling, having worked on many of these projects, that every drug is different. Otherwise, we would all be selling the same drug for the same indication.

So I would prefer to table this question until the end of the summer when everyone says that they will have data. I realize that we are under a deadline to try to meet

an advisory committee that is in early fall but I would suggest that we reconvene sometime in the future when we have data rather than--

DR. ANDERSON: This is my first meeting and I certainly am not prepared to make any decision one way or the other. I made some notes as I listened to the discussion this morning and, as a teacher who is not an elementary-school teacher, these questions are probably very elementary.

These are things that I need to understand in order to make an intelligent decision about this question. I have here, one, presumably, and this is in answer to the question should there be a single content-uniformity standard, my statement here is presumably there is not one now. That sounds like one of the answers to my organic-chemistry questions.

Two, and why I ask this question, what is the consequence of not having a single content standard. I would like some information that would help me answer these questions or at least get more information on them. Having not attended the previous meetings, I would rather wait until I have information in these areas.

DR. BAASKE: The committee is dealing with two distinctly different types of devices. We are talking about a metered-dose inhaler or a dry-powder inhaler and a nasal

spray. Without understanding the capability of the device manufacturers, it is hard to draw, across different devices, one standard.

So I would agree that we need to see data before you could make that decision.

DR. AHRENS: I would agree with the comments that were just made and probably not have a lot to add to that. It is very difficult to answer the question in absence of essentially data as to what is out there with currently existing products.

DR. LANGANIERE: I would like to see more data to put in the proper perspective of variability associated with certain products versus other types of products.

DR. DALBY: I am certainly willing to look at the data but I do think that ultimately what matters is a consistent dose to the patient and it doesn't intuitively make a lot of sense to me to tightly control that at a device level if there are other enormous sources of variability.

So unless the data speaks very consistently to one standardized set of criteria, I am more inclined to say that it should be looked at on a product-by-product and drug-by-drug basis.

DR. GORE: I am very much in agreement with the recommendation that we look at more data. I think we also

need to look more at the consequences of how the drug-content uniformity actually plays out. For example, one consideration that we didn't have to discuss and factor into our consideration is that, in development and also in manufacturing, batches are routinely placed on stability for up to two years.

Just a quick back of the envelope says that the minimum number of canisters is somewhere around 400. Depending on how many go into stage 2, you could get up to well over a thousand canisters for those batches. So I think it is a more complicated picture and we just to need some more time to really understand it.

MR. PAREKH: I would pretty much echo what Dr. Gore just mentioned in terms of I think we need to look at the practical implication of these things. Without data in front, I can't seem to be able to comment on, especially because these devices are so different from each other, and how they are therapeutically used.

So I would like to see more data but, also, strong consideration for what is going to be its practical implication in the end and how industry deals with it, in general.

DR. HAUCK: I guess from a conceptual perspective, the notion, as it is sometimes called, of "one size fits all," doesn't make any sense to me. It hasn't in a lot of

these things. So that takes you down a path towards having different criteria for different products.

The problem there, there is a downside to that which is you don't want an absolutely different criterion for every single product either. Some of the stuff, even what Xi Tsong was presenting, was working toward a notion that using the properties of the reference product would essentially help determine criterion. That sort of approach would be one possibility to consider.

DR. HARRISON: I also agree with Walter that the "one size fits all" concept doesn't seem to really make a lot of sense here. What you need is a dataset. It does seem to be an opportunity to have such a dataset available by the end of the summer. I would also like to see us waiting until that point in time to make a more rational position.

DR. DERENDORF: I basically agree with what was said before. I want to make an additional comment, however, and that is that we are looking at a three-level evaluation here or assessment, content uniformity, the in vitro and then the in vivo assessment. I think we kind of look at content uniformity in isolation. It is tied in with the other two and we need to make sure that they all match. We cannot have more stringent, let's say, in vivo requirements than we have in content uniformity.

So they all interact with each other and need to be put in perspective and we need to identify what is rate limiting. I agree with what Richard said in that sense.

DR. SZEFLER: I think everybody is trying to avoid discussing a difficult question, but, as a clinician, we deal with variability in response among patients. Our assumption always is that the product is acceptable and so we deal with thinking about adherence to the medication and biologic response as the other variables.

So, to delay kind of a movement towards standardization and characterization would be unfair to the clinician and to the public. I think we have to move in that direction. Having said that, I would like to know if there is a problem and is there something worth fixing or is it something that we are moving to.

I guess the impression that we are trying to move towards standards in order to characterize a product in order to get some assessment on bioequivalence. So I would say we need to move in that direction while assembling the data, but making it clear we didn't start out that way by giving examples of problems.

That would force us to move there even more quickly. If we are dealing with products, and having done a study recently where we had a recall on a product, it doesn't settle well when you are investing money into doing

studies and then you find out that the product doesn't meet standards for some reason.

So I would say we need to move there. We need to see some examples where there are problems so that we could set the goal posts because I think the problem is not trying to set the standard, it is trying to set the goal posts in terms of what is acceptable and what is not.

DR. BEHL: First of all, I also agree with my other colleagues that the better off we are in drawing conclusions and making a better guidance. But as a sort of more fundamental issue here, the question was should there be a single content-uniformity standard for all nasal and inhaled products.

If you go back to Dr. Poochikian's slide No. 5, he defines them as eight different kinds of products. Then, when we go back to his eighth slide, we only have values or the specifications given for NDI and DPI.

Dr. Poochikian, are you saying that these also apply to all of the six products, the specifications on your eighth slide?

DR. POOCHIKIAN: That was only an example of NDI and DPI, but, currently, those are being applied also to nasal preparations.

DR. BEHL: The same specs apply then? The same specs will apply to all of them?

DR. POOCHIKIAN: Correct. Yes--where are on the market, for example. Some of those still are not on the market, as you know.

DR. BEHL: If you do that, then the question comes up, in multidose container versus a container that contains less than three doses, your beginning, middle and the end estimate of the dose delivered, the question comes how would we address the issue of a DCU for a bidose or a unidose nasal drug product.

We do have a nasal drug product on the market, a unidose. Is the guidance excluding those special cases at this time, unidose, bidose, nasal drug products?

DR. POOCHIKIAN: That was not the intent of the guidance.

DR. BEHL: So if somebody is developing a bidose or a unidose nasal solution or suspension drug product, then that company is not bound by this guidance?

DR. POOCHIKIAN: In that case, through the container lot, that applies only to reservoir-approach drug products. If you have a single dose, you have a single dose; that's it.

DR. LEE: May I request that you focus your questions to the central question.

DR. BEHL: The central question was about the number of units in the beginning, middle and the end. The

next line of comment is that are we mixing things there in terms of validation versus the Q/C test required to release a batch.

Some of these issues, doing first tier, second tier, on the actual dose delivered, either as a whole, or beginning, middle and end, I believe these things are normally done as part of a product process where we have to do a second validation package to show that the dose delivered, or the actual device used, is, in fact, valid and can be used in an efficacious manner precisely each time you use it.

That, to me, sounds like a validation issue. Beyond that, we shouldn't have very few Q/C tests type of testing procedures. There is no sense in doing a validation type of evaluation on each batch that has to be released for commerce.

DR. LEE: So what is your position about the central question?

DR. BEHL: The central question is that these are too restrictive evaluation tests for each batch released because they are more or less of a validation issue than a Q/C release issue.

DR. LI: I want to just concentrate my remarks on the question of whether there should be a single standard for all drug products with the emphasis on the single. I

can see that, theoretically and, even in some respects, practically, there may be some advantages to moving in that direction in so far as there would be a single standard or single set of parameters that would apply to all these products.

Whether that is, in fact, achievable or appropriate, I think we have yet to determine. I think that, as some of these models and some of the data is accumulated, it will become more clear whether or not the single standard is an achievable goal.

For example, if we assume a statistical approach, does it make sense to have a false-positive rate of 5 percent for one product, 6 percent for another and 10 percent for yet another. As we get more information and look at new products and existing products and how they fit into these standards, I think it will become clear.

Perhaps we could answer those questions product-by-product. So there may be some advantages to having a single standard, but I think as we move toward actually getting practical information, it may turn out that there are some severe limitations to that.

If this is the case, we will limit the down side to having multiple standards based on different classes of products, whether it is nasal products, orally inhaled products, different drug moieties like corticosteroids as

one group and another drug product as another, maybe turn out to be the best approach.

I guess I would say a single standard is a reasonable goal and, practically speaking, I think it will become clear whether or not that is achievable, and multiple standards for multiple products adds some complexity. But I think the downside is limited there.

DR. SHUM: Without the data in front, and without the chance to review all the materials, looking at the two questions, to me, it will be difficult for me to make a decision today to answer question No. 1. We obviously need to review what is out there.

And, I am leaning more to saying yes to question No. 2, which is saying that we should look at other approaches. I also want to remind my colleagues here that, as we are looking into the database, we also can see that there might be other approaches that we should consider. Obviously, we had a presentation from Walter about his statistical approach but I also recall that there was a statistical approach presented by my distinct colleague, Dr. Mike Rebe, in the June workshop last year.

So there are other approaches. Of course, there are also other guidelines, ICH Q6A, ICA Q4, all these approaches that we should also consider before we come to a position.

DR. LEE: So the consensus around the table appears to be that we need more data. I just want to know whether or not anybody knows what kind of data they are looking for. It sounds like a question that I pose to my students; we need more data.

DR. AHRENS: It seems to me that the data needs to address two questions, one probably easier to get at than the other. One is, what is history. That is the products that are currently out there, what kind of variation in content uniformity is there. That would clearly be a floor to what a new product would have to match.

You would clearly not want something that was worse than is already out there. But as I think Dr. Poochikian mentioned earlier, essentially the past is prologue. That doesn't mean, with current technology, it isn't possible to do better than that.

The second question which I think is going to be much harder to get data to answer is what is reasonably achievable in terms of improving on what is already out there. I don't know quite how to address that other than from companies who have tried and what kind of success they have had in improving on content-uniformity variability over history.

DR. GORE: I think probably most of us have not had a chance to speed-read the entire package, but there is

a statement and a proposal in the package from the ITFG/IPAC collaboration on Page 8 which actually makes a proposal for data and data collection.

DR. BEHL: I believe if you look at the QC failure rates of different batches of different kinds of products, one could learn from there as to what the problem is and a resolution or a better method for the future.

Second, the justification for asking for all of these tests for each batch is produced; that, to me, is a central question because is that really necessary? Is the cost justified? In regard to the first point, has there been a Q/C program from various batches tested so far?

DR. LEE: I think that we have heard the opinions around the table about the virtue of a single content-uniformity standard. There is consensus that we need more data and I think we have a vague idea of what the data is.

I would also like to acknowledge comments made by Dr. Jim Li about the need for some kind of guidepost and variations thereof.

At this point, I would like to move quickly to the second question which has been covered kind of in tandem with the first one; "Should the FDA continue development of the proposed statistical approach?"

Very quickly, Walter?

DR. HAUCK: I would say yes and get to say "Off

with his head" later, if you don't like it. [Laughter.]

DR. HARRISON: I would also agree. We should definitely continue and I would like to see some datasets.

DR. DERENDORF: I agree.

DR. BEHL: I also agree but, again, one more time, I would like to repeat that we should look at the justification that we asked in question--because if that justification is not there, then the question can become semi-moot.

DR. LI: I like the statistical approach in part because it forces the agency or committees or even clinicians to concentrate on what the important parameters are. Rather than number of canisters being tested, to concentrate on the guidelines for parameters, for uniformity and variability. To me, that has more clinical significance in terms of protecting patients from out-of-spec products.

DR. LEE: In the interest of time, let me ask the rest of the table, is there any difference of opinion?

DR. SHUM: Mr. Chairman, I just want to refer to that question. To me, that question is a broad question. It is not only applied to what Dr. Hauck has presented this morning. I, again, want to urge this committee, when we look at statistical approaches, we should consider there might be other approaches out there that we should consider.

DR. DALBY: I would say, I think it is quite

important to also make sure we educate people about what the statistical approach means because, to me, nine out of ten passing units sounds much better than a 60 percent probability that the units fall within an acceptable spec. They are both based on the same data so I think it is important not to frighten people with that information.

MR. PAREKH: The only comment I would like to make is that I think the statistical approach makes sense when you are developing the products. To the extent that level of testing that is required to comply with that kind of process controlling the quality of the product, it is very impractical.

So I agree with the statistical approach. How far we can take it, I am not sure at this stage.

DR. LEE: Anybody else? Guirag, do you have enough information to work on?

DR. POOCHIKIAN: Unless there are specific questions that I can enlighten about.

DR. LEE: Any questions? If not, I think that closes the first session of this meeting. We are not done yet, because I would like to move on to the next session. The next session is on bioavailability and bioequivalence.

Dr. Adams, are you ready?

Bioavailability (BA) and Bioequivalence (BE)

Current FDA BA/BE Background and Issues

DR. ADAMS: Good morning, ladies and gentlemen.

[Slide.]

My topic this morning is orally inhaled and nasal drug products for local action, current FDA BA/Be background and issues.

Before starting, I would like to thank the members of the subcommittee and invited guests for participating in this meeting and also to recognize the amount of work that has been done by Nancy Chamberlin and the advisors and consultant staff and also from David Morely and Jim Corey in OPS. This is represented a lot of work in putting this program together.

The talk on BA/BE background and issues, the issues have already been delineated by Dr. Eric Sheinin in the BA/BE questions which he has gone over earlier this morning, so I will talk about background here.

[Slide.]

I would like to start with showing you the Technology Committee, the OINDP Technology Committee, that has been involved in developing two guidances, primarily the nasal BA/BE guidance which has been on the FDA's Internet site since June of last year, and indicate that there are seven working groups that have been involved in that.

Many of the individuals listed there are in the room today.

[Slide.]

The two guidances at issue are both product-quality guidances. One is the BA/BE Studies for Nasal Aerosols and Nasal Sprays for Local Action. The second one, which is in preparation, is a BA/BE Studies for Orally Inhaled MDIs and DPIs and Inhalation Solutions, also for Local Action.

[Slide.]

These draft guidances cover BA and BE. But, on the BA/BE side, they cover only product quality BA which refers to release of drug from the drug product, but, rather, it does not cover additional bioavailability studies which are required by the divisions; that is pharmacokinetic and bio studies in addition to those studies indicated in these guidances.

Of course, bioequivalence is a product-quality issue only. Furthermore, these guidances are strictly limited to locally acting drug products.

[Slide.]

We know that, according to the CFT, the approaches to measure BA and establish BE are pharmacokinetic, pharmacodynamic and clinical, in that order, preferably pharmacokinetic. If that is not appropriate, then pharmacodynamic studies. If they are not appropriate, then clinical studies.

In addition, BA and BE may be established based upon in vitro or in vitro plus in vivo studies.

[Slide.]

The challenge for locally acting drug products is that these products do not require systemic distribution in order to reach sites of action. Consequently, pharmacokinetic studies, in general, are not appropriate for documentation of BA and BE.

[Slide.]

When we talk about the locally acting drug products we have, then, to concern ourselves with both local delivery, which relates to efficacy, and, because these drugs are absorbed into the systemic circulation although, generally, it is not wanted, we have to concern ourselves also with systemic exposure.

[Slide.]

The recommendations for bioequivalence that appear in our nasal BA/BE guidance pertain to formulation equivalence, recommendations that the inactive ingredients be qualitatively the same as those in the reference-listed drug, and that at excipients be quantitatively the same; that is, within plus-or-minus 5 percent of the concentration in the reference-listed drug.

Furthermore, that the devices be functionally comparable. That is because these drugs are, as we all

know, combinations of formulation and the device.

[Slide.]

Regardless of whether in vivo studies are needed, we always ask for in vitro data for BA and BE whether it be a metered-dose inhaler or a dry-powder inhaler or nasal sprays. We are considering confidence intervals for comparative data for selected of the in vitro bioequivalence measures.

As has been indicated by Dr. Tsong this morning, those statistics are under development.

[Slide.]

For the metered-dose inhalers and nasal sprays, the draft guidance lists six tests that we feel are appropriate for characterizing products; that is, dose or spray-content uniformity through container life, droplet-size distribution, drug particle-size distribution, spray pattern and plume geometry, priming and repriming and tail off.

Those six tests are to be provided in the BA and BE portions of the submissions in addition to information in the CMC jackets.

[Slide.]

On the in vitro BE side, statistical comparisons under development are the profile comparisons for the cascade impactor data. Dr. Tsong has talked about the f2

and the chi-square approach. The nasal BA/BE guidance refers only to the chi-square statistic.

But we recognize that there are other possible approaches and we are going to be hearing from Dr. Andy Clark in the next presentation concerning a different approach to profile comparison.

Then, for the non-profile comparisons, we have recommended those for dose content uniformity for container life and certain other in vitro tests as indicated in table No. 1 of our draft guidance. It is based upon a population bioequivalence criterion.

[Slide.]

The proposed bioequivalence criterion for content uniformity requests that the mean performance of the test and the reference products be determined, the variability of the reference products and the variability of the test product, within and between batches be determined. The criterion is based upon differences between test and reference means, differences between test and reference variances, and then scaling of the bioequivalence boundaries to the referenced listed drug variance.

It uses the one-sided, 95 percent upper confidence bound with an alpha of 0.05.

[Slide.]

This is the equation. This is the proposed

equation for population bioequivalence. We simply put in equation form the information from the prior slide showing shown differences in means, differences in variance and then, in the denominator, scaling to the reference product variance.

[Slide.]

Turning from in vitro to in vivo BA/BE, there are concerns about local delivery based upon a clinical study, systemic exposure based upon pharmacokinetic study or systemic absorption based upon PD or clinical study.

Bullets No. 2 and 3 are simply a definitional issue where we are saying systemic exposure is defined as pharmacokinetics and systemic absorption is defined as either PD or clinical.

For nasal-solution formulations, we are requesting, for product quality, BA/BE in vitro data only.

[Slide.]

For nasal sprays, our draft guidance proposes three different types of clinical studies to establish efficacy. It proposes only one of those three studies would be needed, however, not all three. And they are the traditional two-week treatment study, a days-in-the-park study, or an environmental exposure-unit study.

I should indicate that these slides were prepared by my colleague, Dr. Gur Jai Pal Singh with a presentation

he gave recently.

[Slide.]

Now, BE studies for nasal sprays; in addition to the efficacy side of things, there is also the systemic-exposure side of things. For that, we recommend, if possible, that a pharmacokinetic study be used to establish bioequivalence. We recognize that, for some drugs, the systemic exposure may be so low that it may not be possible to measure the drug in the plasma. If that is the case, then we are recommending a pharmacodynamic study.

[Slide.]

Turning from nasal products to inhalation aerosols and, specifically, albuterol MDI, pharmacodynamic endpoints; our present thinking is that pharmacodynamics based either upon bronchodilitation or bronchoprovocation maybe used to document bioequivalence and, in fact, the Office of Generic Drugs has approved generic albuterol MDIs based upon both of those endpoints.

[Slide.]

Our current recommendations for the randomized crossover design for the pharmacodynamic study for albuterol MDI are that, in addition to baseline data, that one puff and two puffs of the test product, one puff and two puffs of the reference product be included in the study design as a minimum although, in order to better define the dose-

response curve, one, two and three puffs of test and one, two and three puffs of reference would be preferred.

[Slide.]

In addition to the efficacy type of study, there are the concerns about systemic exposure of inhalation aerosol products and, for albuterol MDI, we recommend a randomized, two-way crossover study. This is conducted generally in healthy volunteers and the study could be a PK study. We would prefer that, although the current products which we have approved have been based upon comparative pharmacodynamic endpoints for albuterol MDIs.

[Slide.]

And then data analysis for the clinical bioequivalence studies; that data analysis is study-design dependent. For rhinitis studies, those are categorical endpoints and, consequently, the appropriate statistics must be used for those. For pharmacodynamic studies, we have adopted a dose-scale analysis which I won't take the time to go into at this time. For systemic-exposure studies, we use the pharmacokinetics. We use the conventional two one-sided tests procedure.

Thank you.

DR. LEE: Andy Clark is going to talk to us about an alternative view profile analysis.

Profile Analysis of Cascade Impactor Data:

an Alternative View

DR. CLARK: Good morning.

[Slide.]

First of all, I would like to thank the committee for inviting me to come on this morning and talk and, particularly, to Dr. Adams for giving me the job of giving an alternative view on how we should look at profile analysis on the cascade impactor.

[Slide.]

I guess where I would like to start is a little explanation about background. There are three main reasons I can think of that you would want to compare impactor distributions and make some sort of measure of similarity or dissimilarity.

The top two, I guess, are the two we were talking about this morning, releasing batches or bioequivalence between a new product and an innovator. The bottom one is up here mainly because this is where this piece of work and, I guess, along with a lot of other work we have had this morning, this piece of work is still a work in progress.

But this is really where it started, an interest in trying to figure out how good a radiolabel has to be on the product to be able to match the product well enough to tell you what it is doing in the clinic if you measure deposition profiles. I think that the idea behind this one

applies as well.

I guess the question I want to ask is whether simple statistical distance or a measure with some physical significance is actually really needed when comparing these impactor distributions. To be honest, having heard Dr. Tsong this morning, I am not sure simple statistical difference is the right terminology.

[Slide.]

I guess the question is we all know we have got to measure size distributions because we all believe they are physically significant in terms of determining the dose of aerosol that gets to the site of action within the airways.

[Slide.]

What this chart is trying to point out is the approach, so far, appears to be to use this simple distance measure. So this is the distance between a reference distribution and a test distribution. In this particular case, I have chosen log normal distributions, tests 3 microns with a GSD of 3, reference is 3, et cetera.

But the object of the exercise is to measure these distances and either f2 or chi square is really a function of this distance or some of these distances between the two distributions.

The problem I see in taking that approach is if you look at the significance of these distances, it depends

where you are on the size distribution curve whether that particular significance in determining the dose that reaches the lung or not.

For example, the top end, here--this is the top stage of an Anderson cascade impactor--a difference here, in terms of its implication for change in dose reaching the airways, is really pretty small. I apologize for getting this the wrong way around. The 1.2 should be up here and 9 should be here.

It is pretty pivotal. But, if you take this distance, 12 percent difference in the distribution, that 12 percent difference has to be normalized as to how it affects the deposition that reaches the lung. In this particular case, at this particular size, that difference is really pretty small, zero, if you are looking at alveolar deposition.

It doesn't matter what happens up here. All the aerosol is deposited in the upper airways. If you want to go down to a smaller size, around 1 micron, 0.9 or 0.8, in terms of fraction, would be deposited in the airways so, a change here could bring about a major change in lung dose.

So I guess what I am arguing for here is that you have to understand the physical significance of where the change is taking place in the size distribution, not just sum the statistical differences regardless of where it is in

an analysis.

[Slide.]

So what I said about doing this is to try to take a look at f2 and chi square and see if they measure the differences in distribution in any way that is relevant to how a product might perform.

Rather than trying to use real data, the model was pretty simple; log-normal distribution for the reference aerosols, log-normal distribution for the test aerosols and the two variables here are either a change in MMAD, which is the blue line here, parallel so the GSD is the same but the MMAD is smaller, or a change in GSD, which is the width of the distribution--i.e., same MMAD, different angle on here, meaning a different width in terms of distribution.

[Slide.]

If you take a look at f2, which Dr. Tsong defined earlier, and see how that responds to those changes in size distribution, what you get is a nice inverted, almost triangular, function. As you go from the test aerosol, which in this case was 2, move away, either to a courser MMAD or a finer MMAD, f2 decreases.

Typically, for the dissolution-type testing, you take an f2 equals 50. So, for this particular aerosol with a GSD of 2, you can get anywhere between 1.2 and about 2.7 microns, which would be judged by an f2-50 criteria as

being similar.

You could also move the GSD axis. Remember, the test aerosol here was 2 MMAD, 2 GSD. You get the same sort of function in response to the change in the width of the size distribution and, again, an f2 of 50 will give you as GDS of anywhere between about 1.5 and in excess of 3.0, in this particular case, as being similar.

Of course, that is varying two variables independently. You can, obviously, put them all together and build a response surface like this to change in size distribution.

[Slide.]

What we have got here; this is for a test aerosol, again, of 2 microns, MMAD, 2 in terms of GSD. This is varying the GSD. This is varying the MMAD. This is how f2-50 responds. So what you get if you look at a set of log-normal distributions relative to a reference and you slide this three-dimensional picture here at f2 equals 50, is you get an ellipse.

Any size distribution that is inside this ellipse would be judged by an f2-50 criteria as being similar. So the easier way to look at that is actually just project it down onto the bottom axis of GSD and MMAD.

[Slide.]

So what I have tried to do here is put together

five different reference aerosols. So we have got 1, 2, 4, 6 and 8 microns MMADs as the references. They all have a GSD of 2 in this particular plot. And then we see how they respond to changes in either in MMAD or the GSD.

You can see what you get is an ellipse. Anything inside this ellipse, according to this f2-50 criterion, would be judged as similar. One of the problems you get is, for a 1-micron reference distribution, the distance from here to here if we don't vary the GSD, is about 0.7 of a micron. So, an f2-50 criterion would allow you to take an 1-micron aerosol and somewhere around 0.7 microns MMAD would be judged as similar to somewhere around 1.3 microns.

You will see, in a minute, that doesn't make a lot of difference in terms of deposition in the dose that the would reach a patient's lungs. However, if you go up to the courser aerosols, the situation starts to become a little different. 4 microns, if it was in the middle here, would mean that you could get up to somewhere around 1.3 times 4, so somewhere around 6 or 7 microns at the top end and somewhere around 3 microns at the bottom end.

Now, a 3-micron to a 7-micron difference in terms of the aerosol that is deposited in the lungs makes a big difference in dose, as you will see.

Those of you who are confused as to why the ends are flat here, what you are seeing is a limit in the

resolution of the Anderson cascade impactor. This was run as a simulation on an Anderson. If the aerosol gets too big, the f2-50 flattens at the top because all the aerosol is on the top stage.

If the aerosol gets too small, it flattens at the bottom because all the aerosol is in the bottom stage. And then, of course, the f2-50 does not respond because you are looking at no change in sort of seven or eight of the stages and only a big number on one of the them.

[Slide.]

You can plot the same thing for an MLI, which is the other instrument that I have done here. Again, you get the flat ends. They are slightly different because of the way the cascade impactor--the range of sizes that it analyzed. But the difference here is still pretty much the same in terms of what an f2-50 would allow as a pass in terms of a similar aerosol.

Again, at a small size, this difference is not too big in terms of the difference it makes in terms of lung dose. At a larger size, up at around 4 microns, maybe 6 microns, this difference would be substantial in terms of the dose that would actually reach the lungs.

[Slide.]

So that is a rough idea of how f2 responds to size-distribution changes. This is chi square, which is the

other alternative that Dr. Tsong talked about this morning. You will notice the shape is different. It is not an inverted cone. It is much more of a sort of flat mushroom hat.

But, in essence, the ellipses are pretty much the same in terms of if you set a particular value of chi square here to either pass or fail, you would have an ellipse when projected down onto this MMAD GSD axis, which says anything inside the ellipse would pass.

The question is are those response surfaces for those particular statisticals at all similar or relatable to a response surface in terms of how you change the dose that actually gets into a patient's lungs.

[Slide.]

The answer is they are not, but we will go through this chart first. The reason they are not is because it actually matters whether the aerosol is a course aerosol how much change you can allow for a specific change in dose into the lungs or whether it is a fine aerosol.

Typically, what I have tried to do here is choose the 1 micron that we got off the previous slides. f2-50 would say we can go from about 0.8 microns here to about 1.3. The change in dose, and I accept this is a lung-deposition model. I don't believe it is directly applicable in terms of absolute number, but I certainly believe you get

doses that are proportional to these sorts of numbers.

But the change in deposition is really pretty small. It is on the order of 4 percent change down at 1 micron. If you take the same f2 criteria and apply it to an aerosol up at 4 microns, you could end up with a change in lung deposition of somewhere around 150 percent depending on whether you are up at this 6.5-micron end or down at this 3-micron end.

So f2 and chi square, actually neither of them respond in a way that is relevant to the physical situation of what goes on with those aerosols when they are inhaled and deposited in the lungs of a patient.

[Slide.]

Just to try and fill you in again with a three-dimensional plot, this is for a 2 micron, 2 GSD aerosol. All the changes in lung deposition here are actually plotted as negative. In reality, what happens, of course, is the aerosol goes this way, the change is positive. But it is just easier to look at this surface.

So out here, GSD of 2.8, 2.9, MMAD of about 1.2. There is a 28 percent difference compared to the deposition we would get in the lung from this 2/2 micron reference. So the shape really here is sort of a saddle shape.

You will notice the shape for the f2 response surface and for the chi square is much more of a cone or an

upside-down mushroom. So they don't match very well.

[Slide.]

If you do the projection down onto this GSD/MMAD axis again, this is the typical f2-50 plot for a 4-micron and an 8-micron aerosol in this case. This is a 10 percent change in lung deposition. The inside one is an 8-micron aerosol. The outside one is for a 4-micron aerosol. If I was to do a 1-micron aerosol on here, the line would probably be here and up here somewhere.

So, not only do they not have the same response surfaces, but if you try to measure a change here, bounded by an f2 number--and this is just particularly f2-50--you get a channel here where you get significant changes in lung dose that gets to the patient, but you get areas outside by this f2 criterion where you would have a substantial change in lung deposition but the f2-50 would say you have got the same aerosol.

One of these major problems is that it doesn't know whether you are dealing with a fine aerosol or whether you are dealing with a course aerosol. It is merely just the sum of statistical differences.

[Slide.]

The way I propose, and this bit is a real work in progress--the only way that I could think of, having got through that primary analysis to try and correct that

situation, I use the term "weighted" very, very differently from Dr. Tsong's weight in his chi square a little earlier on--was to actually try and weight the importance of the amount of material in each stage.

So, for example, this is a column of deposition weights. It is merely calculated from a lung-deposition model and you will see, in a minute, this is one of the limitations. I think it would probably take us another five years to agree on these weighting factors, but the throat and stage 1, of course, have a very low weighting factor because they contribute very, very little to that part of the distribution that is important in getting into the lungs and affecting an efficacious dose.

Stages at the bottom of the impactor have a much higher weight because there the size fractions stand a high probability of getting in through the mouth and the upper airways and depositing in the lung and, hence, constituting part of an efficacious dose.

Really, all I have done here is taken the median sizes off the stages for an Anderson, calculated some weighting factors based on a pretty simple lung-deposition model, taken the weights--this was for a log-normal distribution on the Anderson plates--and then just multiplied the two together to get a weighted distribution.

I think, at this point, there is a variety of