interesting because we had both genders represented in the test population.

So let's break the problem down to its core elements and look at the drug substance. Well, the drug substance on the surface is fairly simple. Here's the solubility pH profile. It has high solubility at all pH's up to about pH 5 or 6, and that's the PK of the drug. Then the solubility really plummets.

But the key point is even down here, the solubility of this drug was high enough so that the dose could go into solution no matter where it was in the gastrointestinal tract. So there was no issue of solubility.

We looked at the excipients and it was clear that the right-limiting step in absorption was going to be in vivo dissolution. That's the purpose of modified release. We know the excipients were the way that the release rate was controlled.

The mechanisms for each product was different--that is, the release mechanisms--by virtue of the excipients and manufacturing. And the excipient effects in product A were pH-sensitive.

We also look at the formulation and, in particular, in the dissolution, which is the rate-controlling step in the bioavailability for the product.

Picture products A and B going into the gastrointestinal tract. The first thing they hit is an acidic environment within the stomach. And you can see that the dissolution of product A and B--one is indicated in the white circles; the other is in the blue circles, but they're superimposable. So in the upper region of the gastrointestinal tract, namely the stomach, no differences.

When you go down into the upper GI, the duodenum, the early part of the jejunum, the pH changes in a fasting state to 4.5 or higher. And here you can see, I think--I can't see it too well from my angle but you can see that there are differences in dissolution between the products.

Product B is rapidly dissolved in the upper GI. Product A isn't. It has a slow dissolution. It sort of plateaus out and then eventually it continues to be dissolved. This is the excipient differences between these products.

Now picture that dosage form or dose forms moving down into the lower GI. They get into the jejunum, down into the colon area. The pH now rises to 6.8 at that site and you can see these products can be differentiated in terms of their release at 6.8.

This product here, again relatively slow till it gets down into that lower part of the GI and then it increases. The other one has a little different profile.

Much of this drug is probably released in the upper GI and then it plateaus out. So these clear differences in formulation could be demonstrated in vitro.

Now the key to this case, I think, was the gender differences between males and females, and I would say it isn't an unequivocal situation because we have an absence of some information but I think we can put together a reasonable explanation for what we've observed here.

We focussed on the physiological variables and wondered how they might interact with the dissolution properties of these formulations. We realized that many of the physiological variables are under genetic or environmental control. They're highly variable. In the literature one can find subpopulation differences in the distribution of, say, gastric pH, stomach emptying.

Unfortunately, a lot of those papers are somewhat contradictory but there are some differences and we just need more information to sort them all out. But those differences could be anywhere in the physiological variables that I listed here.

What we tended to focus on though, however, were variables that the data was a little bit stronger in the literature in terms of gender differences and that was the intestinal metabolism by CYP 3A4 and the PGP transport processes, keeping in mind that this drug substance was a

substrate for both of those processes.

And here's how the analysis sort of went forward. We took the CYP 3A4 and said okay, what's going on here? We have large intersubject variability and substrate clearances. We know that without a lot of debate. We also know that intrasubject variability is less than intersubject, which happens to be about 30-fold variability in the population. Again that suggests that genetic factors have an overriding presence.

And we also know that the content and expression of CYP 3A4 is not only site-dependent but it's saturable, so as you move down the tract, you get lower content and lower activity from the duodenum on down to the ileum.

Well, can there be gender differences in bioavailability related to this? I think the answer is yes. We know for a fact that oral clearance of drugs like the one I showed you, drug X, is less in females than it is in males. We also know that the first pass effect for these types of drugs is less in females. And we also know that the bioavailability is larger in females for drugs that have the characteristics of this one.

What we don't know specifically is what the mechanism is, but one could speculate that there is less CYP 3A4 as one moves down the tract or perhaps less 3A4 in male than females in general, or perhaps there's some gender-

related homeostatic mechanism that is influencing the activity and content of that enzyme.

Now the second process in the intestinal tract is intestinal PGP. We have limited data on this efflux process. There are some gender differences. And what we know about it is the opposite of CYP 3A4. There's not a decreasing gradient but an increasing gradient in content and activity as one moves from the proximal to the distal gut.

Data show that it's fairly easily saturable. Because it's saturable, one can have dose-dependent effective permeabilities, so something could move from a low to a high permeability status. And we also know that the activity in males is greater than females for the PGP transport.

Now given all of those facts, when you put them together, this is the bottom line of the story. I call it the mechanistic hypothesis of this particular example. We have observed the subject-by-formulation interaction with product B, and why is it? It's because that product had slower dissolution at pH 4.5 in the upper GI, so a greater fraction of its dose was going to be available to the lower GI for absorption.

When it gets down into the lower regions of the gut, there's faster and more complete dissolution at pH 6.8.

We're talking about the jejunum and ileum.

So we think a larger fraction of the dose of this delivery system was released in the ileum than the other product. And with the facts that we have about lower CYP 3A4 activity at that site, we know it's readily saturable. We also know that females have less of a PGP efflux versus males.

We concluded that the greater absorption with product B and the higher Cmax and area under curve is a function of the concentration or the greater percent of dose released at the site and the longer residence time that these dosage forms would enjoy within the intestinal tract itself.

Well, that wasn't enough. We looked for some supportive evidence. We looked at, for example, in this study the metabolite-to-parent area under curve ratio for product B. If more drug was getting past the first pass effect, we'd expect a lower metabolite-to-plasma ratio in terms of area under curve and indeed, we did find that in 10 of 13 females at a lower ratio.

It was consistent because only two of 12 males had the lower ratio, so again it was a signal of the mechanism being rational.

The other thing is we looked for confirmatory evidence and in this case, for this product, we had a

multiple-dose study and we observed a similar subject-by-formulation interaction in the multiple-dose study. So the single-dose study wasn't an artifact. We saw the same thing--higher area under curve and Cmax for product B in females and, in fact, the ratios were the same, pretty much the same--1.33 female to male, 1.54 female to male.

And then there was a body of literature that when we looked at 3A4 substrates and gender effects, we consistently find lower oral clearances in females for this type of drug substance.

Well, from this exercise what we concluded is that we wish we had more data in the database to understand the mechanistic basis of the subject-by-formulation interactions. What we'd like to do is gain experience with replicate BE study designs in subject subgroups to provide some data. And these types of studies, if they're designed along the lines that Dr. Benet presented with the appropriate subgroups, will provide us that sort of evidence.

And then what we intend to do with that information is to take a stepwise analysis similar to what I just demonstrated and dissect the problem on the basis of drug excipient formulation subjects factors and from that, hopefully come up with information that gives insight and predictability in advance to the possibility of subject-by-

formulation interactions.

I mentioned this is the effort of a small working group that works in association with the IBE working group. These are the members of the subworking group. They put in a terrific effort, meeting every Monday at 7:00 for about two hours. Thanks.

DR. BYRN: Questions for Larry from the committee?

[No response.]

DR. BYRN: Okay, the next speaker, Roger will discuss replicate and nonreplicate datasets.

REPLICATE AND NONREPLICATE DATASETS

DR. WILLIAMS: This part of the presentation will cover some real data, if you will, that we have within agency files and that we also have from the published literature that I think tries to deal with Les's statement: Is this a theoretical solution to a theoretical problem?

And in that regard, you have also heard some other statements, I think primarily from Bill Barr, as to evidence in the marketplace and in the literature about subject-by-formulation interactions.

Now the FDA has received a series of replicated bioequivalence studies over the years done primarily because sponsors thought they would be useful in one way or another, and this is a list of those studies that indicate the drug product, the drug substance, the dataset number and the type

of drug that is in the dataset.

We're going to some trouble to preserve trade secret information here, so I'm not going to really talk about the drug substance or the drug product. This will probably be the most you see about it except for some other introductory slides. And you can see down at the bottom we had a very generous transmission of replicated datasets that I won't say anything at all about in terms of the drug substance or a drug product.

Now this is summary information about that dataset. There are actually 31 separate drug products. Thirty-four were single drug products; seven were combination drug products. The study population was healthy males and females. The number of subjects in the studies ranged from 19 to 67 for FDA datasets and 12 to 74 in that generous donation, if you will, that we received.

Many things were analyzed, both parent and metabolites, but we focussed only on the parent drug in the subsequent analyses, which I will discuss with you. The reason for that is that that's what we generally recommend to document bioavailability-bioequivalence, particularly bioequivalence, and also because there's a confounding or if you see something with a metabolite, it might be connected with the observation for the parent drug.

And we focussed on AUC 0 to T and Cmax as our

bioavailability measures. And if you count, when all is said and done, in terms of number of parent datasets we have, the total is 55--34 in our files, 21 from the industry transmission.

Now I'm going to show you some summary information about the findings in these datasets and I'd like to show you some individual cases that we think are interesting and speak to the function of the aggregate criteria.

Now let's look at the percentage of the 55 datasets that showed an important, a possibly important subject-by-formulation interaction.

Now in the criterion guidance, the advisory committee will see a statement that we believe subject-by-formulation interactions could be important if they're greater than .15. That means roughly that 15 percent of the subjects in a bioequivalence study who are above that number would not be switchable, according to our current understanding.

Now if I just show you the percentage that exhibited a subject-by-formulation interaction greater than .15, it's 20 percent for AUC and 33 percent for Cmax.

There's an adjustment going on over here at the right that takes into account the within-subject variability of the reference. The expert panel has pointed out to us that the higher that variability it is, the more likely it

is to see a subject-by-formulation interaction. That was a very important comment we took into account and we adjusted that .15 number based on the within-subject variance of the reference.

Now with that adjustment, and Dr. Hauck can speak to that adjustment if there are any questions--Walter can speak to that--you can see the numbers do drop to 13 percent and 20 percent for AUC and Cmax.

Now that's suggesting that the evidence to say that it's not just a theoretical solution to a theoretical problem goes down when you adjust for within-subject variance of the reference.

Let's look at these numbers down here. These are more general statements about when is the reference greater than .2 in terms of within-subject variability. And you may recall from the criterion guidance that that is the point at which scaling will start taking place. And you can see for AUC it's 46 percent; for Cmax it's 73 percent.

And that suggests that frequently for Cmax we're dealing with highly variable drugs, at least drugs where we would suggest that you should start scaling to give the test the benefit of scaling.

Over here on this side of the thing you've got a comparison of the within-subject variance of the test and reference. In about half the cases the reference is greater

than test and obviously test is then greater than reference I about half. But for the reference-test comparison for Cmax, you see it drops to 40 percent.

Now we did, on these datasets, compute both average and individual bioequivalence using the two one-sided T test approach, as well as the proposed new criterion, and these are some of the numbers. Passed both approaches, 78 for AUC, 62 for Cmax. Passed IBE, failed ABE, 3.6, 9.1. Failed and passed, 12.7, 18.2. Failed and failed, and you can read the numbers there because I can't quite read them.

Now if you look at the seven and 10 failures for AUC and Cmax for individual bioequivalence, the apparent reason for the failures are listed here: subject-by-formulation three and six. Within-subject variability was higher for the test and there are those numbers, I believe, three and one. And then it looked like the studies were underpowered: one and one.

So that gives you some understanding as to why, for either AUC or Cmax, the test failed using the proposed individual bioequivalence criterion.

This seems a little out of order, Kimberly; could you hold it back? I want to get to those graphics of--hold that back, too. There are a whole bunch of slides in there that show the curves. I'm sorry. Yes.

Now you're all experts at learning how to interpret the data based on the numbers and the graphics, so let me see if I can give you some examples. And these are real datasets based on the replicate datasets, where we now look at within-subject variance, the variance of interest here, where it's a simple case for AUC, I believe. I'm having trouble reading these, obviously.

Now you can look at the numbers up here and you should be getting to be experts now in terms of scanning this row of numbers but let me start with it graphically because I think that's where the message is. You can see here the dispersion about the mode for the bioavailability measure of interest for the test is much less than it is for the reference. You can also see a little bit of scaling here. And the final conclusion is that it passed individual bioequivalence and I believe it also passed average. If I'm reading these wrong, tell me because I can't quite see them.

Now this is an example of the aggregate criterion at work, where you have a reward for reduction in variance to the test, you have a little bit of scaling and there's no subject-by-formulation interaction.

Let's go on to this example. Here you can see immediately that the variance of the test is quite large. We would say that the manufacturer did not produce a good product. There's no scaling and there's no subject-by-

formulation interaction and I believe it failed individual bioequivalence and passed average.

Now that's a very interesting example, where we would argue that the criterion is working to achieve the public health objective of having less variable products, whereas the average criterion does not. Let me go on.

This is an example of combined effects for within-subject variance and mean of comparisons. You can see here that the means are about 12 percent on for the test, that the variance of the test is much lower. There is no subject-by-formulation interaction and there is a little bit of scaling. And the outcome here was fail and pass, I believe. Am I reading it right, Rabby? Oh, average fails and individual passes.

Now here's an advantage. If I were speaking to the producer you could see an advantage to the producer in terms of producer risk where individual bioequivalence is helping--the criterion is helping you pass and probably the main reason for that happening is two factors. One is a lot of scaling going on. That 212 number is being driven wider by the performance of the reference. From 137 to 212, that's quite a drive. And you also get a reward for reduction in variability.

Again you see the public health motivations of the performance of the criterion. Let's go on.

Now this story is combined effects of within-subject variance and scaling.

Oh, by the way, I might back up to that prior slide. When we talk about mean variance trade-off, that's an example of it. The means are a little off. Means are off but you get the reward from reduction in variance. So that's an example of mean variance trade-off. Let's go on.

The next one is combined effects of within-subject variance and scaling. Here you see a lot of scaling going on because your variance of the reference is quite wide. It's driving your goalposts very wide. Reduction in variance to the test is obvious. No subject-by-formulation interaction. It passes individual and it also passes average.

Now this is an interesting example that perhaps from a consumer risk standpoint we could be concerned about. There's actually quite a reduction in the mean. You can see the mean of the test is about 75 percent of the reference.

There is a significant subject-by-formulation interaction. That's the blue line. There is no reduction in variance of the test relative to the reference. And there is substantial scaling going on. We're looking at Cmax for this drug. And you can see that it passes individual but fails average.

Now that's an interesting example for my consumer

risk. If we did take into account this new criterion, we would let products into the marketplace that do this.

Now this may be where the rubber meets the road, at least in terms of a regulatory agency, but I think you can see from a producer standpoint that the aggregate criterion is working to allow this product into the marketplace.

And I might say to the committee that one of to questions in the series of topics for discussion will be to allow the use of the aggregate criterion, to allow market access. So you now know what you'll be allowing if you recommend that.

The next one is combined effects of within-subject variance and subject-by-formulation interaction for Cmax for a specific drug and what you see here in terms of graphics is the variance of the test is about the same as the reference. There is a fairly substantial subject-by-formulation interaction and there is some but not a lot of scaling going on. The end result is it fails individual bioequivalence and it passes, just barely, average.

Now this is an example where even with a subject-by-formulation interaction, which we say we care about, the criterion in the aggregate works to allow market access. And probably that occurs by this--I'm sorry. I take it back. The criterion works to impede market access, even

though scaling has occurred based on the variability of the reference.

So I would argue from a public health standpoint it fails there in a way that we would say is good. We are trying to impede products that exhibit subject-by-formulation interactions.

Now I believe this is my last slide in this series and the advisory committee has all the graphics for all the datasets in your backgrounder, so please look at them if you wish over the lunchtime. You can get a sense of how the criterion in the aggregate is performing.

This refers back to a question I believe Arthur asked about downward scaling. The criterion will both scale wider and scale narrower, depending on intersubject variability of the reference.

Arthur, you made the point that you think it's just luck that variability of the reference for narrow therapeutic range drugs is low. We actually don't think it's so much a matter of luck perhaps, but the fact that a highly variable drug would have problems in the marketplace if it were a narrow therapeutic range. And we can certainly talk about that in the course of the discussion. It's a very interesting question. And Les, of course, I think has commented on that perhaps publicly in many ways.

Now let me just show you what's going on here.

The reference here did show substantially low intrasubject variability of the reference. It drives the goalposts for AUC down to 1.02. That's about as low as you can get. And for Cmax it drives the goalpost down to 1.14. For AUC the variability of test reference is about unity, so that's about the same. There is no subject-by-formulation interaction. And if you look at Cmax, the variability of the test and reference are about the same and there is no subject-by-formulation interaction.

So you see the effect of the aggregate criterion can narrow the goalposts here for both AUC and Cmax. And it passes for AUC but fails for Cmax according to the individual bioequivalence criterion, whereas with average it passes both.

Now this is a public health motivation of the criterion, which is for NPR drugs, you would allow the performance of the reference to drive the goalposts always, and we would not allow the Epsilon term.

So this is truly starting, I believe, from 1.25. Is that not right? Okay.

So you see here actually something that would make it more difficult for people to get into the marketplace with an NTR drug and that we would argue would provide a greater assurance of switchability for these drugs that we say we care about more.

Now I want to also now talk about a completely different dataset. I'll be fairly brief about this dataset because it refers to nonreplicate bioequivalence studies that are in our data files. You can imagine that we see a lot of these in the course of a year. This dataset was compiled by staff in the Office of Generic Drugs under the leadership of Dr. Patnaik.

Two hundred and fifty-six datasets came in in 1998; 90 different drug products. They were mostly healthy subjects, some male and female but mostly healthy males. This reflects what we're currently doing now.

The sample size ranged from 17 to 78 and these are the test reference ratios. I'm wondering a little bit about that .75, Dale. How did we let that through? But anyway, you can explain that to me later on. Let's go on to the results.

And the essence of these numbers are on this slide because we have a sense that the ANOVA root mean squared error is a measure of the possibility of a subject-by-formulation interaction. So that if this number is greater than about 1, all we can say about these datasets is that it doesn't exclude the possibility of an important subject-by-formulation interaction. If it's less than 1, we would argue, or if it's less than maybe even about .15, we would argue that the possibility of a subject-by-formulation

interaction is not likely.

Now what are we seeing here? If we take that number, we would say we couldn't exclude from these datasets in a fairly high proportion the possibility of a subject-by-formulation interaction. And it goes down, of course, when you raise the number a little bit, but still the proportion is fairly substantial for both AUC and Cmax. Obviously a very limited look at a dataset but it attempts to look at our nonreplicated datasets in terms of whether they could exclude subject-by-formulation interaction as a likelihood, and these data suggest that for the most part, we couldn't.

This is a set of data compiled by Dr. Chen that looks at PK studies, bioequivalence studies in 26 instances where males and females were included in the study. In some ways this is a more general look compared to the dataset that Larry talked about. All different kinds of dose forms, some single dose and a few multiple dose studies.

Now what do the data show? There was a greater than 20 percent difference in the ratio of geometric means, 35 percent when you took into account both AUC and Cmax. And if you look at datasets versus studies, that number came down a little bit.

If you look at statistical significance, the numbers go down a little bit, so this looks more at the size difference, as opposed to the statistical significance

difference. But I think our bottom line here is that in looking at these datasets, the possibility of a gender-by-formulation interaction appeared to exist in a fairly substantial number. And again Larry talked about a very specific example of this.

I might mention the fact that we see both genders in studies. Probably it relates to our 1993 gender guideline, which encouraged inclusion of men and women in studies unless there was some reason for exclusion.

This is a report from the literature for Verapamil. Many of you know this report. It was a bioequivalence study for two generic products compared to the pioneer, multiple-dose study, eight young, eight hypertensive elderly. The dose was 80 twice a day and the data analysis was, I believe, average bioequivalence.

And if you look at generic 1 versus generic 2, in this one you don't see any evidence of an age-by-formulation interaction. In this one you clearly do between the elderly and young for this particular generic product for all parameters observed, fairly substantial ones.

All the datasets that I have shown so far were numerical observations without any associated clinical findings to suggest that there was a clinical impact of the subject-by-formulation interaction. This is probably our only example where we actually have that and it came to us

in 1998 for Methylphenidate.

We have a coordinating committee called the Therapeutic Inequivalence Action Coordinating Committee that was put in place after Hatch-Waxman to receive reports of therapeutic failure in the marketplace and we began to see reports for the test product here that it was causing trouble in the marketplace. We actually pulled the product out of the marketplace and compared it with the pioneer and what we saw in vivo was more variability than the test, more rapid absorption. It was bioequivalence based on average, but dissolution suggested that it was significantly faster for the test.

Now we actually did do a replicate study on this particular formulation, working with Dr. Myer in Tennessee. Let me see if I can read it from here. I just can't read it very well.

Again you should be used to looking at these numbers. If you look at the test relative to the reference for Cmax you'll see a substantial increase in the variance. So we see an example where the test product is more variable. There is a little bit of a subject-by-formulation interaction in Cmax, .143. The product failed for Cmax when you used individual bioequivalence. For AUC, slightly different results; the variance was increased a little bit but less so. There was no subject-by-formulation

interaction and it did pass according to individual bioequivalence. And recall I mentioned that it did pass average bioequivalence.

So we believe this is an example from the marketplace where there were clinical correlates to the observation and the new criteria would have failed that for Cmax and passed it for AUC, whereas the average would have passed both.

This is the dataset that Larry showed. We think it's a very interesting dataset and I think Larry's team has done a terrific job of analyzing the dataset. What I would argue is that in females they actually showed bioinequivalence, and the reason we were able to document this is because of the '93 gender guideline that encouraged the inclusion of women in bioequivalence studies.

Now if I summarize all the evidence to date, I would say it looks something like this. Replicate study designs. Those are the numbers where we think there's subject-by-formulation interaction. And I would argue I have to recall that these data were performed in healthy subjects. Our supposition is that if you did it in patients or people more representative of the general population, these numbers would go up because you will not see a subject-by-formulation interaction in healthies, or at least you'll tend not to see it.

Nonreplicate study designs. Those are the numbers I alluded to. Gender-by-formulation interactions for both AUC and Cmax, 35 percent, and there was the one very dramatic example for calcium-channel blockers.

Miscellaneous studies, we have Verapamil and Methylphenidate. Mechanistic studies, we have the FDA study in progress and you heard Larry and Ajaz allude to that dataset with sorbitol and sucrose.

And I believe that's my last overhead. Steve, thank you.

DR. BYRN: Questions for Roger? Questions for clarification? Arthur?

DR. GOLDBERG: Roger, on the Methylphenidate where the Cmax ratio is 1.48, that would have passed average BE, with a ratio of Cmax of 1.48?

DR. WILLIAMS: It wasn't 1.48, was it, Arthur?

DR. GOLDBERG: I thought it was.

DR. WILLIAMS: No, that was the variability comparison. That wasn't the caparison of means. Can somebody read the comparison of means? Oh, I didn't give it to you?

I'm sorry, Arthur, we didn't give it but I think they were within plus or minus 20 percent. And they had to be to pass average.

DR. BYRN: Other questions?

[No response.]

DR. BYRN: I think we'll go ahead because we have a full agenda, so we'll go ahead with Vinod Shah's presentation on the general BA/BE guidance for orally administered drugs.

GENERAL BA/BE GUIDANCE ORALLY ADMINISTERED DRUGS

DR. SHAH: Thank you, Steve, and good afternoon everyone. I was warned in the morning before I just came in that my presentation should be ending before noon but I guess I'm starting, so I'll try to go very fast.

As you know, everyone has been talking about individual bioequivalence, the replicate study designs and all that, but how can they put that into practice? There has to be a way. There has to be a mechanism. So these studies at least could be requested from the sponsors, and that's done by using our general guidance for the industry, which is for the bioavailability and bioequivalence studies for orally administered drug products, the general considerations.

This guidance has been posted on the Internet on August 27 and the notice of availability was made available in the Federal Register in September. So the guidance is now out on the Internet as a draft guidance.

This slides provides an overview of all the contents and the table of contents in the guidance. It

starts talking about the background information, the general bioavailability-bioequivalence, methods to document bioavailability-bioequivalence, comparison of the studies and the different types of the dosage forms and the special topics.

Dr. Williams showed a slide in the morning and indicated that we always need to ask three major questions, which is again attributed to Professor Sheiner. I think the guidance also focusses more or less in the same manner, asking what is the question and the question is with respect to the bioavailability and the bioequivalence, what are we willing to rely upon, and that is being addressed in the methods to document bioavailability-bioequivalence and the pharmakinetic studies and different types of the individual studies.

And how confident we need to be, that is addressed in the measures in bioequivalence studies, which talks about the bioequivalence limits, intervals and confidence.

This guidance is intended to provide a how-to information for the bioavailability and the bioequivalence studies to meet the requirements set forth in 21 CFR. It also discusses the biopharmaceutics aspects of the drug product quality; that is, the release of the drug substance from the drug product into the systemic circulation.

The guidance also provides the choice of the

criteria for analysis, which includes the average, individual and population bioequivalence. And it uses the concepts of early, peak and total exposure in the evaluation criteria.

With respect to the replicate study designs, the guidance indicates that the replicate study designs are recommended for pivotal bioequivalence studies for a two-year period using the pharmakinetic measures.

These are the cases where we do not recommend you replicate study design; namely, for the products which contain the drugs with the long half-long, long half-life meaning greater than 96 hours; in case where a steady-state study is needed; and also in case where the excessive blood samples are drawn and that may have a safety hazard.

Therefore in these three cases we do not recommend the use of replicate study design but in all the other cases where a pivotal bioequivalence study is used, it is recommended that a two-year study period would be involved after the guidance is finalized.

The bioequivalence criteria just explain how exactly what we mean by the study, the replicate study design, whether it's going to be an additional burden or what. I'll just give an example here, that when you use the individual bioequivalence criteria using the replicate study design, we are recommending to use 2 times 2 times 12

subjects, totalling 48 treatments. With the average bioequivalence right now you are using 2 times 24, which again ends up in totalling 48 treatments.

So again this gives an indication and shows that no additional burden is encountered when you undertake this particular study. Again this assumes that there is no subject-by-formulation interaction. And as it was discussed earlier by Dr. Williams and Dr. Chen, you can use this study and power it to calculate for the average bioequivalence.

Also the intent of our guidance is to reduce the regulatory burden while maintaining sound scientific principles, which is consistent with the public health policy objectives.

Just to give you some examples as to where we are reducing the regulatory burden or reducing the regulatory requirements are the biowaivers for the lower strengths of the modified release dosage forms. Modified release means either the delayed release dosage forms or the extended release dosage forms. Until now or at present, we are requiring a bioequivalence study for each and every strength of the modified release dosage form but this guidance suggests there is no need to do that. You can just do the higher strength bioequivalence study in a replicate design, and that should be enough.

And this is in addition to the biowaivers, which

we already grant for the lower strengths of the immediate release products, as well as the extended release beaded capsules.

As you heard earlier from Professor Benet, we are also suggesting in this guidance the elimination of the multiple dose bioequivalence studies for the modified release dosage forms. Again this seems to be consistent with the opinion of the expert panel.

We are also suggesting a biowaiver for a higher strength of the immediate release dosage forms and also the reduced emphasis on measurements of the metabolites in the bioequivalence studies.

And I think this concludes my brief overview of the general BA-BE guidance. Thank you.

DR. BYRN: Questions for Vinod?

[No response.]

DR. BYRN: Okay, why don't we take a lunch break until 1:15. So we'll reassemble at 1:15 for the open public hearing.

[Whereupon, at 12:15 p.m., the meeting adjourned for lunch, to reconvene at 1:15 p.m. the same day.]

A F T E R N O O N S E S S I O N

[1:30 p.m.]

DR. BYRN: Okay, I think we can start. I apologize for the late return of some of the committee members.

OPEN PUBLIC HEARING

DR. BYRN: We have, as you have on your agenda, we have a list of presenters. Each presenter will be allowed 10 minutes and the first speaker is Dr. Steve Schachter from the Epilepsy Foundation.

DR. SCHACHTER: Thank you very much, Mr. Chairman, and good afternoon, distinguished committee members, ladies and gentlemen.

I would like to first briefly introduce myself. My name is Steven Schachter and I'm here today on behalf of the Epilepsy Foundation. I also serve on their board of directors and am the chairman of their professional advisory board. I'm a neurologist who specializes in epilepsy in Boston at the Beth-Israel Deaconess Medical Center and am an associate professor of neurology at the Harvard Medical School.

But above all, today I'm here as an advocate for my own patients with epilepsy, approximately 1,500 who are currently under my care.

In addition to these perspectives, I've also been

the principal investigator on over 60 trials of new anti-epileptic drugs and devices and have long admired and supported the FDA and their advisory boards for their roles in regulating the testing, approval and use of seizure therapies.

For the next eight or nine minutes I would like to focus on a subgroup of patients with epilepsy whose health and well-being are dependent to a great extent on their seizure medications and for whom relatively minor fluctuations in serum concentrations could have devastating social as well as medical consequences.

For these particular patients, it is critical that we distinguish the difference between bioequivalence and clinical equivalence with regard to the medications and generic counterparts.

First, a very brief overview of epilepsy. This is a condition that affects over 2 million people in the United States. Approximately 180,000 people develop epilepsy each year and by the age of 75 the prevalence is 3 percent.

The foundation recently determined that the estimated annual cost of epilepsy is $12.5 billion. Of this figure, only 14 percent is from direct medical costs, such as the cost of medication. The balance, over $10 billion, are indirect costs that are due in part to things such as seizures, medication side effects and lost productivity.

A single seizure can have serious ramifications on employment, driving privileges, social interactions. It can also result in serious injury from broken bones to burns to even death.

Now not all people with epilepsy are the same. For the fortunate majority, seizure control is easy to obtain and for this group, varying serum concentrations of seizure drugs would have relatively little effect on their seizure frequency.

However, there is another group of patients, relatively small compared to the first group, for whom seizure control and avoidance of side effects occurs within a much narrower range of serum concentrations. And in my opinion, the range that their blood levels must be maintained is narrower than the range defined as bioequivalent. This characteristic is typical for the patients I see in my epilepsy referral practice in Boston and these are the patients that generate the anecdotal reports of seizure breakthrough or side effects that appear in the literature and that we recognize as clinicians on a day-to-day basis.

The Epilepsy Foundation has taken the position that prior expressed permission of the treating physician and the patient be obtained before one formulation of an anti-seizure medication is switched to another. I would

like to emphasize that as an organization and personally, we are neither pro-brand nor pro-generic; we are pro-choice.

The foundation's view, however, is often at odds with those of insurance companies, formulary committees and state legislative bodies. These groups often make the assumption that the FDA's definition of bioequivalence means that two bioequivalent drugs are clinically equivalent; that is, completely interchangeable without any clinical consequence for any and every patient.

As you know, there are many different seizure medications. The three frontline medications--that is, carbamazepine, phenyoin and volproic acid--are available both as brand name and as generics, and each is classified as a narrow therapeutic index drug.

I would like to focus on the potential economic impact of therapeutic nonequivalence for just a moment. These costs may outweigh the potential savings and costs from generic substitutions. I would like to give you a real-life example from a patient we saw several months ago. He had been seizure-free for years on brand name phenytoin; that is, Dilantin, and with the availability of the Milan version generic, he was switched by his pharmacist from the brand name to the generic without notifying either the patient or the prescribing physician.

Within a couple of days, the patient was admitted

to the hospital with life-threatening seizures and the total bill for his hospitalization was nearly $4,000.

Now the monthly difference in cost between the brand and the generic, according to the pharmacy, was $4. So in other words, it would take over 83 years to recoup the cost of the hospitalization with the less expensive product. Or put another way, a savings of $4 in direct cost was offset by over $4,000 in indirect costs in this particular case.

Now how frequently does this happen? Admittedly we don't have well controlled scientific studies. This is one of the problems in this area. But anecdotally, it appears to happen quite often. In fact, a survey conducted by the professional advisory board of our Epilepsy Foundation documented the frequency with which this occurs. I presented those results to the FDA's Office of Generic Drugs earlier this year.

In summary, the foundation, like the FDA, is committed to enhancing patient safety, avoiding unnecessary medical and social costs, and increasing the safe and effective utilization of generic medications. To this end, I strongly recommend that the committee members urge the FDA to promote scientifically conducted studies to investigate whether there are patients with epilepsy for whom bioequivalence does not necessarily translate to clinical

equivalence. The results of such well controlled investigations will be most helpful in shaping future policies on the interchangeability of anti-convulsants and their generic counterparts. Thank you.

DR. BYRN: Questions for clarification?

[No response.]

DR. BYRN: Thank you very much.

The next speaker is Nevine Zariffa representing PhRMA, Smith Kline Beecham. I apologize for my pronunciations ahead of time.

MS. ZARIFFA: Good afternoon. My name is Nevine Zariffa and I'm here speaking on behalf of PhRMA. The title slide just indicates that Smith Kline Beecham actually pays my paycheck.

On behalf of PhRMA, we do appreciate the opportunity and the invitation that Roger issued to us to come and address the advisory committee.

In terms of an outline, I want to tell you a little bit about the PhRMA expert panel, its formation and mission, its membership, tell you about the position paper that we have crafted at PhRMA and we'll go through the objectives of that paper, as well as very briefly the approval process and then, of course, spend the bulk of the time on the PhRMA recommendations, which are both specific as well as general.

You've already heard from Les about the Blue Ribbon Expert Panel founded in 1998. Now on the Blue Ribbon Panel you have three separate reps from PhRMA--one from the biostats area--that's myself--one from clinical pharmacology, one from drug metabolism, and we were all involved with Les's panel. And it's fair to say that there was certainly at least one occasion where we offered disparate views.

So in order to rectify that, the PhRMA expert panel comprising representatives from the relevant subsections was formed a little later on in January '99 and our mission really was to derive the PhRMA consensus view on the FDA guidance of December '97, investigate alternatives to the proposed methods, draft an expert report for PhRMA that would outline our consensus view. And, of course, one thing missing here is to put it forward for public dissemination, which is part of what we're doing today.

You can see roughly 12 people on the panel from nine different PhRMA member companies.

In terms of the position paper itself, it's split out into four sections. The first is a review of average bioequivalence, its properties and limitations. Then we go into an expose, if you will, of the proposed population and individual bioequivalence criteria from FDA, along with its properties and limitations. We go through point by point

for each limitation that we raise and we offer some recommendation as to how this might be further studied. And then we do have a section on general recommendations.

I'll slip the first three bullets. I'm sure you can read those for yourself. The last two bullets I'll draw your attention to.

The manuscript was cleared through PhRMA itself on August 26 and it was accepted for publication in the Journal of Clinical Pharmacology on the 27th of August.

Now you can see there's a slight issue here we some dates because on the 27th of August the FDA issued its updated revised draft guidance. So PhRMA will be issuing an addendum to comment on some of the other points that have been raised in the newer version of the draft guidance.

I'm going to skip the section on the properties of average bioequivalence and its limitations. I think other speakers have done that and it's probably not a good use of time. I'm going to go right into the properties of the proposed population and individual bioequivalence criteria.

The first point that PhRMA would make is that the clinical relevance of a subject-by-formulation interaction has not been demonstrated. And to date, no association between clinical failure and subject-by-formulation interaction has been demonstrated.

Now we've heard this from other speakers but let

me reiterate. A consequence of the aggregate criteria is that there are a number of numerical trade-offs that can occur between the various terms. The allowable difference between test and reference means in particular is very sensitive to differences between variances, permitting large rewards or penalty, and these differences between variances are likely because estimates of variances in this type of trial can tend to be quite variable.

Now a simple observation. The proposed criteria does not mandate hierarchical testing. We don't first look at means, then variances, then interaction in terms of Sigma squared D. So we don't have any kind of natural nesting order of individual bioequivalence demonstrating population, in turn demonstrating average.

Another point to be made, while IBE seeks to ensure switchability between test and reference products, it does nothing to ensure switchability between two test products--generic to generic switching--which is, in fact, expected to occur in practice.

And last, PhRMA would like to point out that the lack of global harmonization on the subject of bioequivalence for at least a transition period would place burden on sponsors and regulators involved in worldwide submissions.

So let me make a few general comments on behalf of

PhRMA and then I'll go back to each of these points and outline what our recommendations are in terms of studying them.

We believe at PhRMA that the new criteria should be transparent to regulators, prescribing physicians, pharmacists and patients and provide a demonstrable improvement over the current criteria either in terms of the overall performance or simply in the handling of extreme cases, such as narrow therapeutic index drugs or drugs with high variability. And in PhRMA's opinion, the proposed criteria for assessing population and individual bioequivalence do not represent a significant improvement, at least in any demonstrable clinical or public health sense.

Another general comment. Population and individual bioequivalence do address some of the limitations of average bioequivalence but also introduce new limitations which could, in turn, present undesirable characteristics beyond those observed with average bioequivalence.

Going back to the specific points that we raised in terms of the limitations of the proposal for population and individual bioequivalence, the clinical relevance of Sigma squared D and its use as a surrogate for switchability could be studied by a targeted clinical pharmacology trial constructed to provide the best evidence of Sigma squared D.

And then in terms of what had been at least at one point statistical issues in the estimation procedures, these could obviously be studied through the use of simulation techniques.

The trade-offs between parameters, the scaling and this maximum allowable difference could all be addressed through the use of an ordered testing procedure where you would look, say, at means first, then variances, then something to look at switchability.

Now the quantification of the generic-to-generic switching paradigm certainly can be addressed through suitable simulation studies and this has already been done and published for average bioequivalence, so we could do the same under individual bioequivalence. FDA and PhRMA should continue to engage in dialogue with other regulatory agencies and solicit their involvement in any proposed change to deal with this worldwide harmonization.

Now going back to some of the general points that have been batted around at least for the past few years or so, PhRMA believes that while the population and individual bioequivalence criteria proposed by the FDA carried a number of statistical flaws, we believe that these are minor in comparison to other issues outlined above and certainly would be resolved through focus effort and research and I think we've seen that. That was me speaking, not PhRMA.

PhRMA proposes that the current standard of average bioequivalence should continue as the basis for market access until another method is scientifically demonstrated to better serve the public interest.

PhRMA believes that the trial or phase-in period should be replaced by simulation studies. In our view, the regulatory guidance should reflect a set of current practices and not a set of proposed studies to validate the guidance itself.

Now moving on to something a bit more concrete that hasn't been discussed yet today, PhRMA proposes that there may be other, more effective ways of addressing the public health concerns without the burdens of the complexity design and analysis of the proposed criteria, and one such methodology is going to be described by our next speaker, Dr. Larry Gould.

We propose that an evaluation of Gould's method and the FDA proposed criteria be undertaken and we would work with FDA to identify the standards of evaluation, which is, of course, a key point.

I'd like to make a separate comment on scaling. The concept of scaling is appealing and PhRMA is committed to exploring the applicability in performance of any method utilizing it.

And last I leave you with this point. Examining

the performance of the proposed population and individual criteria and its alternative is something that PhRMA and FDA certainly can and should do cooperatively.

Thank you for your attention.

DR. BYRN: Questions for clarification?

DR. GOLDBERG: A fast question. Is JCP a reviewed journal?

MS. ZARIFFA: Yes, it is.

DR. GOLDBERG: And it was submitted by PhRMA on the 26th of August and accepted for publication on the 27th of August? Did I get those dates right?

MS. ZARIFFA: No. Actually, I skipped the first three bullets. We submitted the draft manuscript earlier in the month of August, and it did get expedited, review-through.

DR. BYRN: Okay, the next speaker is Dr. Lawrence Gould, senior director from Merck.

DR. GOULD: While the transparencies are getting ready to be projected I should like to thank the committee for the opportunity to address them and present a few comments on an alternative approach to assessing individual and population bioequivalence.

If one backs off a bit and considers carefully what bioavailabilities are involved, what the statistical issues are in the evaluation of bioavailability, it seems to

me you have to consider the fact that you're getting a measurement on each of two formulations from each subject, and these two observations are not independent. For one thing, they're made on the same subject.

So it's handy, I think, to consider this in a number of ways. This sort of picture of the joint distribution is handy for expressing a number of concepts. One of them is that there are different kinds of bioequivalence and the average bioequivalence simply means that the centers of the two distributions on the test and reference line up.

Population bioequivalence would mean that the distributions essentially superimpose one over the other, and individual bioequivalence means that large differences between the subjects' responses to the formulations are unlikely and certain repeated exposure of the subject to the formulations would be unlikely, would imply that the formulations are switchable.

Now the correlation that is involved here is the correlation between the subjects' effects, the true effect of the subject, to the test and reference formulations. If the responses are not highly correlated with each other then you have what amounts to high subject-by-formulation interaction. Knowing what a subject's true response is to the reference doesn't tell you very much about what his true

response is to the test.

If you have high correlation--this was low correlation--if you have high correlation, one predicts the other fairly well.

Now here are some scenarios that one might encounter. This is the ideal one right here. The reference and test distributions just about line up. And I'll talk for the moment about population bioequivalence, although individual bioequivalence is the key issue, but the scenarios are important.

In this particular situation you have two distributions. They're not even average bioequivalent. Their means are very widely spaced.

Here is situation that I find kind of problematic because here's the reference, which presumably has done all of the work in establishing efficacy and safety; here we now come with a test. Now I realize that this is an exaggeration of some of the material that's been presented earlier but the exaggeration is here to make a point.

The test here has a mean fairly far displaced from the mean for the reference but it is so tight in terms of its bioavailability, so little variability, that it would succeed in terms of, let us say, an aggregate criterion. So one might say very well, we could say the first test certainly was a population bioequivalent to the reference.

But now the point was we might say well, if that's the case, perhaps we ought to consider this test formulation, this new test as the reference because that has got good bioavailability, very tight, very predictable bioavailability. Now you would find that you really couldn't say that the reference was bioavailable, prescribable relative to the test. It wasn't bioequivalent.

You could, of course, keep your reference, and now maybe another test comes along and it's just as nice in terms of its spread properties as the first but it's displaced to the other side. Now again you would, by the usual criteria, decide well, this is certainly population bioequivalent to the reference and we can go ahead and market it.

However, it certainly is not population bioequivalent to these, to the first test. So one might say that this is probably, from the standpoint of prescribability, a highly questionable situation. And if this applied, as well, to the total observations you got on each subject, which would include the within-patient variability, within-subject variability, you might wonder whether that would be safe to switch these two test formulations. This could be two generic formulations that have been evaluated relative to a given reference.

So it would seem in principle--this is not a law

of nature--it seems reasonable to suppose that one might like to avoid asymmetric decision scenarios of this kind.

Now I want to point out here--I'll go through this one fairly quickly; you've seen this before--I'm going to start with the same standard random effects model that's used in the guidance. There's no difference. It's the same assumptions. I'm not simplifying anything. I'm not making anything more complicated. It's exactly the same.

The subject-by-formulation interaction here, Sigma squared D, is simply the variance between the effects of a particular subject to T and R, irrespective of measurement error, and that is this business right here. This is not an exercise in algebra. I know it looks like it is.

Now the FDA, again by review, the FDA population and individual bioequivalence criteria are based roughly on expectations of squares of the test minus reference bioavailability differences. It's a little more complicated than that but that's basically the principle.

And as a consequence of this, which is, by the way, a perfectly reasonable way to start out; there's nothing wrong with it, you combine the mean bioavailability difference in the variance components and you get these particular expressions for population, individual and average bioequivalence as criteria. I've written these in a slightly different way than the FDA has because I thought

perhaps it might be a little bit simpler and it certainly takes up less real estate on the slide. Lambda is a constant or scaling factor which could be either a constant or the within-subject variability.

Now this particular approach to evaluating individual bioequivalence certainly requires three-or four-period designs.

This does, however, raise some issues. The question is is this a justifiable regulatory burden? Is it necessary to be quite so precise for most drugs? I'm haunted by Dr. Gretter's comments this morning that differences in compliance or lack thereof probably has far more significant an impact on what one sees in terms of bioavailability of a drug for a patient than variability of absorption or metabolism.

Prescribability and switchability are intuitively sensible in principle but there's no published evidence of clinical problems from substituting formulations that are truly average but not population or individual bioequivalent. And I know this to be true because I went and I looked very hard for these in the literature in Medline over the past 20 years and found none.

The point here, the big take-away message is that the FDA criteria are an approach to evaluating individual bioequivalence--perfectly reasonable approach. But the key

point is it's not the only approach that one might take for this purpose.

Now as an alternative approach one might require as a principle that if you have individual bioequivalence--if you're switchable, you ought to be prescribable. So individual population ought to imply population bioequivalence. And if you're prescribable, it ought to be, on average, bioequivalent. So you would avoid scenarios of this sort where you had situations where you could demonstrate individual bioequivalence but not average bioequivalence.

This is a matter of a principle that one imposes on the picture. It's not an essential feature of it but it does seem to be reasonable to require this.

Back to the picture. If one looks at this picture from a statistical point of view, it turns out, and I will spare you the grubby mathematical details, that individual and population bioequivalence can be evaluated using standard regression and correlation calculations on data from 2 by 2 cross-over designs. The statistical properties of these estimators are well known in the normal case, and nonparametric and robust analogs exist. In effect, not only do the methods work in the normal case but they're actually flexible and fairly robust. Variations exist that are flexible and fairly robust.

Now just to walk through what's involved with the calculations, you simply take the sum of each observations on the test, sum of subjects' observations on the reference. And I say the sum of this because you, in fact, could apply this to four-period design, so it's not restricted to the application of two-period designs only.

The correlation between the observations on test and reference provide an intuitive measure of individual bioequivalence.

Now what this sample correlation coefficient consistently estimates is the true correlation between the true effects of test and reference of the reference formulations for the subject, but attenuated possibly by a factor that depends upon the relative variability.

What this really means is that if you have a large within-subject variability--in other words, if you were to administer reference formulations with subject and then administer it again and then again, you would find a very large degree of variability. Then, in a sense, you ought to be penalized for trying to determine those individual bioequivalents because if the observations you made on the reference formulation didn't really well predict the subsequent observations on the reference, it's not entirely clear to me what individual bioequivalence means in that context. And, in fact, this is related to the subject-by-

formulation interaction because here is the Sigma squared D, here is the row. It's just that we have something a little more complicated.

I'm going to do this quickly. Basically the slope of the regression of the sum of the measurements on the difference consistently estimates the difference between the variances. That's essentially what population bioequivalence is about. And, in fact, having applied this through a number of circumstances, the conclusions appear to be close in most cases to the FDA method.

Key points. Population and individual bioequivalence are intuitively appealing concepts, although there doesn't seem to be any evidence that these are needed for the evaluation of most drugs. They can be evaluated in various ways. That's really the key point.

The guidance proposal has some statistical appeal but there are some issues that need to be resolved. And, in fact, the bottom line is it is possible to assess population and individual bioequivalence for all practical purposes using data from conventional 2 by 2 cross-overs with giving results that, at least in the applications that I've tested them with, are consistent with the findings from the guidance. Thank you very much.

DR. BYRN: Questions?

[No response.]

DR. BYRN: Okay, thanks very much.

The next speaker is Dr. Michael Spino from Apotex.

DR. SPINO: Thank you, Mr. Chairman, and thank you, Dr. Williams, for suggesting that we present to the advisory committee. I'm doing this as chairman of the Scientific Advisory Committee for the International Generic Pharmaceutical Alliance.

You have a hand-out of this presentation and it's a little bit different than what I'm presenting right now because I've deleted some slides and I've added some elements from this morning's discussion to more directly address some of the issues that were raised.

IGPA is comprised of generic associations in Europe, the U.S. and Canada, and the conclusions of this position paper were presented at the meeting three weeks ago in Montreal, the IBE workshop.

The first conclusion that we arrived at was that the scientific and clinical basis for implementing a new system for the regulatory assessment of bioequivalence, employing the approach of individual and population bioequivalence, has not been demonstrated.

The view that a subject-by-formulation interaction, that these are substantial and they're prevalent and constitute a regulatory concern is not supported by published scientific data. I feel like I'm in

an echo chamber here, restating what has been stated repeatedly.

But the Levothyroxine issue from this morning, I submit, is not a convincing piece of evidence regarding subject-by-formulation interactions. Rather, the way I read the data it's more convincing to me that we have a highly variable dissolution of the brand product, resulting in variable influence on the thyroid-stimulating hormone, not a subject-by-formulation interaction. The blood levels of the drugs were constant.

Based on the similarity and the release characteristics for the majority of products, and that is immediate release products, that demonstrate average bioequivalence, there's little scientific rationale to expect important subject-by-formulation interactions for studies conducted under conditions of average bioequivalence.

This is not to say that there are no conditions where you cannot demonstrate that various factors alter the absorption of drugs; they do. But in those conditions where you have average bioequivalence and you do not have a modified release dosage form, there isn't even a scientific rationale, in my opinion, to consider the existence of such an interaction.

The modified release products and the data in

particular, presented, I thought, very elegantly by Dr. Lesko this morning, merit further exploration. But recall this was an immediate release--this was not an immediate release; it was a modified release product under special conditions in which they found some sort of group effect. I think it merits further exploration. In a study's period, to explore in greater depth such a potential phenomenon, I believe, is worthwhile and it's consistent with what the expert panel suggested.

And I think this is particularly important because, in fact, I presented to the last advisory committee meeting here some data that we had published a few years ago in which we did a comparison of Verapamil sustained release, modified release preparation in which there were 18 subjects. Nine subjects were tested twice on the brand and once on the generic and nine the other way around.

And what we found was that if you just took the sustained release Verapamil product and did a comparison for bioequivalence of itself on the two separate days, the product failed. In fact, it failed quite miserably as the mean difference for the AUC is about 25 percent, something in that neighborhood.

Now here is one example and there were several--here is one example of a subject given that reference product on two different occasions, and that's a

serum concentration time profile. What we heard from the group this morning, that would be a subject-by-formulation interaction. Well, it's the same product out of the same bottle on two different occasions.

The current understanding suggests to the IGPA that an observed Sigma D greater than .15 might not represent a true subject-by-formulation interaction. And since there are many factors, and I was pleased to see Dr. Williams comment that the variability of the reference product seems to be correlated with the detection or the potential observation of subject-by-formulation interactions, there may be others, as well, and we need to explore this further.

If there were true subject-by-formulation interactions detected under the conditions of average bioequivalence, we have no idea of how large these would need to be to have any clinical significance whatsoever. Therefore whatever that number is, whatever that number greater than .15 is, we don't know what it would be and we don't know what the clinical relevance would be.

Newly proposed modifications of the methodology, such as have come out in the second iteration of the draft, need to be assessed by scientists in academic and industry before their possible adoption. And I say this because the original wave was found to have a number of matters that

needed to be addressed and I submit to you that probably the current iteration also needs some substantive tuning before any implementation could go into place.

It is unnecessary to perform replicate design studies with drugs exhibiting low residual variation in two-period investigations of bioequivalence, and this was addressed earlier.

An interim experimental period for regulatory submissions requiring a replicate design for all bioequivalence studies is unwarranted, in the opinion of IGPA, based on the current level of evidence. Such a directive would be disruptive to the industry and add a further financial burden and time delay that would not be offset by benefit of possible discoveries.

And I want to reiterate this is not an issue of a concern of finances. If the scientific merit and the clinical merit are there, we would support it. We believe they are not.

However, an interim experimental period might be reasonable for the regulatory submission of certain--not all--bioequivalence studies with replicate designs if the selection of the products were limited to those few that were considered to have a scientific rationale--not a fishing expedition--for subject-by-formulation interactions or if they were at the discretion of the sponsor. And I

note that this is completely consistent with the recommendations of the expert panel this morning.

Any interim experimental period, even if only voluntary, must not be considered until there are clear statements regarding the purpose of the experiment, the study design, how the data will be analyzed and how the data will be used. That is if we are going to embark upon an experimental period, then we should know what is to be gained from that experimental period.

Three weeks ago there was a workshop hosted by AAPS/FDA--I believe TPP was a co-host of that--and there was in that meeting overwhelming opposition to the implementation of IBE as proposed in the preliminary or the draft guidance. In fact, I think it was noteworthy that Dr. Bill Barr stood up and said, "Roger, it looks like I'm the only one up here that's supporting you."

I think there's something to that. I do not know of any time there has been such strong opposition to a proposal by FDA and that they've proceeded with it.

Please note that the people who have most to gain from this are the CROs because their economy would substantially increase with replicate designs, and yet almost to a person, what I've heard is that the scientific rationale--this is from the CROs--the scientific rationale does not convince them of the need for IBE.

So Mr. Chairman, I would leave these thoughts with you emanating from the International Generic Pharmaceutical Alliance.

DR. BYRN: Questions for clarification?

[No response.]

DR. BYRN: Thank you very much.

The next speaker is Laszlo Endrenyi from the University of Toronto.

MR. ENDRENYI: I would like to comment on the primary motivations as suggested by FDA in the presence of estimated variabilities; that is in the presence of random variations.

The primary motivations, as Roger Williams said today, were to evaluate the subject-by-formulation interaction, to consider the effect of reference scaling, and to provide reward for the reduced variability of the test product.

I shall be considering the last two motivations. I have discussed the first motivation, the interaction, in Montreal, so that's too much.

Now about the reference scaling, it's mainly about highly variable drugs where the reference scaling widens the apparent bioequivalence limits, and narrow therapeutic range drugs where the scaling narrows the apparent bioequivalence limits.

Now a few comments about this. In the case of highly variable drugs, scaling by average bioequivalence, as I shall illustrate very briefly, would be probably more effective than scaling individual bioequivalence criterion.

As Dr. Benet said, the expert panel has actually asked for such a procedure for scaled average bioequivalence in October '98 but such a procedure has not been forthcoming.

I shall not discuss narrow therapeutic range drugs, just to note that there was a paper by Masson and Yacobi in Montreal which demonstrated how very restrictive that scaling can be for NTR drugs when the variation is small. So in short, it doesn't pay to have small variation because you pay for it by a small bioequivalence range.

In this slide the results of simulations are presented with 24 subjects, coefficient of variation of 40 percent. The red curve shows the results for scaled individual bioequivalence.

Now you notice it shows the percentage of acceptance, the trials under different conditions, as the difference between the logarithmic means. And you notice that this declines very shallowly, very gently, and this means that it does indeed permit large differences between means with a fairly high probability, and that is expected, but you see it in action.

The blue curve is unscaled bioequivalence, is the present procedure in four-period trials. It has comparatively little power, as we know, so we need more subjects.

The green curve is the curve for scaled average bioequivalence and the point is that it certainly has been characteristics for this specific purpose than the red curve, the scaled individual bioequivalence. It has high power. The concern about large difference between means still there but much less than with IBE.

I would like to turn now to the main concern of mine, that of reward. You have already heard about the trade-off between means and variances, so I don't repeat it. This was discussed by Walter Hauck and colleagues mainly from the FDA and under ideal conditions; that is, without the consideration of errors. And indeed there is a possibility of reward here. This is a consequence of the aggregate criterion.

So if I could comment about the aggregate criterion, this has, as you have seen many times today, it has three components put together and the sum of the three should be less than the bioequivalence criterion.

There are problems, difficulties with the aggregate criterion. This is very nice under ideal conditions. So it's attractive in principle; there are

major difficulties in practice, as we have had shown in a paper with Drs. Amidon, Midha and Skelly.

There are conceptual differences and you have just heard them from Nevine Zariffa and Larry Gould, the hierarchical problem.

I would like to be concerned with the technical problems. And anybody who discussed it outside FDA, the issue of aggregate criterion, was against it. And you have the names there.

If I could go on, the effects of random variations, first, not only rewards are present but also penalties. If you have truly equivalent, the two variances, inter-subject variances are truly the same, then there is a 50 percent change that the test variance estimated is smaller and a 50 percent chance that it is larger than the reference variation.

So there is a 50 percent probability, just by random chance, that there is reward and a 50 percent chance that there is a penalty. And that's the first thing. Actually I shall demonstrate this.

The second is that these rewards and penalties dominate the difference between the means. I shall demonstrate that, too.

So actually the usual concern about average bioequivalence gets very low priority because of the

aggregate criterion.

And thirdly, the rewards and penalties can be large. They're not just present but they can be large due to random chance. If I could go on to the demonstrations.

But this is from our theoretical paper and it illustrated that if, for example, you have a coefficient of variation of 30 percent, then the probability that you would see a 10 percent change in the difference between AUCs, which is quite a change, is 73 percent, half of which is reward and half of which is penalty. So that's quite a large probability, but his is theory. Let me go on to look at the FDA data, and this is the data of 55 datasets, which Dr. Williams has presented.

In the first column you see the rewards., all together, 49 cases that were reward, and 61 cases that were penalties. So apparently rewards and penalties indeed occurred at random.

In the next slide this is about the magnitude. The ratio of test over reference variance differed by more than 40 percent, which is quite substantial, all together, in 21 cases, which is a fair proportion, and the lower shows that the same kind of probability percentage occurs with statistically significant difference.

So in these cases you have very substantial effect on the outcome of decision.

This is about the mean variability trade-off. This is the '98 dataset because I didn't have the opportunity, didn't have time to massage the '99 dataset. It shows that the mean differences are less than the difference of variances in an overwhelming number of cases. So indeed the concern about the difference between means gets smothered away because of the quantitative feature of the criterion.

So in summary, in the absence of random variation, the aggregate IBE criterion is attractive, combines three features, but in the presence of random variations, there are conceptual difficulties, as already you heard earlier, and technical difficulties and I'm particularly concerned with the technical difficulties.

First is that there is reduced efficiency, as you have seen on those curves which I presented, and problems arising from the mean variability trade-off, and this is really what I am particularly concerned with, that apparent rewards and penalties can occur by random chance; large rewards and penalties can occur with fairly high probability by random chance; and consequently, favorable and unfavorable regulatory decisions can be reached by random chance. Therefore, the consequences of this trade-off amounts to a scientific and regulatory lottery. Thank you.

DR. BYRN: Questions for clarification?

[No response.]

DR. BYRN: Okay, the next speaker will be Dr. Russell Rackley from Purepac Pharmaceutical.

DR. RACKLEY: Thank you for the opportunity to make a brief presentation on a few points today.

I am employed by Purepac but I come here mainly as a research pharmacist in the industry. Thus I'd like to throw in the following disclaimer. Views expressed in this presentation are those of mine and not necessarily those of Purepac and its employees.

Briefly for an introduction, the proposed individual bioequivalence methodology may have some scientific merits. There are a number of unresolved issues apparently regarding implementation and use. Ultimate adoption of the methodology may or may not prove to be in the best interest of the public.

I'd like to cover the following points. Question first: Is there a problem? Is the current methodology--ABE, that is--protecting the public? If so, should a change be implemented? If not, is there proof that a change should be made?

In my opinion, a convincing case for IBE has yet to be made.

First, the method is complicated. The method and criteria cannot be easily conveyed to the public or even

health care professionals. Theoretically, information from a conventional two-period cross-over study may give the information needed to make such assessments as subject-by-formulation interaction, as would a four-period replicate design study. I refer you to the method of Gould.

There has been some disagreement on the actual mechanics, although those are claimed to be resolvable. From the recent AAPS workshop there was some disagreement. However, if one believes a new method is justified, little consideration apparently has been given to alternate proposed methods.

Subject-by-formulation interaction may be misinterpreted. It could be affected by random variation, and I give you the following example. If the response of a particular test reference comparison is clustered in one particular area and in one particular subject and a response of one of the treatments is removed from the other responses, this might be viewed as an outlier possibly, which could affect interpretation of the data and mislead a subject-by-formulation interaction conclusion.

What does subject-by-formulation interaction tell us? One of the driving arguments for individual bioequivalence has been to identify this. For example, response in a particular subject might be such that those test treatments are fairly close together or the response

level of the reference might be far removed. When this type of result occurs, it would appear to me anyway that you would get the same information from this as you would from average bioequivalence.

Limits on mean ratios or point estimates might be the more appropriate thing to say, seems to contradict the method. When the test formulation is found to be less variable, then the criteria may be scaled to the variability of the reference. However, it is proposed that some limit on these point estimates or mean ratios be implemented, which to me seems to contradict the theory of the method.

For some products, even reference-versus-reference ratios or estimates could be fairly divergent.

Further points on limiting mean ratios or point estimates. Limits on mean ratios with individual bioequivalence might negatively affect the public as follows. It is conceivable that pioneer companies may attempt to make formulations more variable. With even tougher criteria of putting criteria on mean ratios, there might be fewer generic formulations ultimately available.

The method as proposed, the guidance as proposed, would indicate reduced population sampling; that is, reducing the population sample from 24 subjects with average bioequivalence to 12 with individual bioequivalence reduces the potential to see or identify subgroups showing

significant differences between formulations.

It's my opinion that there may be some impact on generic competition. For a variable pioneer drug, drug formulations, there may be cases where only population bioequivalence passes. I think there were some presented at the last workshop.

If we move forward with this method, then pioneers should be held to similar requirements, at least for any significant formulation changes relative to the clinical formulation, pre- or post-approval.

And I think it has also been stated that it would be a good idea to have individual bioequivalence results appear in the labeling of pioneer drugs.

Acceptance at the state level. In certain states where formularies exist for substitution of generic drugs, it is sometimes a tenuous task to gain approval. How will these state agencies react to approvals under the proposed method?

There already is data available for us to make an assessment, roughly 55 to 60 replicated design datasets. My question is will an additional 400 significantly add to our understanding of the method, the utility of the method as we know it now? And it's also been pointed out that two-period cross-over datasets might also be evaluated to determine if the problem really exists. Again see the method of Gould.

In summary, the problem may be stated in theory but I feel convincing evidence is lacking. The method appears to be complicated, leading to implementation of multiple rules and conditions. In some instances, interpretation may be misleading. An example is subject-by-formulation interaction with respect to outliers. And potential to further evolve a brand-defense tool may exist. Data, I think, is currently available to assess the utility of the method.

Finally, a convincing case that the public will benefit from the methodology cannot be made based on existing data or even that envisioned for a trial period. Thank you.

DR. BYRN: Questions for clarification?

[No response.]

DR. BYRN: Okay, thanks very much.

The next speaker will be Dr. Leon Shargel for the National Association of Pharmaceutical Manufacturers.

DR. SHARGEL: I would like to thank everyone on the committee and Dr. Williams for allowing us to make a few remarks at this hearing.

My name is Leon Shargel and I'm the vice president and technical director for the National Association of Pharmaceutical Manufacturers. I'm also an adjunct professor at the University of Maryland School of Pharmacy. The NAPM

has been highly involved in legislative and regulatory technical issues concerning the generic pharmaceutical industry.

I'm sorry I did not prepare slides for the audience. The advisory committee though has a copy of my talk. And on behalf of NAPM and its members, including our generic drug product manufacturers and the contract research organizations, and we do have a large number of CROs as members, I'd like to discuss some of these recommendations for performing individual bioequivalence studies and specifically I want to chat about clinical significance, ethical concerns and cost considerations.

If we consider clinical significance, and we concur with many of the speakers already today, we agree with FDA's position that the prescriber and patient should be assured that the newly administered drug product will yield comparable safety and efficacy to that of the product for which it is being substituted.

The question is in our minds whether switchability, as defined here, is a clinically significant problem which we need to be very much concerned with.

We agree the use of replicate studies in determination of individual or population bioequivalence is useful and certainly we would consider that this would be a useful tool to look at in the future of bioequivalence.

Now on January 28, 1998, and there's a typo error on the hand-out that I gave to the committee, Dr. Stuart Nightingale, associate commissioner for health affairs, wrote a letter to health practitioners that was prompted by concerns about the interchangeability of certain products characterized as narrow therapeutic index drug products.

In this "Dear Colleague" letter, and I'd like to quote four points, one, "Additional clinical tests or examinations by the health care provider are not needed when a generic drug product is substituted for the brand name product."

Two, "Special precautions are not needed when a formulation and/or a manufacturing change occurs for a drug product provided that the change is approved according to applicable laws and regulations by FDA."

Third, "As noted in the Orange Book, in the judgment of to FDA, products evaluated as therapeutically equivalent can be expected to have equivalent clinical effect whether the product is brand name or generic drug product."

And fourth, "It is not necessary for the health provider to approach any one therapeutic class of drug products differently from any other class when there has been a determination of therapeutic equivalence by FDA for the drug products under consideration."

Now in the same letter Dr. Nightingale also wrote, and I quote, "To date, there are no documented examples of a generic product manufactured to meet its approved specifications that could not be used interchangeably with the corresponding brand name drug." That was written in January 1998 and I would assume by 1998 there had been a great deal of information at the FDA.

Now at the meeting in Montreal that was just referred to, Mr. Eric Ormsby of the Health Protection Branch in Canada gave a presentation and in one slide he reported that 2,500 products on the Canadian market were approved using the AB standards. He mentioned, and I quote, "Is post-marketing surveillance really so insensitive that clinically important problems can't be detected?"

Thus Mr. Ormsby, who was then representing the Canadian Health Protection Branch, has indicated that in Canada there has not been any observation of clinical safety problems due to switchability.

What then apparently is a problem, we feel certainly that the current approaches for determining therapeutic equivalence by the FDA is certainly working and the generic substitution of AB-rated drug products is safe and efficacious.

And, of course, this doesn't preclude that we shouldn't look at other methods. And certainly over the

last 20 years we have seen improvements in how we do our bioequivalence studies, how we do formulation and we're doing a better job, I believe.

However, do we really need to be performing individual bioequivalence and determining subject-by-formulation interactions on every bioequivalence study? NAPM does not feel that this is the case.

To date, approximately 50--I hear 55 now--datasets have been published, have been looked at, and at the annual meeting of APS in Boston in 1996 and at the IBE workshop in Montreal it's apparent to a nonstatistician, such as myself, that there is a lot of controversy and concern whether subject-by-formulation interaction is really a safety and efficacy problem. We feel that the scientific literature would be replete with clinical studies or at least case reports if this were such a major problem.

Let's move on to ethical concerns. A fundamental caveat in clinical studies in humans is "Do No Harm." And the Declaration of Helsinki has a number of basic principles and I'd like to recite a few of those and see how it fits in.

Principle number one is "Biomedical research involving human subjects must conform to generally accepted scientific principles and should be based on adequately performed laboratory and animal experimentation and a

thorough knowledge of the scientific literature."

And as I mentioned, at this time we do not have scientific literature that really indicates that switchability is a significant safety or efficacy problem.

The second principle in the Declaration of Helsinki states, "The design and performance of each experimental procedure involving human subjects should be clearly formulated in an experimental protocol which should be transmitted for consideration, comment and guidance to a specifically appointed committee independent of the investigator and the sponsor, provided that this independent committee is in conformity with the laws and regulations of the country in which the research experiment is performed."

With noted exceptions that were listed in the draft guidance, as mentioned by Dr. Shah this morning, FDA is recommending that all bioequivalence studies should be designed as replicate studies and that the applicant may use average population statistics or criteria for establishing bioequivalence. And this data then is also going to be used and collected by FDA for further analysis.

We're concerned that the request for additional studies and extra datasets from human subjects should not be obtained without peer review protocol describing exactly how the data is to be analyzed, the risk-benefit assessment and how the data is going to be used.

The third principle that I want to mention from the Declaration of Helsinki is principle number four as listed in the declaration and that states, "Biomedical research involving human subjects cannot legitimately be carried out unless the importance of the objective is in proportion to the inherent risk to the subject."

Now in terms of ethics, we're concerning always a risk-benefit ratio and a four-way cross-over replicate design always has a greater inherent risk to the subjects. It doubles the drug exposure compared to a two-way cross-over study. The chances for an adverse drug event is certainly greater and the fact that we're taking more blood samples per subject may also increase trauma to the subject and risk of damage to blood vessels.

We should be concerned about the subjects. I haven't heard anything yet this morning or in many of these seminars about concerns for the people who are actually going to be involved.

There is also--the last item I want to emphasize is cost consideration and burden to the industry. If we just consider the financial cost to the pharmaceutical industry, I don't think it really matters much if our objective is to make better, safer, more efficacious drug products. I would not stand here if I felt, gee, it'll cost a little more money; it's going to make it a little harder.

Cost should not be an issue if our objective is that we have safer, more efficacious products.

However, the proposal as written in the draft guidances is going to increase cost both to the finished dosage manufacturers, as well as to the contract research organizations. They will not be making lots of money, according to our membership.

First, the cost for replicate design studies is much higher than a two-way cross-over. I pulled two of our--actually I pulled several others but I got a reply from two and you should have in your file three drugs that were looked at: warfarin, indapamide and diltiazem, and these costs of studies were greatly increased.

From a CRO point of view, there are recruitment problems, problems of getting subjects, institutional review problems, drug monitoring problems. In the case of diltiazem you're going to be four-way cross-over electrocardiogram or other kinds of monitoring and a need for increased clinical capacity.

So in summary, so I don't go over my allotted time, we do not feel that switchability is a clinically significant safety or efficacy problem. The risk-benefit should be concerned in performing replicate design studies. We should carefully consider this. The replicate design will put an additional burden to the industry. And we're

also concerned with how the data will be used, whether this data will also be used perhaps naively by consumer groups, state formulary commissions and others, as well.

We do commend or compliment FDA for looking at methods for reducing burden in terms of a single dose study for modified release or use of the VCS system for highly permeable, highly soluble drugs. However, for the studies in terms of individual bioequivalence, we feel that a well designed with objective statistics analysis should be available for peer review.

I thank you for the time.

DR. BYRN: Questions for clarification?

[No response.]

DR. BYRN: Okay, thank you.

We have two speakers that have asked to be added to the list. I'd like to give you each two minutes but that may be too little, so let's try to limit it to five minutes apiece.

First speaker is Lew Sheiner.

MR. SHEINER: My name is Lewis Sheiner. I'm a professor of clinical pharmacology at UCSF and I am one of the early proponents of individual bioequivalence and probably--I don't know who else wants to speak but if the mystery speaker is not of my opinion then I'm the only person who speaks for it.

I want to justify the criterion a little bit and talk about the bias variance trade-off because these are two issues that have come up in all of the talks this afternoon that it seems to me there's another point of view on.

So I'm in opposition to the concerns expressed that we don't know what it is, what we would do about it and what this interaction means and so on.

Taking my clue from Roger, I'm going to try to make this very simple so here's the set-up. I drive to work every day and I stop off at the convenience store and I buy a cup of coffee and I put two teaspoons of sugar, put the top on the thing, get in the car, put it in the little hook in the car and drive way toward work and drink the coffee on my way to work. But I decided that my paunch is getting a little big and take the easy way out. I'm going to put something like this in instead.

So I ask people, I said, you know, this generic substitute for sugar, tell me something about it. They said one of these is a teaspoon of sugar.

So now I'm going to get in my car, I'm going to buy this stuff and throw two of these in; that's the way I'm going to start, and take it in the car.

Now the reason I talk about taking it in the car is because I want to make this thing be symmetric. If I oversweeten it, obviously I'm not going to like it but I

can't do anything about that. If I undersweeten it, I'm in the car and I'm not going to take the top off and tear open one of these and try to--so my problem is I'm stuck with what I put in to start with.

So what am I going to do? I'm going to put two of these in because people say one of these is one teaspoon of sugar. And then what'll happen is I'll taste my coffee every day and I'll titrate. And I'm going to discover perhaps that one and a half is right for me, rather than two.

Well, what if the mean difference, that statement was correct, that there was just one of these to one sugar and that there was no inter-individual variability. Then I'd be just right the first time and wouldn't have to titrate or anything. I'd be right on.

What if the mean difference exists but there's still no inter-individual variability? In other words, it's not true that one of them is one. It's some other ratio.

Well, I'll still have to titrate but remember in that case if I have some other friends who've tried this and they titrate, they say, you know, it turns out you need three packets for two of sugar, then I can just do that the first time and I'll be fine. So I don't have to titrate if I have other people's experience and that delta exists.

What about if inter-individual variability exists

but no mean difference? I'm going to have to titrate. I'm different than anybody else. Nobody can tell me what to do, so I have to titrate. So I'm going to go through that titration period if both of those aren't zero.

So in that particular circumstance I claim--oh, there's one more thing. What if there's day-to-day variability? They don't put the same amount in the packets. Every packet has a different amount. Day-to-day variability of the stuff. Then I'm never going to get it right. Some days it's going to be too sweet; some days it's going to be too bitter and that's it; I'll never get it right.

So I look at that and I say that's the worst case. That kind of variability is worse than anything.

Second worst is difference in the means because if my friends can learn about it they'll tell me about it and I can get it straight the first time.

In the middle is that inter-individual variability. I can learn about it and get it right and have my coffee right every day but it'll take me a little while to titration.

So that's variability versus mean trade-off on the starting. What about switching?

I come in one day in my store and they ain't got the pink packets; they've got the blue packets. Suddenly I don't know what to do. Again I'm told one is one.

Okay, what happens? If the mean difference is there isn't any, one is one and I've gotten myself to 1.5 packets remember on this stuff--I've found that that was right for me--and there's this perfect correlation between the blue and the pink, that means there's no subject-by-formulation interaction, then I'm right on. One and a half of those, one and a half of these; I'm great.

What if there's a mean difference but no subject-by-formulation interaction? Again I have to titrate. But again I can learn from my friends, since they did it before me, and they say no, the blue and the red aren't exact equals. It's three to two.

What if there's subject-by-formulation interaction? That means my particular ratio is different than somebody else's ratio of the pink to the blue. Then I've got to titrate, but I can get to where I'm going. And after a few days of having the coffee wrong one way or another I'll get there.

Worst case again. The between-packet variability in the blue stuff is worse than the between-packet variability in the red stuff. Then I'm getting weird coffee more often on that than I did on the original. That's the worst case.

So the worst thing that can happen is big within-individual variability. Second worst is subject-by-

formulation interaction, which acts just like between-individual variability when you're starting. And the least important is the difference in the means.

So the lesson is if you care about the mean, if you think you care about the mean, you have to care more about those variabilities. They produce the same problem as the mean only worse. Either you can never get the thing right or it takes you some titration to get it right. But in the mean case remember you can learn from your buddies.

So that gets me to say finally that the only question then really is are the differences in between-individual variability between innovator and generic big enough ever or are the ratios of the within-individual variabilities not one or far from one ever? Or is the subject-by-formulation interaction ever large enough to worry about, to cause me to have those problems?

And that's a question that's settled by fact. It's not settled by lack of fact, the non-dead bodies in the street, which is just not an argument. And it's not settled by argument. It's settled by fact, and the FDA is proposing to get some more facts on this issue. I think that's almost an unexceptionable notion in the scientific age, to do that kind of thing.

Finally, let me just say a word about Laszlo's point. There may be technical issues; I'm not sure. I

think it really actually, those technical issues will depend upon again defining what it is you care about.

Now I believe in the aggregate criterion. I believe in these distance measures, but we can all agree on what we care about, the what do you want to know question, and then we can settle the technical issues. We can settle them through simulation and we can design a study format and design an analysis system that has the right performance characteristics. We can design a test that has just the right performance characteristics and have no question about that.

I think we're actually pretty close to that but maybe we need to do some more of that. But the point is that's a purely technical problem and that can be settled and you have good people working on that.

So the issue you have to think about is if you care about the means and you care about the variability, the question is do we know enough to know that those variabilities are of no concern?

DR. BYRN: Questions for clarification?

[No response.]

DR. BYRN: Okay, thanks.

Next speaker is Robert Buice.

MR. BUICE: My name is Bob Buice. I represent the Bioequivalence Focus Group for the American Association of

Pharmaceutical Scientists. We have submitted a position paper on this topic of population and individual bioequivalence and I think you have a copy of it now. I would like to briefly, in my two minutes or five minutes now, highlight some of the points.

I think all of these have been made already but I'll run through them very quickly anyway.

Variations of the average bioequivalence approach have been used for more than two decades now. Recently we've seen a few subject-by-formulation interactions pointed out. We've seen a few variance issues raised.

Overall, nothing really jumps out at you as being a serious clinical problem. The problem, we feel, is still largely theoretical. Now there might be something there but there's just nothing jumping out as saying that.

I could go on and on about the complexity of the replicated designs, the increased clinical costs, the problem with drop-outs, the increased exposure to the subject. This has all been pointed out.

A key point we'd like to make though in this subgroup causing the subject-by-treatment interaction, you're talking about doing studies with 12 subjects, 24 subjects, even 36 or more. What are your chances of picking up that small subgroup in one study? You might do four or five studies and never see it and pick it up in another one.

There just doesn't seem to be that many of them.

Also, this Sigma D of .15, as has been pointed out, that can occur by chance. It's also been pointed out that there's no evidence that that suggests any kind of clinical significance.

And regarding this two-year trial period, a lot of data have been submitted over the past two decades and Larry Gould has already suggested a method of analyzing these data.

Also, the FDA has replicated data. There's a lot of data already available. It just doesn't seem warranted to jump into a two-year trial period without a little more reason to do that.

Now if there are isolated problems, if there are minor problems picked up, and that's about what we suspect will happen, maybe 5 percent or so or less, if that many, we suggest you treat those as isolated problems. If you see a subject-by-formulation interaction, identify the mechanism, the physiologic mechanism, the pharmaceutical mechanism, whatever, and treat that as a separate problem.

And finally, we suggest that you keep the present method in place until a serious problem has been identified. Thank you very much.

DR. BYRN: Questions for clarification?

[No response.]

DR. BYRN: Okay, thank you.

Okay, let's take a 15-minute break and then the committee will be--oh, I would like to thank all the speakers for staying on time. Then the committee will begin our deliberation.

[Recess.]

COMMITTEE DISCUSSION

DR. BYRN: Okay, what we're going to do is deviate from the agenda just slightly. Dr. Vince Lee has to leave so I've asked him to make a few comments. Then Roger will go ahead and introduce the discussion topics and the committee will continue.

So Vince, the floor is yours.

DR. LEE: Well, thank you, Steve. I have to be brief because the FDA shuttle is going to leave in a few minutes.

I think the idea about IBE is a very forward-looking approach for drugs coming through [inaudible] chemistry. I'm speaking as an academic and I have the suspicion that the new drug candidates coming off the pipeline as a result of [inaudible] chemistry will have more and more challenging delivery problems that are more prone to variability.

And I think that even though ABE might be addressing and might be serving us well for the time being,

it's about time to take a long-term view about what to anticipate in the future.

So that's my view. So I'm in support of the concept. I do agree that we need to do more work to substantiate the concept. On that note I will close. Thank you.

DR. BYRN: Thanks very much, Vince. Good luck. We wish you safe travel and we'll now go back to our agenda and let Roger introduce the discussion topics.

INTRODUCTION TO DISCUSSION TOPICS

DR. WILLIAMS: At this point what I'd like to do is really now start working very closely with the committee to assist the committee in any way possible as they deliberate on the six discussion topics that you see in your agenda. And associated with each of those discussion topics will be an overhead and Kimberly, I think you can go to the first one, which is a question for the committee. You'll see that all these six questions are interrelated and sort of flow sequentially one to another.

Associated with some of these questions I may show another overhead in an intent to clarify the question. And I have to be very careful in terms of not influencing the committee here, but I will say if the answer to the first one is no, we can probably all go home.

So I don't want to give you an incentive but we

put this very important one up first. I think you can see why it's so important.

Now at this point in time I think I will just sit quietly, Steve, and if there are people in either the working group or the expert panel who I think could provide some assistance, I'll make that statement.

DR. BYRN: Okay, thanks, Roger.

I've had a couple of questions and maybe we'll start with Kathleen. She had an early question.

There was also a request by the committee that we would be able to ask speakers questions directly to clarify certain points related to these topics, so I think we will do that.

We ask people in the audience if they are asked a question to simply answer that question, to not engage in a debate with each other or with the committee.

So with that, we can start with Kathleen, who had a question for clarification.

DR. WILLIAMS: Steve, could I just say one more thing before Kathleen starts, that I meant to say?

DR. BYRN: Sure.

DR. WILLIAMS: You know it's obviously up to the committee how they want to provide their recommendations to the agency, which can either be done by some kind of consensus process or a vote. And I have absolutely no

opinion about how you handle that, Steve. If you feel at a certain time a vote would be justified, I'd leave it up to you as chair.

DR. BYRN: Okay, I think the general thinking of the committee is that we'll try to do it by consensus. It may be necessary to take a vote but we don't plan to at this time.

So with that, we can go ahead with Kathleen's questions.

DR. LAMBORN: I had a couple of points of clarification resulting from some of the statements this afternoon.

I'd like to ask Dr. Hauck if he could comment on the perceived differences. A number of people suggested that Larry Gould's method could be used in lieu of replicate designs and obviously the working group has been considering this. Could you comment on that?

DR. HAUCK: I wrote down a couple of notes, as I was given at least a couple of minutes advanced warning. I apologize. You will see why. I don't normally hand-write anything for public presentation but hopefully this will be helpful.

Let me, by way of background, say that the FDA, in getting to the criterion, has recommended in the working group, we considered every single criterion that was out

there and available at the time, including disaggregate criteria. The aggregate-disaggregate terminology actually comes out of the working group efforts. So this is not really new.

Specifically now with respect to Larry's proposal, there's a couple of things in there. One is he talks about--and actually Laszlo did, too--the hierarchy of criteria, that if you have a criterion that shows IBE, it should also show a population bioequivalence.

I think if you actually looked at the material that has gone by you this morning, and unfortunately that kind of means a little bit of dealing with the Greek letter aspect of it, you'd actually see it's not sensible. The two sets of criterion end up depending on different variances.

The hierarchy question has actually been there since--actually, this is actually the 10-year anniversary of the initial presentation on individual bioequivalence, and I suspect the hierarchy question was there one month after that. So this is again--it's been around. I've been looking at it since about that time and every single time I see an individual bioequivalence criterion and a population bioequivalence criterion, they don't satisfy the hierarchy.

So how does Larry do it? Well, if you actually look at what he does, he does it by not doing individual bioequivalence. There's two things in there where he fails

to do it. One is he's looking at the wrong variance, and that's really the main thing. When you look at population, it's total variance. When you look at individual, it's within-subject variance. He looks at total variance. So within-subject variance doesn't get covered.

And then he also only looks at a piece of to subject-by-formulation interaction and I'd like to thank my colleague Terry Hislip, who's been working with Larry's method a little bit for the following graphic.

When you look at the subject-by-formulation interaction, as actually Larry had pointed out, it has two pieces to it. Part of that piece is whether the between-subject variances are equal. In fact, his approach is not sensitive to differences in between-subject variance, and that's what this is attempting to show. The Y axis here is his correlation coefficient and the X axis is the within-subject variability. And you see the sensitivity is totally a function there of the within-subject variance and drops off considerably. And that's very different. That's for one between-subject variance, twice the other.

Even more so, I think there's a more fundamental problem with the approach that Larry's taken is that there's really no hypothesis there. He starts with a test statistic. You then have to reengineer and reverse-engineer, rather, to find out what the regulatory

requirement is that's associated with the procedure that's put forward.

And if you happen to do a three-period design or a four-period design, you get a different answer. That is in the fact that the regulatory criterion established by Larry's approach would be different depending on what design the sponsor chose to do.

Now I'm a statistical consultant, not the regulatory person, but it strikes me as somebody who's informed about some of this, that that seems unsatisfactory in a regulatory context.

And this would be kind of new information. Obviously the appeal of what Larry has proposed is that it can be done in two-period designs. I want to mention and I'd like to thank Terry for this; we have found that at least for the low variability products--that is probably in the 10 to 15 percent range of lower--that, in fact, the aggregate criterion that we have proposed can be done in two-period designs. So some of that appeal goes away.

I think there's general appeal to the disaggregate approach and I think Larry is to be credited for taking what has been vaporware in the disaggregate area in the sense that people have been saying "We want disaggregate" and not really proposing anything and he at least was willing to put something on the table. But I think for the reasons I've

outlined, it's not really a viable alternative.

DR. BYRN: Any questions? Kathleen?

DR. LAMBORN: I must admit that just hearing it now, I can't totally follow all your points except to say that that you feel that you've looked at it and that it is measuring a different--something different than the individual bioequivalence. But then if that's the case, why are you saying that depending on variability, you can measure it?

DR. HAUCK: We can test the aggregate individual bioequivalence criterion with a slightly conservative test in a two-period design. We cannot separate out the components in that circumstance but if somebody were to say can you do a 5 percent, or actually it would be slightly less than 5 percent test of the individual bioequivalence aggregate criterion, as proposed in the guidance and that would be the constant scale piece, yes, that can be done with a two-period design without paying too heavy a penalty in sample size for the conservatism.

But no, we can't estimate subject-by-formulation interaction. We can't do a separate comparison of the variances. I mean it's just the aggregate piece as an aggregate.

DR. BYRN: So I think the answer is that you could do some limited work but you couldn't do the complete study

with the two-period design.

DR. HAUCK: That's correct.

DR. BYRN: Could I ask Roger a related question, which would be, and this relates to all of these models; if the data was requested by the FDA--that is, if we answered discussion topic 1 yes--would there be any way that that data could be put, blinded, of course, be put on the net so that people could perform analyses of different types to try to understand other ways of analyzing the data and so on?

DR. WILLIAMS: Yes, Steve. As a matter of fact, you're reminding me. Kimberly, I had a second overhead, I believe, with question 1. It should be following that. That's a series of steps that kind of give a motivation for the replicate designs and what would we do with them in this interim period.

And when we get to the sixth question, what we're trying to do is to find our analyses and protocols and approaches during this interim period.

I will congratulate the expert panel. I think some of these suggestions came from them and we want to further elaborate it when we get to topic area 6.

But Steve, you'll see in step 4 there's an attempt to get to just what you're talking about. We'll try to give progress reports as this interim period moves along. To the extent possible, we will share publicly the available data.

DR. BYRN: Okay, other questions for clarification on question 1? Robert.

DR. BRANCH: The proposal right now is to accept a smaller sample size for bioequivalence studies, half that sample size, and do replicate measures in those individuals.

Can we hear some reassurance from either Les or Walter about the impact that will have in trying to identify an already small subset who we're suspecting has a subject-by-formulation interaction? It seems to me that going down to 12 subjects, if you're now saying what is the effect of gender comparison or what is the effect of age comparison, within such small subgroups the statistical power is going to be extremely small.

DR. WILLIAMS: Steve, may I comment on that? That actually is an excellent question which comes up, I believe, in topic area 5. We kind of put that a little lower down, Bob, but it's a great question and we would like to hear from the advisory committee on that.

DR. BOEHLERT: Maybe this is a clarification question as well but for sponsors who submit these kinds of studies during this two-year interim period, are the conditions of approval then dependent on the outcome of those studies? And if indeed flaws are identified, then what?

DR. WILLIAMS: Judy, I think you got us to

question 3 I want to say, which relates to a very critical question about market access. Are we going to be willing to rely on the individual criterion for market access?

We do have methodology now that always allows us to use the average approach with a replicate study and Don Sherman is here in the audience and Stella Machado if you have questions about that.

DR. LAMBORN: I think a related question though, is suppose you see what appears to be an interaction which might raise concerns. Will you ignore that for the purposes during this interim period? So do you truly use the average and take the other as being a research component?

DR. WILLIAMS: Kathleen, that's such a good question. I feel like this is what we do now and we don't even know about it. So I guess what would we do if we saw a huge subject-by-formulation interaction? I would say the agency has a right to ask an applicant about it and further discuss it.

MR. BOLTON: One possible answer to that question is that when you analyze the average bioequivalence using a replicate study, you would be looking at the interaction term as the error term, and that would really--if you had a large interaction, that would really widen the confidence interval and perhaps cause it to fail unless you use a very large number of subjects.

So there is some protection for that.

DR. BYRN: You need to identify yourself for the recording. That's okay?

And I appreciate the input but anybody that speaks needs to be recognized by the chair. That was Sandy Bolton. Thanks.

Other questions for clarification of discussion topic 1? Kathleen?

DR. LAMBORN: I guess I'm having a little concern that--I understand why Roger put it first because it's sort of the bottom line but I think as I think about the discussions and some of the individual comments that we've had back and forth and also the specific recommendation from the expert panel, if we take this as being is it reasonable and appropriate to recommend replicate study designs for some specified drug products with some conditions yet to be determined--in other words, if all we're saying at the beginning is is it worth discussing the additional questions, then that's one thing. Because it seems to me that we are coming back to a lot of those specific questions in terms of a comfort level in saying that we could recommend that we move ahead with replicate designs.

DR. WILLIAMS: I think Kathleen's suggesting that we maybe, and I like this idea, discuss question 1 in terms of something like is it in principle a reasonable approach?

And in a way, that's the way it's worded. Do we think that under certain conditions, undefined as yet, is it appropriate for FDA to recommend replicate study designs for specified products?

Maybe we can start with that discussion, just sort of an in-principle discussion. Robert?

DR. BRANCH: I liked Dr. Sheiner's comment that this area really would benefit from having some data on it. But as I heard it presented, particularly by the expert panel, what is being proposed is a two-year experiment.

Usually when you propose an experiment, you not only have your entry criteria but you have some methods, you have some analytical criteria, but you have some objective in terms of what the outcome would be.

It seems to me that--I would just like a level of reassurance, and I think that last slide that Roger was showing was the first that I really saw about where the process would go if this does start, that this is actually considered as an experiment, which means that the FDA is open to the proposition that this is not contributing to overall public health and that this is something that can be rolled back. If this is gone into in the full expectation that once started, this is inevitable and will always continue, it's not an experiment.

So I would just like some reassurance and

clarification of okay, if industry is requested to provide information in this format for two years, what criteria would justify rolling back that position and going back to the current status quo?

DR. BYRN: Maybe Roger can comment on that.

DR. WILLIAMS: I think again an excellent question and this slide that you see here, Bob, is an attempt to begin to have the agency think about the protocol, if you will, for the interim period.

And then I think when we get to topic area 6 we'll get more into a discussion of that as to how the committee would give us recommendations as to the specific elements of the protocol.

And I think it's a good question. I think it's a very fair thing to ask the agency to do.

Now I think the rollback concept I might state in a slightly different way because there's always the thought that based on a better mechanistic understanding, the way Larry Lesko talked about, we could move more products into the Biopharmaceutic Classification I System.

Now I don't know that you would call that a rollback but it's more a deviation to say as we get the data to understand these things better, we don't have to do in vivo studies at all.

Now that's the essence of the Biopharmaceutic

Classification System. And the Biopharmaceutic Classification System rests on the assumption that you will not see subject-by-formulation interactions.

Now that's not only a rollback; I would call it a roll-forward.

DR. BYRN: Roger, who would write the protocol if this went forward? Who would write the protocol? Would that be done by the expert panel or would that be done by the agency?

DR. WILLIAMS: I think it can be something done by the agency subject to review by the expert panel and perhaps a further endorsement by the advisory committee.

DR. BYRN: Kathleen?

DR. LAMBORN: I guess I'm coming back to the idea that was mentioned, and I think it links to what we just said, that it's important that we know exactly what we expect to learn when we come out of the two-year interim period. And, for example, does it make the most sense, rather than saying we're going to look at all products unless there's a safety reason or other reason not to, to start with the instances where there is some reason to think that if there's going to be a subject-by-formulation interaction, that that's the group where it's most likely to be found.

So in other words, do we need as broad a brush to

move into this interim period or can we perhaps come up with a narrower definition of the group which is worth studying first?

DR. BYRN: Arthur and then John.

DR. GOLDBERG: Kathleen, I'm a little concerned about limiting the scope because I think that you might bias the outcome. If we look, for example, only at highly variable drugs and the agency would like to apply this to other drugs, we won't have any other data on any other drug other than the highly variable.

So I would like to see, if we are going to go through with this, I'd rather see it not be limited but to go across the board.

DR. LAMBORN: I was thinking more of the instances not of highly variable versus not highly variable but the issue of whether we really do have substantial cases for subject-by-formulation interaction.

DR. WILLIAMS: Steve?

DR. BYRN: Yes, Roger?

DR. WILLIAMS: May I just say that that is the second question. So if there can be some agreement in principle, we will immediately go to that question.

DR. DOULL: I guess I share Kathleen's concern in that this morning when Dr. Gretter was talking, he talked about eight or nine drugs, something like that, and then we

heard about Cyclosporine, Levothyroxine and so on. But somehow I have the feeling that we're devising a system that will be applied to a huge number of drugs, all the drugs, and really the problem is a more defined problem. And somehow I have the feeling we haven't well defined the problem yet.

I appreciate what Arthur is saying, that we need to look or we certainly won't find anything but if we're devising a system that applies to everything because we have a few bad actors, then the information I'm not sure will justify that effort.

DR. LAMBORN: I think I got asked to modify the first one to see if we had something that we could vote on, which is just another way of making this as--is it possible that there's any place where we're going to want to recommend replicate study designs? And do we agree that that's at least worth exploring? And then we could go to the more specific. I think, Roger, that's what you're asking us to do.

DR. BYRN: Okay, does everybody understand the change? So we're trying to refine the question a little bit to address it. Does everybody understand the change that Kathleen has proposed?

DR. LAMBORN: And Roger, is that consistent with what you intended?

DR. WILLIAMS: Yes, I think these seem like excellent modifications.

DR. BENET: What's the difference between sum and specified?

MS. TOPPER: Go to the microphone, please.

DR. LAMBORN: The question was what's the difference between sum and specified? Probably not much. It was where I sort of started from so I just followed through with it.

DR. BYRN: Okay, what's the thinking of the committee? I'm hearing that there's some thought that maybe we should discuss some of these other topics before we go back to the first topic? On the other hand, we have a fairly general question here that we can try to determine whether there's some consensus as to whether there's consensus on supporting this question.

Does anybody on the committee want to go ahead to some of these other topics? Do you think we should further discuss this question and see whether there's consensus?

[No response.]

DR. BYRN: Okay, let's further discuss this question and see whether there's consensus.

DR. LAMBORN: Could we just maybe have a show of comfort level on this?

DR. BYRN: That's what I'm trying to do. Anybody