0100

And this equation here just shows the linearized test for bioequivalence, including a factor for the would have been referenced variability. And both AUC and C max must meet the bioequivalence acceptance criteria using this approach.

Now some advantages of using this approach and thinking about the plot for the simulations that Dr. Haidar showed earlier this morning, his simulations certainly confirm these features of the approach in that if the test variability, the test product variability is less than the reference product variability, then using the scale of average bioequivalence approach will benefit the test product.

If the test variability is greater than the reference variability, there should be no benefit to the test product, and this was shown by some of Dr. Haidar's simulations and we believe that this approach, by using this approach this will help discourage conducting sloppy studies or not give the highly, not give the scale, reference scaled average

0101

bioequivalent advantage to poorly formulated products or sloppily conducted studies.

Now one question we had is what about borderline highly variable drugs, drugs for which they don't always consistently show a within subject variability of greater than or equal to 30 percent.

As our simulations were presented earlier, they did confirm that for a true borderline highly variable drug, either a scaled or unscaled bioequivalence approach is suitable. In other words, the outcome of a three-way cross-over study would be the same whether a reference scaled average bioequivalence analysis or an unscaled average bioequivalence analysis is conducted. So in other words, for a true borderline highly variable drug, there should not be a problem with using the three-way cross-over study design approach.

Now when the scaled average bioequivalence approach is unsuitable, we believe that this would be when high variability is due to the generic product itself as opposed to the drug substance or the conduct of the study.

0102

If the variability is due to the affects

of the generic formulation, then the product is not going to benefit from scaled average bioequivalence. In other words, if the test variability exceeds the reference variability.

If the studies are poorly performed and it appears that the reference variability is high because the study was poorly performed, then we believe the burden should be on the applicant to prove to the Office of Generic Drugs that the drug substance is highly variable. And we can conclude in individual cases that the scaled average bioequivalence approach is unacceptable.

Our reviewers do routinely confirm all the calculations that were done by industry, they run their own calculations and they would certainly routinely start doing the calculations for studies that are submitted using this approach.

Now there's several concerns that we have about reference scaled average bioequivalence used for highly variable drugs and these concerns have been alluded to by the speakers this morning.

0103

The first concern is that firms will conduct a replicate design study, then submit

results with both scaled and unscaled bioequivalence analyses and maybe the two different analyses will give different outcomes, in that one will pass and the other will fail. In other words, this is the pick the winner approach.

Our proposed solution is that to evaluate, in these cases to evaluate the within subject -- the within reference variability very carefully and basically if the within subject variability of the reference product is greater than or equal to 30 percent, we'll use the reference scaled average bioequivalence approach. If the within subject variability for the reference product is less than 30 percent, then we will use the unscaled average bioequivalence approach.

A second concern is that scaling can allow the resulting AUC and C max geometric mean ratios to either be unacceptably low or unacceptably high. Our proposed solution is that acceptance criteria can include a point estimate constraint and

0104

this has been discussed this morning.

And a final concern is what should be an appropriate number of subjects for a bioequivalence

study that uses this approach. In other words, should the FDA recommend a minimum number of subjects.

And finally, I'd like to acknowledge the efforts of a great many individuals that contributed to this project and contributed to these presentations this morning, the Office of Generic Drugs, highly variable drug working group and the division of bioequivalence research group, all of whom collected a great deal of data from over 1,000 studies in a very short period of time.

I'd like to thank everyone who worked on this and thank you all for your attention.

DR. COONEY: Thank you. I'd like to take a few moments for any questions around the presentation, but I will suggest that we have discussion on the proposal after the open public hearing period.

Are there any? Marv?

0105

DR. MEYER: Two brief questions. Your slide that was entitled when scaled average BE approaches unsuitable, I object to unsuitable really because if the generic doesn't benefit, that's

tough.

It's not really an unsuitable design, it just doesn't help the generic get passed, so some other word than unsuitable, perhaps.

And the, you mentioned a group sequential design, is that essentially an add-on?

DR. DAVIT: No, that's not an add-on.

DR. MEYER: Okay.

DR. DAVIT: That's what I mean, that the study has to be in place apriori. In other words, the protocol is designed so that there is the option of adding on, but the statistical -- well, okay, there's the option of conducting a second cohort or a second study.

DR. MEYER: So that's an add-on?

DR. DAVIT: It's not an add-on in the sense that, I guess in Canada it's added on, and I guess I think the difference is that there is a

0106

difference in how the statistical are evaluated and this has to be set at the beginning.

DR. MEYER: Right, but I was unaware FDA would even accept, quote, an add-on design, I thought that was discouraged?

DR. DAVIT: We've actually been encouraging it for the last two years.

DR. MEYER: Oh, okay.

DR. DAVIT: But we haven't seen any protocols to date. I think basically because of the complexity of the study and the fact that, you know, the, to maintain an alpha of .05, one might have to adjust the competent intervals to 94, 95 percent.

DR. COONEY: Ken.

DR. MORRIS: Yeah, I think there's a lot to discuss for this afternoon, but just one question on slide 4, when you say some reasons for high variability in BE parameters, drug substance obviously and then in drug product you say inactive ingredient effects and manufacturing effects.

Is this, are these data that you're referring to implicitly or is this just, is this

0107

just by inference?

DR. DAVIT: These are data that we're referring to and we've inferred it from the data. In other words, we've seen differences in the formulations and it's possible that some of these formulation differences could be contributing to the

variability.

DR. MORRIS: But I mean are you seeing it in the tests as well as the reference?

DR. DAVIT: That's a very good question.

No, we don't know. We don't know. The reference is constant and then the variability is in the test product.

Like I said, it's pooled right now because all we have is two-way cross-over studies.

DR. MORRIS: Right. Right. Thank you.

DR. COONEY: Meryl.

DR. KAROL: I would just like some clarification, are all the study results reported or just those that are successful, because we've heard a difference?

DR: DAVIT: That's a really good

0108

question. That's an excellent question.

Unfortunately, this is very much a biased sample because the, at present applicants submitting ANDAs are not required to submit all their bioequivalent studies. They are only required to submit an in vivo study, and the decision -- well generally, generally companies do one fasted

bioequivalence study and one bioequivalence study under FED conditions, and generally all that we see are the passing studies. So, we don't have a sense of the failed attempts.

DR. KAROL: (Not talking in mic) of the number of tests that are conducted even if you don't see the results, you just don't know?

DR. DAVIT: That's correct. Yeah, we just don't know.

DR. COONEY: Okay, thank you very much.

I'd like to move -- were there any more questions from the committee?

Okay, I'd like to move to the next presentation and we will come back with adequate time for discussion of this topic later.

0109

We have scheduled 45 minutes for an awareness topic on risk management of complex pharmaceuticals and Steve Kozlowski will make this presentation.

DR. KOZLOWSKI: I just want to start off by making a comment about relaxation, so you can relax by meditating and making your mind blank, but you can also relax with yoga, which has a lot of

complicated, active positions. So we'll see what type of relaxation we're looking for.

So, basically as an overview of what I want to talk about, so some background, and I think we heard a lot about risk management yesterday, so I will try and move through this relatively quickly.

Some ideas of how risk management or risk assessment could be applied to complex products, less than the whole nine yards. Are there parts of this if it's impossible to do the whole thing that make sense to do and then finally, what kind of considerations would we need for the future for this.

So, to start off with I'll show a slide

0110

that we saw yesterday, this is from ICH Q9, there are lots of different risk assessment tools and they may fit particular jobs and will not all be useful for all things.

We also saw this table a number of times yesterday in which risk management is a complex process with many components and what I would like to focus on is the risk assessment issue, because I think that, at least for complex products with many

attributes is the biggest problem.

How do you really assess the risk of the attributes, not so much how you deal with them once you know what they are.

So, again, risk is defined as probability times severity. There are questions you ask what could go wrong, what are all the different things you need to look at and for each one what's the likelihood and what's the consequences of those things going wrong.

Now the use of this was discussed again yesterday and the first topic, inspections and audits are the two main examples that we were given,

0111

so this is clearly an area where risk assessment has value.

But also in the guidance it talks about facilities and equipment evaluation, materials management, manufacturing and change control and then finally assessments, including product quality. So how would you begin to use some of these systems for some of these products. And again, this is not answers, but how to begin to think about it.

So, again, we've heard about different

kinds of risk management, so failure mode effect analysis is a bottoms-up risk assessment. It looks at individual things and then it assesses the impact of what they, of what goes wrong and how severe it is and the frequency and it's semi-quantitative and basically assigns categories for probabilities, categories for consequences and then makes boxes which might be considered low risk, high risk and in between risk. And again, a very qualitative set of assessments.

There are opposite risk assessment tools, like a fault tree analysis where you start

0112

with the disaster is and you work your way down. And so this is an example in terms of such a risk assessment about a car crash. Cars at both junctions -- I would actually say I didn't get a chance to work this out, but to do a risk assessment tree like this for not being able to attend an advisory committee, so you could have inability to fill out the paperwork or unwillingness or you could have rejection of the paperwork and for rejection of the paperwork you could have, you know, conflict or you could have appearance of conflict.

So I think that if you looked at all of the numbers of those it would be an interesting project, but that's not, not my agenda. But in such an assessment where you look at severe outcome and then you look at probabilities for severe outcome, you can begin to quantify those and actually put numbers on that. So this would be a quantitative risk assessment.

And again, you can use a similar graph of, similarity graph with probability and instead of having just broad categories, you actually have

0113

quantities. So you can have probabilities that go from 10 to the minus 7 to 10 to the minus 1, or whatever they are, and severity measured in a more quantitative way. And this generates a curve based on risks you don't want to take or risks you need to deal with and risks you don't on the other side of this.

But quantitative risk assessments, and this question I think was brought up by Dr. Benet yesterday, is -- also has uncertainty associated with it and any number you get by putting such a quantitative risk together involves some level of

variation and if that variation is large, your ability to trust that is less. And so that needs to be taken into consideration.

And finally I want to talk about what's called probability -- probabilistic risk assessment, and people may have many different terms for this, but this has a number of features. And so what does this mean. And this is taken from presentations by NASA that uses this type of analysis a lot.

So some of the parameters are it

0114

includes the uncertainty of quantitation, which we mentioned before. It models for unknown information. It assumes that you can't really estimate most of the P values. You need some way of estimating them or guessing them. And it involves not looking at each mode independently and ranking it, like you might do in an FMEA, but actually trying to integrate all these different risks into overall risk judgments.

So the modeling part is interesting, how are these models. So for NASA in modeling the space shuttle, again published in Aeronautics journal, they used two methods. One is similarity, so

similarity is you have a component.

So the modeling part is interesting. How are these modeled. So for NASA in modeling the space shuttle, again, published in Aeronautics journal, they used two methods. One is similarity. So similarity is you have a component, you have no idea what the risk is, you look for the closest component like it and then you look at the risk and the probabilities you know for that and then you

0115

extrapolate.

And then what they called at least, and again, there may be different names for this, a sort of more structural probability analysis and that is you, more from first principle, look at all the variables you think matter for this component, you vary them with simulations like Monte Carlo simulations and then you derive numbers theoretically for the risk of these components.

So, again, taken from a NASA publication, just to go over some of the general inputs in this. So here is the space shuttle and it's, all of the different components and then some areas which or which may not contribute to failure.

Then selecting one of those elements, manifold, a manifold weld failure and then looking at all the different ways that failure could impact ignition; is it small enough not matter, is it detectable, and then making some sort of logical graph based on the role of that component. Then inputting, then taking all that data and putting it into a tree that actually assigns P values to all these things and

0116

looks at how they affect success or failure of the mission.

And again, the probabilistic issue would be you're looking at the distribution of the initiating event, not just its frequency, and you're also adding many things to the equation, tests, modeling, similarity analysis to try and make these estimates. And finally, you're integrating all the different components, be they a failure of a manifold weld to a sealed failure and getting an overall idea of the risks associated with the shuttle.

So, how can we apply this to complex products? So again, I've shown this slide many times before, a lot of proteins have a lot of

complexity in addition to their primary sequence. There are many different ways of combining attributes to give you a large number of possible parameters and combinations, how would you deal with this.

So, I'm quoting a humorist who says some problems are so complex you have to be highly

0117

intelligent and well-informed just to be undecided about them.

And I think this problem kind of enters into that domain, but nonetheless, let's think about it a little. So for probabilistic risk assessments, people tend to think that lack of data is a reason not to perform one.

But most people who do this, and again maybe they want -- business would argue the exact opposite is true, that a probabilistic risk assessment is in fact desirable when you don't have exact data. It's generally used for low probability, high consequence events for which you don't have enough statistical data and enough data -- if enough statistical data exists to fill in all the trees, then you don't need to do this. This

is really where you're lacking information. And again, this comes from a NASA quote.

Also from Bilal Ayyub, who's worked with the agency previously on some risk assessments pointed out to me that even if you don't know things and it's not that useful to predict, often when

0118

you're dealing with large amounts of complex data, you need a way of organizing it at least. And as a minimum, such a risk assessment begins to look at how to organize all this uncertainty associated with complex products.

So, again, taking a mimic of the space shuttle diagram and starting off with a protein, so again, you have what its structural attributes are at release, the expected stress effects on that, to look at just the at release ones, you have issues involving container closure, excipients, impurities, primary structure, three-dimensional structure and non-aggregation and quadinary structure, and again, these are select examples, these trees would be huge.

Pick one of them, primary structures, you may have glycosylation, deamidation, oxidation,

glycation, truncation, on and on for any change that could exist associated with any one of the amino acids in this structure.

Take one of them, oxidation, which often happens at methyianines, you have encytes in your

0119

protein that could be oxidized and then each one of them becomes a potential parameter in this assessment.

And then you go through the next stage that NASA did for the space shuttle is what's the impact, what's the tree involving the outcome and such. So you have an oxidation at site one, does oxidation not matter. And for many proteins in many situations it may not matter at all and if the answer to that is yes, then you're okay. If the answer to that is no, then is the level low enough not to matter. Again, a very low level of oxidation, even if it does matter, may not be important.

So both of these would lead to acceptable product, despite the presence of this change. Is it detectable. If it's, you know, not at a low enough level not to matter. If it is, then

it's an unacceptable product but it's batch failure, which is again, it's not the best outcome, but it's not the worst outcome.

Finally, if it's not detectable and it

0120

matters and it's at a high enough level, then one would be in a position of clinical failure. So, you have three potential outcomes depending on these relationships.

Now each site would have many relationships, so for safety you might have to worry not as for that activity, but worry about, for instance, immunogenicity. And again, similar questions, is it low enough not to generate immunoresponses. Immunoresponses have high zone tolerance, not that I think we'd ever want to use that, but nonetheless, there could be too high to generate an immune response.

Does the immune response have any clinical significance or not and again, is it detectable at a level where you get an immune response. And again, for each of these you have different outcomes from acceptable product to batch failure to clinical failure.

So going back to the activity diagram, you can then organize that into an event tree and try and quantify all this. So the frequency of an

0121

oxidation would have some probability at that site. If, in fact, it had no impact and then the other variables don't matter, you'd have one scenario and your end state would be acceptance.

If, on the other hand, the inverse of no impact probability, in other words, a probability of an impact, you would have a second scenario but if its level is too low, and again, down means no to these questions, so, in these, then you'd also have acceptable product.

On the other hand, if you would have a product that was above the level that would matter but was detectable, you'd have a batch failure and that would have a probability. And finally, the probability of not being able to detect a significant level of oxidation that had impact would be a clinical failure.

So the big issue here is not how to organize this. You have hundreds of attributes, you have hundreds of trees. It's how to fill in the

blanks. And I think that's where the dilemma for these complex products exists.

0122

So how do you assign probabilities to initiating events? That may be somewhat easier sometimes and certainly how do you apply probability to these event trees that the outcome of these initiating events, so there's actual data. You can have very little actual data for many products. Certainly novel products you're not going to have a lot of data.

There's a similarity method that NASA used, and so same product, but in non-clinical models or other models and this might be similar to a comment that again was in a NASA publication on this, is that they said that if they don't look at the ground tests, this is almost useless.

If you just look at launches, your numbers are so low it doesn't mean anything. But if you take all the ground tests and you use that information, you become much better able to assess. So again, all the different models you might have about the product.

And then with a different product, maybe

you can extrapolate. And this would be a much

0123

tougher argument. Certainly the agency doesn't like the idea of extrapolating from different products, but there may be ways, again, assigning greater variability since you again, you must control not only what you put in but how much weight you assign to it by what you think the variability is, can you also use data from related products or components.

And the last possibility I want to throw out, because I think there may be a time when we know the 3-D structure of every receptor for every protein and you just model what it looks like and you get a probability of an interaction, but I don't think we're there for a long time, I think we have to live with similarity for the time being.

And again, a slide I've shown before, for product, itself, you have whatever clinical data you have, but you have a lot of data from developmental lots, the lots you threw out that you used in a variety of assays and there's information there.

But again, you would want to use other things and so assessing relatedness or -- of a

0124

related product for assigning probability so you might have products that are different with the same primary sequence, you might have products which are different sequence but align in domains that matter or don't matter for the mechanism of action, and there's commonly people do molecular biology use blast searches which are ways of looking for small segments of amino acid similarity.

Would it be a value, and I throw this out without knowing, if you looked at every oxidation, you looked at the sequence of amino acids flanking that, is there any correlation between that?

There's probably certainly correlation with three-dimensional structure and accessibility and again, that would be the next level. Do you have protein structure databases and I say T cell like because T cells recognize primary sequence in the immune system and B cells recommend -- recognize three-dimensional structure, that's an analogy, not a way of analyzing this.

But protein structure databases, you

0125

know we've heard work by Stephen Kramer about using molecular descriptors, small pieces of structure to predict chromatography, could you use small pieces of structure in this way.

And there's certainly one example which again is experimental, but in Silico, ways of looking at immunogenicity based on binding to MHC molecules which are key determinants in immune responses.

Each of those things has much larger uncertainty than the agency would accept to make an extrapolation for a product. But if you were doing a broader risk analysis to look at this change-over to a lot of things and you associated the uncertainty with each measure you used, would this information be useful in that way?

I want to take a moment to talk about Monoclonal antibodies. Yesterday when we talked about quality by design, we talked about platform approaches as one strategy to help deal with developing these products and minimizing some of the work involved.

0126

And so the earliest Monoclonal

antibodies were murine, 100 percent of the sequence was non-human. They had huge amounts of immunogenicity unless they were one time products like, okay, T3, the first licensed therapeutic antibody or limited use, it wasn't very useful, these antibodies.

Then through genetic engineering they were made chimeric where most of the antibody was human and just the variable regions were mass and these actually faired much better in terms of immunogenicity, and then eventually to push a good thing forward, although how much of an actual reduction in immunogenicity this does is not so clear, is to humanize them, to basically make the entire antibody human structure except for the small amount of amino acids that determine the binding sight, in which case 95 percent of your protein is human. And then, again, they are fully human antibodies, but the variable regions are probably antigenic, too, and different, so I don't know if there's, you could model those necessarily better.

0127

But if the humanized antibodies, if you look at what the agency has approved, talk about

this, so 50 percent of licensed Monoclonal antibodies are IGG 1s, and presumably many of them probably come from the same framework that's shared, so is that going to be true of new antibodies.

Probably people are going to make more sophisticated things, but it's probably a good estimate that more than half the antibodies under development, and you're talking about a few hundred are humanized or human IGG 1s, a huge wealth of product that shares 95 percent of primary sequence and probably has a similar range of heterogeneity.

And again, since I mentioned before valuable human in vivo data exists for some examples of Monoclonal antibodies, often IGG 1s. There's some examples, certainly at least one case made public so far of looking at PK not just for the presence of antibody, but for mass spectrometry of the molecular weight which can give you Glycoform variance, so you can get the PK not just of the antibody, but of five or six or more Glycoforms

0128

variants and then you can look at how different the PK for each component of those with the same study that you would do for PK anyway, just a different

way of analyzing the product. If that data existed for lots of IGG 1, that would be a wealth of information about PK effects.

Large safety database for shared attributes. I don't think this is necessarily compiled together, but we know there are this many IGGs in the clinic and there are this many licensed IGG 1s and many of the side effects are primary mechanisms of action which is what it binds to, so it wouldn't count. But probably there's still a large safety database that the rest of the molecule. And since these are in vivo proteins IGG 1s, you can look at polyclonal proteins and look at all the range of oxidations and Glycoforms that exist in vivo, now those may make a difference and you don't want to say because they exist in vivo that's your attributes base.

But it certainly tells you from a safety concern if this variant exists in vivo at a certain

0129

level, that a certain level of confidence associated with knowing that for these products. And again, specific mechanisms matter.

I'm not saying this is a general

criteria for everything, but if you start assembling this and looking at risk, it can potentially be a very powerful tool.

And then antibodies are often re-labeled and looked for imaging. The problem is most of these aren't humanized because people want very fast clearance for these as opposed to not. But again, the idea of labeled product gives you an idea not only of their systemic PK for Glycoforms, it may even give you an idea of their tissue distribution.

So, again, more and more information about this related group of products.

Also, a lot of bioassays relate to the primary mechanisms of action, they would not be shared, but on the other hand, FC receptor binding and effector functions which are the sort of backed on to the antibody, those are all shared and the assays that are looked at are now different by

0130

different companies, different ways, but again, there's a wealth of data that they have, mostly not public, but that industry has about the effect of variations on all these assays.

And finally, I think if you group this

by risk you would certainly need to categorize them by target interaction and obviously the same target would tell you the most, but a soluble target would certainly have different risk factors than a cell expressed target, whether that target signals or not, where that, what target, where that tissue and the accessibility of that target and the role of effector function, the mechanism of action and finally cross-reactivity of the particular binding.

So there are a lot of other product specific factors, but if know what they are, you can begin to try and assemble this map.

So, I think that this is a very complex process. I don't know if this is necessarily a way to help these products move forward or not, but I think it's something that needs to considered and certainly one could tell the blanks as best as they

0131

can and it may just turn out to be an organizing structure for data as opposed to something you can use for answers, but even that has value.

Now what about less than that, what about less than the whole nine yards? Is there a role for risk assessment short of a full

probabilistic assessment that really takes every structural attribute and thinks about it.

So, so even without a, it could be even without a prayer, but even without a probabilistic risk assessment, there may be alternatives to do that and I'm going to put some more quotes here.

That nothing is more difficult and more precious than to be able to decide. So we make decisions at the agency all the time and we often don't have as much information as we would like. Are there tools that could help make those decisions better even that are imperfect tools.

And, again finally, a correct decision is wrong when it's done too late, by Lee Iacocca, but again, there's real pressure on us to decide and we need to use the best tools we can.

0132

So, we mentioned FMEA, so Patrick Swann prepared, and I think he took this from things industry's presented, so not on, you know, only his input, but one could look at product quality in a broad way in an FMEA and assign relative severity, you know, on a scale of 1 to 5, 5 being worst, then this sort of semi-quantitative thing, and you could

look at the, in terms of the occurrence of a particular problem, no known occurrence possible to almost certain documented occurrence, severity, no effect on performance to linkage directly to an adverse effect. And obviously possible linkage to an adverse effect is somewhere in the middle there and then detectability, is this something that a lot release would show so you'd know it all the time, is it something characterization would show, is it something that a good, you know, QBD process would make sure it doesn't change. So I think that, again, there are lower levels at which such things could be taken on.

And I wanted to talk a little bit about combination products and the way the agency

0133

determines combination products, where they go, this is both a logistics issue and also a resource issue is primary mode of action.

And I think that this is, again, driver of jurisdiction between FDA centers, toxic component would override a targeting component and there are a variety of information that's used in assessing these primary mode of action decisions. And

certainly this has a lot of administrative ease and may be the best choice for the agency. Certainly now because risk assessment is relatively new and how to apply it would be tricky.

But I would argue that risk assessment is really the way one should allocate resources, because that's what it's designed to do, and not necessarily just primary mode of action. And I'll give you an example.

I hate using military examples, but we'll talk about a cruise missile. The primary mode of action of a cruise missile is a bomb, its payload, TNT or C4 or whatever that is, I don't know, and there's risk associated with that.

0134

Would it go spontaneously, you know, is it linked I guess to its fuse around a variety of risks associated with it and the manufacture of it. I think probably the greatest risk is that it doesn't work for some reason, but there are risks associated with that.

On the other hand, the cruise missile has a guidance system, you know, a propulsion system, probably orders of magnitude more moving

parts than the actual explosive. If the explosive fails, most of the time you would leave a lack of efficacy. If the guidance fails, the cruise missile hits a hospital instead of a target it was intended to, so even though the payload is the primary mode of action, I would argue that if one did a risk assessment on a cruise missile, you would spend a lot more time error checking the guidance system than you would spend error checking the explosive.

And so, again, I think whether, whether this makes sense for combination products or not, I think it makes sense about how we think about complex products. And so to pick an antibody

0135

conjugated to a toxic moiety, which is a common product being developed now, if you think about such a product and you make an event tree and you think again about these scenarios, conjugated to the antibody is just the product itself, linkage, is it okay, is the conjugate separated from the antibody, is the antibody okay, does it target where it's supposed to go and is the toxin okay, does it work, is it toxic, does it deliver. And if there's a failure in any of these things, could you detect it.

So if you think about all these things being okay, obviously you have acceptable product. If the toxin is not okay, the likely outcome is that that -- is that there, the batch would fail if you can detect it. If you can't detect it, you would have a product that wasn't efficacious.

If the Monoclonal antibody failed, if you could detect it, again, you would fail the batch. If you couldn't detect it, you might have all this payload delivered to the reticular endothelial system in a bolus and have rather significant toxicity.

0136

Finally, if the conjugate breaks up systemically, which is a problem with both components, a shared problem, then you have free toxin and free antibody, again, a likely source of toxicity if you used a very toxic component which is what you tend to do when you can target specifically.

So again, I think how we think about combination products is tricky and is there a way to do a probabilistic risk assessment. I think for making standard jurisdiction cuts it would be very

hard to do this, but I think in terms of thinking about these products, risk assessment's really the way to think about what effort needs to go into a product.

I want to throw out a, sort of Rube Goldberg product and this is very artificial and you can sort of make fun of this example, but it's off the cuff.

So, this is endothelial cells lining a blood vessel and you have atherosclerosis, you have inflammatory cells, they are releasing enzymes and

0137

other things and the endothelial cells around them are expressing receptors based on inflammation, such as ICAM one or something like that. And then you have super duper product, which is a magnetic bead so that you can aid in its delivery and retrieve it, which has cells associated with it that are genetically engineered to be resistant to a toxin. That toxin is on this bead and is released by inflammatory enzymes at the site of inflammation and is able to endocytosis debris to prevent it from becoming clots and to repair damaged endothelial cells and it has a targeting mechanism to inflamed

endothelial cells.

So, this goes to the endothelial cells, the enzymes release the toxin and release some polymer which is enzyme sensitive that keeps the cells associated with the beads and then the toxin kills some of the inflammatory cells and these replacement cells endocytosis the damaged cells and maybe repair the endothelium.

So, very science-fictiony, but the truth is we don't know what products are going to look

0138

like in 10 years and there may very well be products with lots of moving parts and lots of complicated scenarios and I think both as industry manufactures them and thinks about the risk associated with the different components, the agency also needs to think about how to review them.

Because, as always, we're, you know, you've heard we're resource limited now, you know, hopefully we won't be resource limited forever, but it's hard to think we'll have all the resources we need. So there's always going to be some question of how to make choices and in any situation.

Talk a little bit about what this would

mean for the future and again, this is an awareness topic, so we're not asking specific questions, but just to think about this. So I'm, there are many programs that have started to deal with this. We've heard about inspections. (Inaudible), actually was involved in, working with Bilal Ayyub when he was at the University of Maryland and there's actually a draft report which came to the agency on Transdermal patch risk, or risk assessment.

0139

So, some of these things have started and he was certainly interested in conferences on applying risk management to pharmaceuticals. So there's some previous interactions.

Also, you know, in discussing with Helen Winkle, we've talked about, you know, future education on risk management. The OPS talks a lot about risk-based things, we need to really understand how to better use that.

And then how would we manage this for complex product, is it worth doing at all, is it worth doing in a limited way, are there pilot things like antibodies where you have so many shared things that maybe it's a good target to start with to begin

to look at how to deal with this.

And again, if an antibody platform is used, how would you, how would you best work that out and who would do it and how would it be organized. And I have a homework assignment, I guess, and this is my homework assignment.

This imaginary product I made, so I expect a fault tree analysis from all of you with

0140

the probabilities for every possible thing that can go wrong and 5:00 p.m. today.

Anyway, no, I'm just kidding, but I do think that, that the homework I would say is what's the relative importance, we had this question yesterday which was deferred, how much should the agency be involved in quality risk management in cases of limited resources and maybe it shouldn't be the agency, maybe this should be something that industry should be doing together with the agency, but how much effort should go into this, are we right for doing this for complex products and what are the potential benefits if they do this.

And one thing I'll mention which I think is something I've mentioned that other -- this

depends on a lot of sharing, because just like NASA's examples, it's the tests on the ground that have the volumes of data. The launches are few and far between and so for pharmaceuticals, it's going to be sharing of information for risk assessments and that the feel is that the risks of sharing this in a separate risk assessment are lower than the

0141

risks of not using all this information.

DR. COONEY: Steve, thank you.

I'd like to take a few minutes for comments and ideas from the committee.

It certainly seems that there are multiple questions here that you've put on the table. One is around the need for and appropriateness of risk assessment that can be used in risk management. Another is how you generate the knowledge to populate that approach. And a third, at least a third is who would participate in this exercise, because I think as you appropriately pointed out a moment ago, there's part of this that the industry has a unique knowledge of, particularly in the design and synthesis of, manufacture of the products and there are parts that the agency has

some unique experience with.

Mel.

DR. KOCH: Yeah, I'd like to say that it was a very enjoyable, relaxing presentation.

What I'd like to do, though, is the molecule you have up there, very sophisticated with

0142

the magnetic particles and the protein and everything. Taking two or three steps back, it almost looks like a formulation and when you think of it with the excipients and the more we learn about interactions, et cetera, I think one could expand on the diagnosis of a complex product and begin to look at some of the formulations of things of what we used to think as simple, simple molecules.

DR. COONEY: Ken.

DR. MORRIS: Yeah, definitely yoga.

But the question I have is, you know, as much as I like the ab initio approaches because you say it's going to be a few weeks before those are all done, is there an analogy to be struck here with the small molecule tox project that's ongoing between you guys and academia, essentially, to say,

you know, given the, I'll, granted there's probably a paucity of some data, but given the relative success of that approach and given the lack of the ab initio understanding, is that really how to have to start to generate what we would call the

0143

short-cut order of magnitude models before you can start to concentrate on more mechanistic, mechanistically-based risk assessments.

DR. KOZLOWSKI: Right. I think certainly a true first principle risk assessment I think we're really far away from.

The question is using similarity tools, you know, would be good and I think if that's similar to small molecule approaches, then it's a reasonable way to go.

I mean I think, there's certainly things for, the agency for a long time has been interested in comparability and Tony Meyersis was involved in suggesting comparability databases that industry share and a lot of those things don't always move forward. And sometimes it may be, you know, agencies, again, overworking doesn't push it, but I think also there's a resistance to necessarily share

some of these things, and I don't know what the results have been with small molecules.

DR. MORRIS: Well, I think it's actually your program, right, I mean this is --

0144

DR. WINKLE: Yes, where we've been looking at the tox studies and putting up tox information and stuff like that to use for comparability and stuff.

DR. MORRIS: Right, and actually using your data, I believe.

DR. WINKLE: Right. Right. So I don't think Steve is completely familiar with that. But I agree with you, I think it's another part, but I think you're right, I think there are some things there that would be very relevant to us if we were to have had.

DR. COONEY: Meryl.

DR. KAROL: Thank you for the example because I really appreciate something about immunology coming forward.

My question is how do you evaluate the quality of the risk assessment? It's going to be so complex, how do you begin to evaluate how successful

it is?

DR. KOZLOWSKI: Well I think the true test is its predictability, but that's obviously,

0145

you know, information that you would only gather way, way after the fact. And I think one of the ideas, again, as presented by people who do this so they are in some sense marketing what they do but is, is that often information isn't useful because of its organizational status, that it's there, but you really don't see it.

And one thing about these risk assessment methods is whether or not they become predictive. They first become organizational, right. You start looking at all the different attributes and you may collect a lot of data that you wouldn't extrapolate from, but you would say, you know, in hundreds of methylamine oxidation in this domain of an antibody, you know, nobody's seen anything.

Doesn't mean we'll will, we won't, we'll say the next one doesn't matter, but it changes the way you think about it.

It may not be predictive yet, but it

begins, and again, this risk assessment always happens, I mean, you know, sort of going to say I'm

0146

talking about risk assessment, you know, and I've never even played a risk assessor on television, but the fact is anybody who manages anything is doing this. It's just you do it anecdotally, your reviewers do it, so we're doing it all the time. There's organizing and in some way, now maybe the effort and expense of organizing it this way is more than it should be. Maybe there should be simpler ways of organizing and sharing it.

DR. COONEY: Cynthia.

DR. SELASSIE: Yeah, you know, with all this data that you're collecting or could collect --

DR. KOZLOWSKI: Could collect.

DR. SELASSIE: -- like looking at the blast sequences and all the descriptors, have you all thought of using something like multi-variant analysis and MPLS to solve?

DR. KOZLOWSKI: I think all, there are a lot of potential tools that could be used to try and correlate what matters and that would probably be a good approach, too.

DR. SELASSIE: Yeah, because it would

0147

give you a level of reasonable predictiveness.

DR. KOZLOWSKI: Right, or at least tell you whether something matters in a lot of cases.

DR. SELASSIE: Yeah, right.

DR. KOZLOWSKI: So, again, I picked an example because again we looked at who, who looks at complex things without data. And so NASA does this, I think the Nuclear Regulatory Commission does this, there's a number of groups where they have extremely, you know, catastrophic outcomes and limited data when they replace a system, so they use a tool. It doesn't mean that's the best tool. I bet, you know, current academics on this would say those papers are old, you know, there are better ways of organizing it.

But it's the conceptual issue, you know, should there be a systematic way of trying to apply this, not just simple questions, you know, this company has been audited three times and failed once versus a company who's been audited six times and failed not at all. And the importance of the product, again, those are very important

distinctions, but could you start making it for actually quality attribute decisions.

DR. COONEY: A couple of additional thoughts, Steve, on this.

One is it seems to me that first of all developing methodologies for risk assessment in a formalized way is a very positive thing to do. It just makes fundamental sense.

However, I, it should not be done I believe in isolation and should be done collaboratively between the agency and those, those who are dependent upon the interaction with the agency and the industry in particular.

And it seems that approaches using CRADAs which are I believe having a very positive impact in the area of PAT, for instance, in quality by design, would be very appropriate here.

Second, that when you think about using these structures for risk analysis, they can be very useful for enhancing the quality of your design of experiments.

So not just using them in retrospective

analysis, but actually to assess where, where the greatest risks are and then to use that to put the experimental work and to direct it to the hot spots.

DR. KOZLOWSKI: Right.

DR. COONEY: As opposed to those things that are, that are less important.

DR. KOZLOWSKI: And I think even one could say directed where the uncertainty is, because where you know there's risk, may be easy. And where you know there's no risk is easy, it's all those holes in the system which make it not useful for prediction, so there may be so many uncertainties that it's hard to prioritize, but then there may be some sense of what's an unknown that's more likely to be -- we would think in some general sense is more likely to be associated with a risk.

DR. COONEY: Well, you had the opportunity to identify the points of uncertainty linked to the points of high impact, which is the combination that you point out that you, that you want.

So it seems to me that the use of these

0150

frameworks very early in a project is much more

desirable than trying to use it retrospectively at the end; hence, the suggestion that they be done and try to develop them collaboratively with those who are in the early stage of many of these projects.

The last point is the, some of the methodologies that are corelative as opposed to mechanistic and the extent to which you can base your analysis of uncertainty, assessment of probability distributions of the relevant parameters around mechanistic considerations I think is far more powerful than simply corelative approaches, which again fits in with other initiatives within the agency.

Are there any additional comments from the committee? There seems to be an encouragement to think further along this path.

DR. KOZLOWSKI: Is it fair to conclude that there's encouragement and obviously involving industry in doing this in a general way.

DR. MORRIS: Yeah, I would say not to ignore academia in this, but I think the, I think

the reality is is that you're, the joint programs you have on small molecules are largely academic and

the agency using industry data in a blinded fashion in many respects, so it's not always easy to get all of the industrial folks to commit the kind of effort that it takes to collate, share and blind and do all of the work that has to go along to it, but if you already have some of it, that makes it a lot easier.

DR. COONEY: Okay, thank you. We're going to take a break for lunch, but before we do that, one, no two announcements.

I've already noted that we will re-convene promptly at 1:00 for the period of the public hearing, and immediately following the public hearing period, which I believe will be brief, we will have then the discussion on the first topic of this morning on highly variable drugs. So, to please keep that in mind.

The second schedule issue, we are going to swap the discussion of critical path initiative and the discussion on nanotechnology because of some individual scheduling issues and we will begin at

0152

2:00 discussing the nanotechnology, its use and definitions, followed by the critical path initiative.

So, if you would keep that in mind as you plan your lunch and your afternoon.

Now, I'd like, I have one, one announcement, statement to read. No this isn't -- one additional point to make. Just to remind everyone that the lunch break is not a period for extension of discussion of the advisory committee activities, but it's a time to discuss all those other things that you wanted to talk about, so I would ask you please not to discuss either amongst the panel members or between panel members and guests the topics of the advisory committee meeting.

And we will re-convene promptly at 1:00.

(Lunch recess taken)

0153

DR. COONEY: I'd like to call people back to order. Before beginning the 1:00 open public hearing, I'd like to read the following statement for individuals presenting at the public

hearing.

Both the Food and Drug Administration and the public believe in a transparent process for informed information gathering and decision-making. To ensure such transparency at the open public hearing session of the advisory committee meeting, FDA believes that it is important to understand the context of an individual's presentation.

For this reason, the FDA encourages you, the open public hearing speaker, at the beginning of your written or oral statement to advise the committee of any financial relationships that you may have with any company or any group that is likely to be impacted by the topic of this meeting. For example, the financial information may include a company's or a group's payment of your travel, lodging or other expenses in connection with attendance at this meeting.

0154

Likewise, FDA encourages you at the beginning of your statement to advise the committee if you do not have any financial relationships.

If you choose not to address this issue of financial relationships at the beginning of your

statement, it will not preclude you from speaking.

We have I believe one, one speaker who will, Laszlo Endrenyl who has 10 minutes to share with us some thoughts on determination of bioequivalence of highly variable drugs. Laszlo, please.

DR. ENDRENYL: I appreciate the opportunity to be here and to make a presentation. I have no financial interest involvement. I came out of my pocket.

I would like to consider two issues, what kinds of replicate designs should be applied and whether there should be a constraint on the estimated ratio of geometric means that is under GMR.

I would like to skip these two scissors slides, they are, they involve definitions of the

0155

average bioequivalence and unscaled and scaled average bioequivalence you have heard about.

I would like to turn to the question of experiment and designs. The scaled average bioequivalence we are talking about and by definition or dogma, it's referenced product scaled.

One can argue about that, but that is in the general commercials, so it's the within subject variation of the reference product according to which we scale.

So for this purpose, a three-period design with single cycles is sufficient. You replicate the reference product and you can estimate the evidence object variance from that.

(Inaudible) an additional goal is very important and that, this design is unable to consider, namely, to compare variations of the two drug products. This way one could identify highly variable drug products, that is to a certain -- where one product has a higher variation or a substantially higher variation than the other and it's not necessarily the test product that is bad, there have been examples, strong examples when the

0156

reference product was a bad product and the test product was much better.

So, highly variable drug products ought to be investigated and ought to be identified and therefore this design in which both the reference and the test product are replicated would be able to address this issue. And this is a very basic

fundamental issue in my opinion.

Even better is the four-period design which permits the estimation of the two within subject variations for the two product in the same subject and average those into your estimates, you know.

And that is achievable in a four-period design in which those products are replicated in each subject (inaudible) or the other. Also, some outlying observations can be identified with that.

Moreover, since the three- and four-period designs, design require approximately same number of observations, actually the four-period design can afford a better estimate of the RR within subject variation because there are

0157

more of them than the three-period corresponding two sequence design.

So, there is a strong merit in my opinion to consider the four-period design. It has several merits and they ought to be consider, in my opinion.

Now the second issue we heard about that Dr. Benet was concerned about the large possible

deviations between the logarithmic means, estimated logarithmic means. And the concern as he expressed it as political. It has to do with interpretations of the results to physicians and patients and that's a varied and strong reason.

Now, when we have highly variable drugs as we do here from 15, 35 to 50 percent, obviously the distribution is wider and wider and as you get a wider distribution, the difference between the logarithmic means also gets, it fluctuates the estimated value.

Now as you go higher variation, it fluctuates more, like that, so it is, indeed, possible to get large differences. Now if

0158

artificially you cut down those differences, the estimated differences, then you especially truncate the distribution.

Now there is a line in your handout which is not on my slide and that in my opinion is important, that by doing this kind of truncation, you actually are committing a scientific faux pas. The outcome is scientifically incorrect.

Not -- and I would like to emphasize.

Now this was eloquently demonstrated in Dr. Haidar's slide when he showed you the results of coefficient of variation of 60 percent, that's high variation, and showed the results with or -- and without a GMR constraint. The GMR constraint dominated the result, therefore, in effect, the outcome was a GMR criterion, that is, you want to determine the, that there is a, the results would not be different and not a bioequivalence criterion. That's in my opinion is very wrong, very incorrect.

Moreover, we could go back to the basics of the purpose of bioequivalency investigations, the goal. Is it political control or is the goal mainly

0159

a (inaudible) to service that. Now biointernational -- in '94, the meeting, very diplomatically this determined that it should serve both goals.

Now, ever since there has been an awful lot of confusion because the two goals require totally different conditions and considerations. For quality control, you would like to ask for high sensitivity and high statistical power.

For therapeutic surrogate, you would

like to have clinical relevance. They are very different. The twos are very different. For example, in, for quality control you would like to have young, healthy volunteers in the sample because they provide certain results. Clinical relevant, you would like to have heterogeneous study propagation and this was eloquently argued by Dr. Levy some 10 years ago when he talked about the (inaudible) of bioequivalence because in his opinion it wasn't irrelevant because it was on the surrogate side.

For quality control, sensitivity, you

0160

would like to ask for single dosing. Clinical relevance, if appropriate, you would like to look for steady state.

The difficulty with current guidance is FDA and other, that they serve board's masters and therefore, indeed, there is confusion.

Now, if you consider the, could be able to separate the two goals fairly easily, because if you consider the investigation of generic drugs, then you probably would like to put emphasis on quality control and this could include various

conditions, but essentially for generic drugs your primary goal is probably quality control most of the time, verse when you develop new drugs, then your aim is to think about the therapeutic conditions and therapeutically the new product of the same drug should have the same effect. So, for the development of new drugs, I think the emphasis ought to be mostly on the therapeutic side.

So, as a result, I think that applying the same condition, the GMR constraint makes, is irrelevant to the therapeutic consideration. It's

0161

even for political purposes, not only is it scientifically wrong, but politically irrelevant because it's within the same product, there was one product after the other of the same drug. So here we have again the one size fits all, or both conditions, or we are back to a problem because of this kind of confusion.

So at least in my opinion, this or our opinion, the secondary condition, secondary criteria ought not to be generally involved, maybe on the quality control side for generics. Politically, yes, scientifically, no. For new drugs, new drug

politically, no scientifically.

So to conclude, in our opinion, three- or preferably four-period studies in which both products are replicated are advantages and the four-period design in my, in our opinion is more favorable than the three-period.

In four-period, you can get away with 24 subjects and not 36 and so you have essentially the same kind of results consideration, but better outcome. And the concept of GMR in our opinion

0162

ought not to be introduced, not even for the sake of politics or publications.

Thank you.

DR. COONEY: Thank you very much.

Do any the committee members have questions on this speaker?

DR. MEYER: Maybe just a quick one, Laszlo. FDA I think is going to propose a Sigma WO of .25.

Do you concur with that as a logical choice or not?

DR. ENDRENYL: Not in my opinion. The Sigma W corresponding to a coefficient of variation

of 30 percent in my opinion corresponds to the current definition of highly variable drug, coefficient of variation of 30 percent, at which there is actually continuity, the mixed model would use a constant unscaled average bioequivalence up to that point and from there on there is an expansion.

So, I think that criterion would correspond to the current definition of highly variable drugs, in my opinion.

0163

DR. COONEY: Ken.

DR. MORRIS: Yeah, I'm not a clinician, or a PK person, so you'll have to forgive me, but in the distinction between the criteria for therapy versus quality control, I mean the way I understand it, though, you're doing it for a new drug, the pivotal clinicals wouldn't be a BE type study anyway; is that correct?

DR. ENDRENYL: True, there are clinical studies that are there, indeed. But new products are developed different coating, different --

DR. MORRIS: So either you're saying within formulation changes and things?

DR. ENDRENYL: They are evaluating

against each other.

DR. MORRIS: But by that time I think the therapeutic part is -- I guess my question is.

DR. ENDRENYL: This early in the game.

DR. MORRIS: Yeah, that's basically what I was saying is that the therapeutic value of the compound should be determined by other types of clinical studies.

0164

DR. ENDRENYL: Absolutely.

DR. MORRIS: So I guess then this question is actually more for FDA, but I mean are we, are we -- for the discussion.

Okay, that's fine, thank you very much.

DR. ENDRENYL: Okay, but a different division of FDA as it turns out mostly.

DR. MORRIS: Okay, well thank you, though, I just want to make sure that I understood that.

DR. ENDRENYL: Yes.

DR. COONEY: Okay. No other questions then?

Thank you very much.

DR. ENDRENYL: Thank you.

DR. COONEY: We will now proceed to a period for discussion of the proposed questions from the FDA, which probably should be presented.

MR. UNIDENTIFIED SPEAKER: I think we need to go back.

MR. CONNOR: Hi, I'm Dale Connor, I'm director of the division of bioequivalence in the

0165

Office of generic Drug of FDA and I have a word to the wise when you're in this kind of situation, don't sit in the front row because you'll be, you know, within just a few minute's notice you'll be called to run up and give your comments on something. Always sit in the back.

I just have, first off starting off before we get to the questions, I have a few comments. Laszlo and all the other speakers are always extremely interesting. I've heard them speak on similar topics many times. I'm always quite amazed with the depth of their thought and their insight into this.

Just since it's fresh in my mind, I'd like to do a few comments on some of the things Laszlo and others have said, in no particular kind

of order, just a few things to point out.

There's a practical aspect when, first off, when you look at two-way versus three-way versus four-way studies, and in a sense you know whether you're doing one type of approach or another or when you have the same number of treatment

0166

periods, say, with a two-way, two time -- two times say 20 subjects would be 40 treatment periods and to get the same amount of power for a four-way study, it would be roughly half as many subjects but twice as many treatments for each subject. So, it really seems like it all comes out evenly.

But for those who do these type of studies like CROs and sponsors, they know that these are not exactly -- that when you study a person, an individual more times, they have a much higher likelihood of going out.

So it's not a straight, you know, wash that all things are equal expense, because you actually have to bring in more subject alternates because there's going to be a higher drop-out rate with a four-way study, in a three-way, than a two-way. So it's not safe to say that it's all

equal as far as cost goes because you do have to account for the higher drop-out rate.

So, if you're looking at expense or the number of subjects that you're going to potentially study, it's not strictly equal. So that's just a

0167

little practical thing so you don't believe that that's the case.

Just what, there's some misconceptions about the bioavailability and especially bioequivalence as it relates to how it's used in NDAs and how it's used in ANDAs. And I worked as a reviewer and as a team leader on the NDA side looking at these type of studies in NDAs and I also obviously am in OGD.

So, I have knowledge of how it's used both ways.

First off, there are things that are very late in the development of a product where bioequivalence techniques or types of studies are used.

The most common and perhaps one of the ones that's closest to what some refer to as generic is frequently the, a formulation is developed for

clinical trials and it's used in the major clinical trials. Often it's a smaller scale type of batch or manufacturing. The firm, you know, it has proven that the product is safe and efficacious, you know,

0168

and they believe that the FDA will be happy, but then they go to scale up the product for commercial purposes and they may have to actually make changes in the formulation to get it to scale up to large batch, perhaps it's too expensive, perhaps it doesn't scale exactly as they made it in smaller, so they have to make sometimes small and sometimes, you know, not small changes to the product. And generally in their NDA they will do a bioequivalence type trial to see how the clinical trials formulation compares with the to be marketed formulation.

There's no, there's no legal requirement that I'm aware of that that study passed our strict bioequivalence criteria. It is done in the same, in very much the same way the generic sponsors do it. It may actually be done with less subjects or more, but it is merely a demonstration of how those two products differ and it's up to the clinical

division, both the OCP, which are the biopharmaceutics, clinical pharmacology people and the clinicians in that division to decide whether

0169

that difference that's shown by that study is really significant or worth worrying about.

So it's not a strict criteria that's used in general where, you know, if you're beyond a certain set point you fail and you either have to re-formulate the product or re-do the study.

It is very much a judgment on the -- of looking at the data, if it passes the usual criteria, everyone's usually happy, but if it doesn't, it does not mean that that study is a failure or that formulation can't be approved.

So that's, that is very, very late in the development and is actually quite frequent in an NDA, so that's probably the closest that an NDA sponsor will get to bioequivalence.

Other types of cases are if you've developed your product on a capsule and at the last moment your marketing people say no, we don't really want to sell a capsule, we're going to do a tablet instead and you want to connect the tablet

formulation that you want to market to the original capsule formulation where you've done all your

0170

clinical trials. That's another very, you know, not uncommon thing in an NDA.

So just to say that NDA people, NDA sponsors do do these type of studies, but they don't have the same rigid criteria that the generic sponsors do.

Also, I mean, I found very interesting that Laszlo likes, depicted the split into two categories, either clinically relevant or quality control. I wouldn't use the term quality control because I, I literally, when I conceive of these two things, I consider them as different viewpoints of, to achieve the same end point.

What we're trying to achieve with generic drugs and bioequivalence is therapeutic, in the end, therapeutic equivalence. In other words, the generic switchable product will be therapeutically equivalent to the original so that you can go into, you, as a patient, can go into your pharmacy and without the doctor's intervention, the pharmacist can switch you back and forth between AB

rated generics and ideally you will see no objective

0171

difference in your treatment.

If you're having side effects, you'll have the same amount of side effects. If you're having successful treatment, it will continue as successful treatment. That's the ideal in what we're, what we're trying to achieve.

You can look at this from two different ways. You can say, well, first off, I'm going to do a test that clinically relevant, so I'm going to do a clinical trial with, a comparative clinical trial with patients and see how the clinical response to both of those formulations comes out and see if they match. We do have to do that with certain types of products.

A lot of topical products, locally acting products, we really have no choice because pharmacokinetics and other kinetic or direct measurement type of methods are really not suitable for that type of, for looking at drug appearance at the site of activity or bioavailability, so we have no choice but to do comparative clinical trials which we term bioequivalent trials with clinical end

points.

Those trials are extremely difficult to do. They are extremely large. They are not single dose studies, they are often studies that go for weeks or months. They involve patients and they can involve as many as 6 or 700 patients. So next time you get a generic cream or ointment, you should -- a new generic cream or ointment, you should appreciate what the sponsor had to go through to get that approved.

Same thing for inhalers for asthma, for nasal sprays, it's a huge amount of data, different sets of studies, both PK and these large clinical trials, so when we look at what, you know, when I look at in that context at 6 or 700 patients in a bioequivalence trial and then I look at, you know, what may be 60 or 65 normal subjects in a normal bioequivalence trial, you know, it doesn't seem all that bad to me.

But one of the things that you have to remember is what we, people use the term a too many subjects or an unreasonable number of subjects, but

that is all, that's a relative judgment.

A firm who has to pay the bills for this may consider any more than 24 an unreasonable number of subjects. Someone else in the academic community may say, oh, you know, 50 is not too bad, but I don't have to pay the bills for it. You know, it doesn't seem like that much to me.

So the judgment, you know, we all have to come to some type of consensus, what is too many. Is 100 too many or do we want to really restrict, you know, the overall sample size, do we want to target a method that can get things done, demonstrate bioequivalence for those products that should rightly demonstrate bioequivalence with some set reasonable number of subjects or, which may be everyone in the room may have a different opinion of what's reasonable. So that's part of what we're doing.

And the ones that go up to 60 percent bioavailability are very much in the minority, fortunately for us. As Barbara Davit showed in the data collection, most of the products that we've

seen come in, successful products that are highly

variable are in the 30 to 40 percent range.

So, this is going to have a little bit of relief as far as number of subjects and expense, but the very few, the small minority of products where you really have 50 or 60 percent, those are the really, really costly ones and that's in a very, very small minority.

That's where this will have the most impact and also, you know, I found it very interesting in Dr. Haidar's talk that the point estimate constraints would actually predominate in that, you know, high percent, because I'm not 100 percent sure that is exactly desirable.

I mean I think the scaling is something that is very appealing and very elegant, but to simply overshadow it with what's admittedly a political constraint doesn't exactly seem to be a very desirable thing. But that's, again, for debate. So, those are my comments on that.

Lawrence, did you have anything else?

DR. YU: I guess, okay, this why you

0175

were picked. I have a number of comments, a number of issues with respect to study design, the variable

Sigma zero, the number of subject, the point estimate for geometric constraint, those decisions which as regulatory agency we will have to make.

With respect to study design, as you can see from Barbara Davit talk, normally right now it's two-way cross-over study design, we call average bioequivalence study. They have to meet the bioequivalence standards with interval which is in 80 to 125 percent.

However, even with that, we normally accept a replicate study design, for example, four-way cross-over studies design and the agency never suggests or never requires that you only can use two-way cross-over study. There's no other study design you should be used.

At this point, I guess this morning we talked about a number of things with your feasible, in terms practicality of study as well as feasibility, cost effective, we were thinking three, three-way cross-over studies. Nevertheless, sponsor

0176

has always options to use others as long as you justify it.

The value, with respect to the value of

Sigma zero and I think the three value we're discussing right now, .20, .25 and .294.

If you assume the CV is about 30 percent, which is the definition cut-off for highly variable drugs, if you use actually in myself, one of the scientifically we discuss internal myself in favor .294 simply the curve will be smooth from, from average bioequivalence to scaled average bioequivalence.

However, if you use that as .294, one of the major drawbacks of the availability, for example, 31 or 34 or 32 as most mentioned, most drugs will have those variability will not have a benefit from this approach. In fact, as Sam has showed this morning of a CV exactly 30 percent, an average bioequivalence is better off than scaling bioequivalence, scaling bioequivalence, the average bioequivalence, so you have to consider that.

In terms of numbers of subject, we have

0177

considered extensively whether 24 or 36. At this point we are suggesting 36, but we, I wait for 24 depend on the committee's suggestion.

Finally, geometric mean ratio

constraint, that's, we recognize, long time ago, this is not a today study, we recognized and statistically speaking this may not be very good choices, but in terms as Les point out, in the communication it makes our life a lot easier.

We were talking about 80 to 125 percent confidence in the four lasted 20 years and I believe, we all very good communicated and we give many, many talks and various, I even don't know how many scientific meetings, nevertheless when we receive certain petitions, always that the different between generic and innovator is 45 percent, 40 percent. In other words, no matter how you talk about, the message is not crossed.

If we have a point of system, make life a lot simpler and then make our communication a lot simpler. So I have to consider that.

With that, thank you.

0178

DR. COONEY: Thank you. I would like to invite comments from the committee and questions.

Art, please.

DR. KIBBE: The temptation to -- well, first, Dale's right, going from two to three to

four, even if you cut down or have the same number of subjects cost you more and you'll lose them. And for a couple of years I was with a company that were doing 10 bio studies a month and to manage that and to manage the people, we, we'd have to, if you wanted to make sure you had everybody left after a four-way cross-over, you'd have to do it at Guantanemo. So, it cost more.

The second thing is I feel a little bit like Joshua at Jericho, I think for about seven years we've marched around Jericho playing our horns and saying we should do replicate studies and we're still waiting for the wall to fall down.

And Les made good points, Kam made good points. It's almost to the point where we should be asking for replicates as a way of avoiding repeated studies.

0179

You say you don't see a lot of 60s and I'll tell you why you don't see a lot of 60s is because when we did them, they never sent them and then they'd re-do them and re-do them and often you, we would actually do a replicate on the innovator and find out that's where all the problems were and

then we'd throw up our hands and not know what to do. So, we've got to get past that.

I think that we should consider also asking on the new drug side for at least one replicate study with the product that they are going to market with and that's because if we're really serious about quality by design, then the innovator ought to care about how available their product is and design with that in mind. And that information ought to be available for the agency years before they have to start adjudicating potential therapeutic equivalence, even within products that the innovator might bring out subsequent to the original one or what have you.

So I would argue that the agency on the new drug side ought to be looking at requesting of a

0180

replicate study in the process to start with.

But I think we just need to get off the dime on this one, I'm not, I don't have any strong feeling one way or the other about the .25, .294, .2876, whatever, I think somebody has to look at all the numbers. I think we did a lot of statistics and let the numbers kind of help you.

And if we really are committed to good science defining the therapy, then why are we being anal with regard to rules when we expect that the FDA scientists and the industry scientists can agree on what the study means and the outcomes can make a decision.

We do it with new drugs before they hit the market because there's not four other companies trying to fight over that part of the market, and so if there were three or four other innovators reaching for that market at the same time, then I'm sure they come down and your a little bit of flexibility on the new drug side would go away because there would be lawsuits and all sorts of citizens petition groups and affected citizens with

0181

certain diseases that would all be campaigning for all sorts of different things.

Let's go back to what we really wanted to accomplish for the last however many years I've served on this committee, and that is to make sure that the decisions we make are based on good solid science, are fair and can be easily applied by the agency in that respect.

Replicate studies get rid of the variability effect because you can tease it out, you can separate it out, you can control for it. And if we don't start to allow the companies to submit either three- or four-way replicated studies, and we don't have to insist on four or three, if they want to try it with two and they think they can make it, good luck to them, but we know this is going to be a lot better advice. They are far better off with 24 subjects in a three-way study than trying to figure out 96 studies -- subjects in a two-way study.

And if they want to turn in a four-way to make the point of how much better their product is, more power to them. So, let's move forward.

0182

Thank you.

MR. CONNOR: Just a correction, or just to make it clear, we accept replicate studies, but it is strictly the sponsor's option at the current time.

So, people are perfectly free, a sponsor at their own choice to put in a four-way, usually it's a four-way that they do, if they choose to do it, a four-way replicate design.

It does seem to make things -- even though we are doing average bioequivalence, we aren't really teasing out and using the inter-subject variability. It seems to have a bit better go at a highly variable drug in passing our current, you know, fixed criteria.

And we've seen, there's one drug I'm thinking of where virtually every sponsor chose to do a four-way replicate design and of course the ones we saw all passed, you know, obviously, and it had been one chose to do a very large two-way cross-over and that didn't pass. And so that we had a lot of discussion with them saying, oh, well, you

0183

know, could I drop subjects and, you know, the usual types of things that people try and do to get their studies to pass when they are in trouble.

But, you know, it really says that, even handled in the static non-scaled way we do know, a replicate design does help a little bit in this.

People mention before other techniques or other ways to address this, like the sequential design which is really, sequential design to me is not an add-on because I define that differently.

It's being able to break your study up into several groups and take a look at one point and see how you've done and then have a decision role, whether you go on and study the next group or not. But if you've met the criteria, then you stop and you don't go any further.

This is used in clinical trials quite a lot and, you know, we haven't, people have asked us to accept this for bioequivalence trials and we've finally said, yeah, we're open to it, send us some protocols, tell us how you're going to analyze it and then nobody does it, so.

0184

But the problem is that's not really a solution for this. That's really more of an efficiency, because we're still going to be using the same criteria and although you'll hopefully focus in on the right number, the right number still will be large of subjects.

And so it doesn't really deal with the fact that I need, you know, to do this right. I need 100 subjects. What the, what the sequential design was saying is I will probably end up with 100 subjects and I won't, you know, overshoot and do

20 extras that I might have done if I just, you know, did it all at once for safety.

So it really isn't the solution to this, it just makes it a little bit more less wasteful, I think.

DR. COONEY: Marv, then Ken.

DR. MEYER: I'd like to get back to the questions that were posed to us. I gather from the way they were phrased we're not being asked whether we need -- whether we're supposed to do a three-way or a four-way and we're not being asked whether we

0185

should do a scaling.

We're being told we'll do that and then I want to ask some details on it, which is fine, I don't have any problem with that.

I think that I'd like to address the first one. I think it's essential that we have scale -- that we have control over the point estimate and I think that those of us who have been in the business for a while and know that every patient advocacy group in the country will ask Dale Connor and Gary and probably the commissioner to come visit them at their meeting and explain why

you're allowing an 80 and a 125 spread on mean values when you've just about convinced us that, okay, 80 to 125 is a confidence limit, but that really means that the means can't differ anywhere close to that and still pass.

If you move to a point estimate that allows 80 to 125, you're going to have hell to pay and I think that the brand names will exploit this and I don't blame them. If I owned stock in them, I'd expect them to do that.

0186

And so I think if you, the closer you can get to the, if you can point to the products that are being approved now and their mean ratios are in the neighborhood of, let's, I'll pick a number, 90 to 110, then that's a reasonable thing you can defend. If you go down to 80 to 125, you're asking for trouble.

MR. CONNOR: Well, I just want to give you a little bit of history. Up until, let's see, January 31st, 2005, which is when the FED bioequivalence studies guidance, which is the guidance that talks about FED -- you know, specifically about FED bioequivalence studies, and

it covers both NDA and ANDA, up until that point in time, those studies were just based on a point estimate criteria of 80 to 125.

So, there was no calculation of confidence intervals. Those studies prior to that date were not powered to look at confidence intervals, so they usually had less subjects than what would be required if they actually had to calculate confidence intervals and pass our usual

0187

criteria.

And that was an evolution. And I don't want someone to quote me in saying that there was anything wrong, necessarily wrong with that or that was dangerous in any way, because I think we made good, you know, solid decisions based on that.

But in the evolution of thought about, you know, what these FED studies meant to us, we finally decided that they were true, not just a supportive study to put peoples' mind at rest about the effect today on bioequivalence, but they actually were a bioequivalence trial in the true sense of the word. And so we evolved to the point where we finally formulated a guidance over many,

many years of discussion, public and within FDA that we would take this to the next level and actually do the 90 percent confidence interval equivalence methodology and bring this up to date to be a real equivalence trial. And that's what we did from that date onward.

So, we do have a history, you know, it's not too many years in the past of actually using

0188

just this criteria, you know, just the point estimate 80 to 125 and not the rest of this. I mean that was the only criteria for passing, so that's just a little bit of history about how we've evolved and that we still have many products on the market, very successful generic products with no problems whatsoever that we're, no problems whatsoever with food that were approved on that basis.

DR. MEYER: But the point is you did see the error of your ways and correct it?

MR. CONNOR: Right.

DR. MEYER: And now you're proposing to go back to that same error and then maybe 10 years from now correct it.

MR. CONNOR: Well, I wouldn't go out and

say that. You would be correct if this was all we were, if point estimate criteria was all we intended to do here with, you know, but this is really kind of an add-on to the more important talk that we're having, which is scaling the average bioequivalence. So it's not just we're only doing a point estimate criteria.

0189

DR. YU: I guess this only apply as Barbara point out the 10 or 15 product which is safer, it's not applied to all the product on the market.

DR. COONEY: I'm going to take just a moment and I'd like to read a, read a series of comments from one of our committee members who was not able to be here at the last minute. These are the comments of Jurgen Venitz.

The proposal should study, bear with me while I read a blackberry, the proposal should consistently refer to the drug product, not drugs, since the high WSV may arise from the drug product formulation or device rather than being intrinsic to the drug or API.

This would particularly be true for more

complex dosage forms such as Transdermal patches, pulmonary inhalation devices, et cetera, that are intended for systematic delivery. It is my assumption that the proposal would apply under these circumstances as well.

The second point is that the proposal

0190

ought to define what evidence is needed to qualify a product as a highly variable drug product. Is the WSVR greater than 30 percent based on a replicate design study. The most accurate way of assessing its value are based on previous bioequivalency, bioavailability studies as part of another ANDA or NDA. Usually it's part of a non-replicate two-way cross-over study where the residual variance can be a poor estimator for WSVR as discussed in the background paper.

The definition of HVDP has to be unambiguous and feasible.

Third comment. Overall I'm in favor of using WSVR as a means to scale the goal post for the test product along with additional constraints on the point estimate. I think it implies that WSVT can be no more than WSVR while maintaining

equivalence of the means. One of the things that always troubles me about IBE as you may remember was the fact when WSV product differences could be canceled out by differences in point estimates between products, leading to cases where a product

0191

could pass IBE and fail ABE as some of Laszlo's cases demonstrated a few years ago.

In addition, I consider the -- okay, formulation, interaction of uncertain -- interaction of uncertainty at best, clinical significance and more likely a statistical artifact. The proposed approach clearly separates the two criteria.

In addition, IBE had the scaling factor equal to the fudge factor whose choice in value would determine F over P of the bioequivalence and there was no rationale way of selecting a value other than considering the bioequivalence consequences.

Nevertheless, I have the following comments about the proposal.

The fourth point, the clinical significance of this widening beyond .8 to 1.25 needs to be reviewed, discussed and approved for

each drug. For instance, does existing ER information such as a flat ER support the notion that S&E are likely unaffected by wider goalposts. I agree with Les' arguments at the previous meeting

0192

that the true NTI drugs are unlikely to show high WSV.

What is the intended use, for instance, are they given as fixed dose or dose titrated. In the latter case I would be more comfortable with widening. Period. What are the stakes of S over E for oncology drugs where the stakes of underdosing may be very high.

High stakes may make us more conservative about why, I think we're getting close, this is fifth, you should at least consider typing the constraints on the GMR such as .9 to 1, rather than the proposed .8 to .12 -- to 1.2. This may be justified based on the previous point.

But I don't have a sense of what the consequences would be in terms of P over F of the bioequivalence in the simulations or the real world. Again, the clinical significance may be the tie breaker.

Six, I need to see more information about the proposed minimum sample size requirement of greater than 36. For instance, WSVR estimate,

0193

powered, cast, et cetera, before I could answer that question.

I assume that the sample size is based on the supposition that WSVT is less than WSVR, equalling the maximum allowable WSV, which would appear reasonable, but other assumptions may be less reasonable, or the sample size estimate quite sensitive to one or more assumptions.

Seven, regardless of minimum sample size, what happens if the study is underpowered, do you need the failure and bioequivalence on the CI, even with the scaled goalpost?

Did you consider increasing the sample size incrementally, if pre-specified in the protocol? This outcome would be quite possible if the original WSVR is underestimated.

Finally, I believe that the current bioequivalence guidance does allow the use of replicate design if pre-specified. Would that still be the case of this new proposal, if this new

proposal were to be adopted? If so, could the current proposal be modified to achieve the same

0194

goal with a replicate design, RT/RT or RTR/TRT, namely, ensuring that WSVT is less than or equal to the WSVR and the GMR passing bioequivalence.

Food for thought.

MR. UNIDENTIFIED SPEAKER: That was a banquet.

DR. COONEY: Ken.

DR. MORRIS: I think I forgot what I was going to say. No, I think the scaling part makes perfect sense based on the concept that actually you had raised which is the therapeutic equivalence. If we're referencing it against the demonstrated therapeutically efficacious reference product, then I think there's no question that that makes good sense.

There, the mean -- I was a little, I had to listen to what Mel was saying, I haven't thought quite about, much about that, that variation, I'll think about that as the discussion goes on a little bit.

The one thing I wanted to raise as an

issue is that if it turns out that, in fact, there

0195

is this relative insensitivity to a formulation with some formulations, does that mean that we now have to start re-visiting things like dissolution specifications that may not be important? Does this open that discussion?

MR. UNIDENTIFIED SPEAKER: That's like another couple days' work of advisory committees right there.

DR. MEYER: No, exactly, I have a vested interest in asking that because I'll, you know, have to sit here for two more days.

DR. KOZLOWSKI: No, I do dissolution all the time and I'm a great believer in the clinical realm of dissolutions.

DR. MEYER: Well, but that's my point, if this is actually a clinical -- I mean if this is actually supposed to demonstrate therapeuticals, then it sort of raises the issue of just as a caveat of what this may be --

DR. KOZLOWSKI: Well the dissolution question actually makes this look, you know, kind of small and compact. I'm pretty sure we don't have

0196

enough time.

DR. MEYER: No, I'm just raising this as a point of discussion, but on the other hand, I don't think it's, I don't dread re-visiting that question. I think it's high time to get to it, but for this category of compounding.

DR. COONEY: Paul.

DR. FACKLER: I just had a couple of points.

One was when, Dale, you were describing NDAs and how applicants at the point of finishing clinical trials and maybe to be marketed formulation do the BE studies and that the same stringent criteria aren't placed on them, I was going to say that there are are several examples we're aware of from the Freedom of Information summaries where, in fact, they weren't able to pass under the stringent criteria and no surprise to anyone, those are the same products that some generic companies are struggling with trying to with 100 subjects show the bioequivalence.

So, I don't disagree at all and think

0197

that while we've been able to do replicate design studies and have turned them in, what we haven't been able to do is use this reference scaling approach, which I endorse, I think it's a step in the right direction, it's making a proposed generic product compared to the existing reference product. To me it makes very common sense.

And the last point I want to make was with regard to the point estimate, even if it were only a political benefit, I think it might be worthwhile, but I thought Dr. Haidar presented data that for the very, the extreme highly variable drug, those with 60 percent CV or higher, that it actually had scientific merit.

So, I wouldn't want the committee to characterize it as just a give-away, you know, to the public or to physicians or that it does have scientific merit and maybe it's not for a large percentage of the products, but I think for some products it actually provides a bigger constraint than the --

MR. CONNOR: I'd say it a little

different, I'd look at it a little differently.

It's not scientific merit, it's an effect. I mean the -- and what he showed was that if at a certain percent when you made a certain choice as far as the constraint, that it would, its effect would overshadow it.

That doesn't mean it has scientific merit, that simply means that, you know, what we're looking at, we really need to assess when we say this, we say, oh, well, we're just going to arbitrarily, you know, plop this constraint on it. What is the true effect. I mean it does have an affect on the acceptable at some point for certain products and we have to, we just have to figure out what that is and what we're comfortable with.

DR. FACKLER: That's stated better than I did, yes.

DR. COONEY: Carol, then Ken.

DR. GLOFF: Thank you. I think Marv said very well my thoughts on this, that I think it does make sense to work to apply this scaled bioequivalence proposal.

0199

I also am somewhat uncomfortable with the 80 percent to 125 percent on the point estimate

and I don't, I can't really quantify that very well, but for the reasons that have already been discussed from political, scientific, whatever, it seems to me that we should seriously consider a bit narrower range than 80 to 125. 90 to 110, 85 to 115, I don't have exact numbers to put on that, but 80 to 125 is going out to the limits of what's acceptable for the confidence intervals now and I understand all the reasons why we need the wider range for the confidence intervals for the highly variable products, but I'm uncomfortable with the point estimates going out to those extremes.

DR. COONEY: Ken.

DR. MORRIS: Thank you. Yeah, I don't know, actually I had said Mel and said but Marv, it was actually you, so I didn't mean to pick on you, Mel, but, yeah, this is, at this point I'm a little unclear on it. I guess the idea that this is, I mean -- I mean Les was probably being a little editorial when he said it was political, but what's