DEPARTMENT OF HEALTH AND HUMAN SERVICES

FOOD AND DRUG ADMINISTRATION

CENTER FOR DRUG EVALUATION AND RESEARCH

ADVISORY COMMITTEE FOR PHARMACEUTICAL SCIENCE

CLINICAL PHARMACOLOGY SUBCOMMITTEE

Monday, November 17, 2003

8:30 a.m.

Advisors and Consultants Staff Conference Room
5630 Fishers Lane
Rockville, Maryland

PARTICIPANTS

Jurgen Venitz, M.D., Ph.D., Chair
Hilda F. Scharen, M.S., Executive Secretary

MEMBERS:

David D'Argenio, Ph.D.
Marie Davidian, Ph.D.
Hartmut Derendorf, Ph.D.
David Flockhart, M.D., Ph.D.
William J. Jusko, Ph.D.
Gregory L. Kearns, Pharm.D., Ph.D.
Howard L. McCleod, Pharm.D.
Wolfgang Sadee, Ph.D.
Lewis B. Sheiner, M.D.
Marc Swadener, Ed.D.

Efraim Shek, Ph.D., Acting Industry Representative

GUEST SPEAKER:

Peter Bonate, Ph.D.

FDA STAFF:

Hae-Young Ahn, Ph.D.
Albert Chen, Ph.D.
Joga Gobburu, Ph.D.
Peter Hinderling, M.D.
Shiew-Mei Huang, Ph.D.
Leslie Kenna, Ph.D.
Peter Lee, Ph.D.
Lawrence Lesko, Ph.D.
Stella Machado, Ph.D.
Ameeta Parekh, Ph.D.
William Rodriguez, M.D.

C O N T E N T S

P R O C E E D I N G S

Call to Order and Opening Remarks

DR. VENITZ: Good morning, everyone. Welcome to the Clinical Pharmacology Subcommittee Meeting. As you know, we have a full agenda both for today as well as for tomorrow. So, I would like for us to get started by introducing the members and the FDA staffers around the table before Ms. Scharen introduces the conflict of interest.

My name is Jurgen Venitz. I am the chair of the committee and I am an associate professor at Virginia Commonwealth University.

DR. D'ARGENIO: My name is David D'Aregnio. I am professor of biomedical engineering at the University of Southern California.

DR. FLOCKHART: My name is Dave Flockhart. I am a professor of medicine, genetics and pharmacology at Indiana University.

DR. SHEINER: I am Lewis Sheiner, clinical pharmacologist from the UCSF.

DR. SWADENER: Marc Swadener, from Boulder, Colorado.

DR. JUSKO: William Jusko, Department of

Pharmaceutical Sciences, University at Buffalo.

MS. SCHAREN: Hilda Scharen, FDA, Center for Drugs.

DR. KEARNS: Greg Kearns, clinical pharmacologist from Children's University Hospital in Kansas City, Missouri.

DR. DERENDORF: Hartmut Derendorf, Department of Pharmaceutics, University of Florida.

DR. DAVIDIAN: Marie Davidian, Department of Statistics, North Carolina State University.

DR. SHEK: Efraim Shek, Abbott Laboratories, the industrial representative.

DR. MCCLEOD: Howard McCleod, clinical pharmacologist, Washington University in St. Louis.

DR. HUANG: Shiew-Mei Huang, Deputy Director for Science, Office of Pharmacology and Biopharmaceutics, CDER.

DR. LEE: Peter Lee, Associate Director, Pharmacometrics, Office of Clinical Pharmacology and Biopharmaceutics.

DR. LESKO: Good morning. Larry Lesko, Director of the Office of Clinical Pharmacology and Biopharmaceutics.

DR. VENITZ: Thank you. Let me turn over the microphone to Ms. Hilda Scharen. She is the

executive committee secretary and she will provide us with the conflict of interest statement.

Conflict of Interest Statement

MS. SCHAREN: The following announcement addresses the issue of conflict of interest with respect to this meeting and is made part of the record to preclude even the appearance of such at this meeting. The topics of today's meeting are issues of broad applicability. Unlike issues before a committee in which a particular product is discussed, issues of broader applicability involve many industrial sponsors and academic institutions.

All special government employees have been screened for their financial interests as they may apply to the general topics at hand. Because they have reported interests in pharmaceutical companies, the Food and Drug Administration has granted general matters waivers of broad applicability to the following SGEs which permits them to participate in today's discussion: Dr. David D'Argenio, Dr. Marie Davidian, Dr. Hartmut Derendorf, Dr. David Flockhart, Dr. William Jusko, Dr. Gregory Kearns, Dr. Howard McCleod, Dr. Mary Relling, Dr. Wolfgang Sadee, Dr. Jurgen Venitz.

A copy of the waiver statements may be

obtained by submitting a written request to the agency's Freedom of Information Office, Room 12A-30 of the Parklawn Building.

Because general topics could involve so many firms and institutions, it is not prudent to recite all potential conflicts of interest but, because of the general nature of today's discussions, the potential conflicts are mitigated. We would like to note for the record that Dr. Efraim Shek is participating in today's meeting as an acting, non-voting industry representative.

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

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

DR. VENITZ: Thank you. As you can tell from the agenda, we have three main topics for discussion today, end- of-phase-2A meetings; PK/PD modeling of QTc prolongation; and pediatrics. The

person who put the agenda together, Dr. Larry Lesko, is going to introduce the topics for the meeting and the outcomes that he would like for us to achieve. Larry?

Introduction to the Meeting

DR. LESKO: Thank you, Jurgen.

[Slide]

Good morning and welcome back to another Clinical Pharmacology Subcommittee. In particular, I would like to welcome some new members, Dr. D'Argenio and Dr. Davidian. Thanks for joining us and bringing some expertise in you areas to our working subcommittee.

[Slide]

What I am going to do today is really introduce the topics for today but I am also going to review the topics that we covered in the first two meetings, and link those to today's topics to try to illustrate the continuity in issues that we have been bringing before this advisory committee.

[Slide]

So, let me start by saying that this is the third meeting of the Clinical Pharmacology Subcommittee. As you can see, it has been about 12 to 13 months since our first meeting, back in

October of 2002. We had our next meeting in April of 2003 and this represents our third meeting.

I have to say that the input of this group has had a significant impact on the progress that we have made in each of the general topic areas that I first introduced back in October of 2002, those four or five broad areas. As I go through a kind of synopsis or review of what we have done to date, you will appreciate where that input is coming into play.

[Slide]

Back in October I had indicated that a major emphasis of this committee is going to be risk, and I subdivided risk into risk assessment which we defined as a quantitative or science-based estimate of risk in a special population who is either under- or over-exposed to drug treatment. This, of course, relates to dosing adjustments that are pertinent to labeling of a drug product.

The second broad area of risk was risk management, and that was defined as taking action to reduce the risk through appropriate label language related to dosing adjustments. As you recall from our prior meetings, we talked about a two-fold approach to dosing adjustment. One is

identifying the magnitude of the risk involved with under- and over-exposure and then trying to determine an appropriate dosing adjustment to minimize that risk.

[Slide]

It isn't by accident that have covered these topics so far. In fact, approximately on August 30 of this year, the FDA's new strategic plan was released. It is on the website. One of the key parts of that strategic plan that relates to the objectives of this group--the key element of FDA's new strategic plan is efficient risk management. Secondly, to use the best biomedical science to achieve our health policy goals. Third, to make new treatments and technology less risky with greater predictability and less time from concept to bedside. I would say all the topics we will talk about come under the umbrella of the strategic plan, and in particular these elements of it.

[Slide]

So, let's talk about the scope of topics that we have covered to date and will continue to discuss: Quantitative risk analysis using exposure-response regulations; pediatric PK and

analysis of the FDA pediatric database; pharmacogenetics--we have talked about improvements in existing therapies and at the last meeting we introduced the topic of metabolism- and transport-based drug interactions.

[Slide]

Now let's take a look at each of those topics and see what we have accomplished to date and where we are going today. Well, basically, the methodologies that we presented to this committee both in October and April have basically resulted in a finalized, systematic pharmacometric methodology to apply to dose adjustments. We are and we have applied the methodology to both assessment of efficacy and safety biomarkers; in some cases clinical endpoints; and it has been helpful as a methodology or an approach to assess risk-benefit.

We are currently integrating the methodologies we talked about at our meetings into the routine NDA reviews and will in the future in early meetings with sponsors that I will talk about when we get to the end-of-phase-2A meeting.

We talked on several occasions about the utility function. This continues to be a work in

progress. The approaches that we have discussed at prior meetings have raised awareness and also the issues. I think our next step as a work in progress is to have some future further dialogue with our physicians and statisticians. There still remains an unresolved issue, namely, how to determine the appropriate utility function for relative efficacy and safety endpoints.

[Slide]

At today's meeting, thinking of the broad topic area, what we are going to do is talk about a new proposal for an end-of-phase-2A meeting between FDA and industry. What we would like to do is discuss topics at this meeting that revolve around the evaluation of exposure response and prospective dose selection.

We are going to show you some case studies of exposure-response analysis. These come from the NDA reviews but we think they are models for the type of analysis that we can conduct at the end-of-phase-2A. The idea is to look at these models and get a feeling for how the analysis at an earlier stage in drug development would have benefitted the quality of the new drug application.

[Slide]

Also related to exposure response we will be talking about a methodology for evaluating QT. This has become a major issue, as many people are aware. We will talk about points to consider for PK/PD or PK-QT study design. We will talk about the use of clinical trial simulation to optimize the study design for this evaluation, and we will show you some case studies illustrating pharmacometric considerations arising from NDA review of QT data. We are beginning to get a lot of experience with this but, looking ahead, what ought to be the important aspects of study designs for the next study that might be conducted?

We have talked about pediatric PK and the analysis of our FDA database. We basically have completed the PK, as we call it, study design template, and we have utilized it in interactions with sponsors as an alternative to determining full sample strategies in looking at the PK in pediatrics.

We have further work in progress on simulation to further optimize the number of samples, the sampling times and number of patients--basically the design of the study, and that is an ongoing work.

Last time in particular we talked about our pediatric database analyses. We are going to look at the database retrospectively. We presented some ideas on that. We got your input on it. But that has been a challenge for us, and it hasn't been a very successful initiative.

Over the last three or four months what we found is many incomplete data sets for the analysis that we want to undertake. We have non-optimal study designs because they weren't designed for the type of analysis we wanted to conduct. We haven't given up however. We have begun to look at the database more selectively, picking on drugs for case-by-case analysis and comparing pediatric and adult data for similarities and differences in exposure response. We have picked drugs were there is a more full data set and we will probably bring some of that information forward in the future. However, today we will talk more about that this afternoon.

[Slide]

So, today's meeting topic, number three, we want to revisit the clinical pharmacology principles of the pediatric decision tree with some case studies. This is a decision tree which is

always evolving as new information becomes available. But you will see in the decision tree that there is a point at which we talk about comparing similarities and exposure-response relationships between adults and pediatric patients. We haven't really adopted any methodology to compare that similarity so today we will present a method to be used in the determination of similarity of exposure-response relationships.

You are also going to hear some perspectives. There will be new perspectives. You will hear an FDA perspective from the medical side and you will hear an academic perspective from the clinical pharmacology side. Both of them will be based upon experiences with the pediatric decision tree and applying it in the development of pediatric drugs.

[Slide]

We have talked about pharmacogenetics, and the emphasis has been on the improvement in existing therapies or approved drugs. We focused for the most part on polymorphism in metabolizing enzymes that determine variability in drug exposure. We are going to stay in this area for a

while. Our emphasis in prior meetings had been on TPMT and the polymorphism that affects dose response for the thiopurines.

Since we met in April we have had additional discussions of the TPMT issue and the possible modifications of the thiopurine labels. We presented a lot of the information that we presented to this committee, including the input of the committee, to another subcommittee, which was the Pediatrics Subcommittee of the Oncology Drug Advisory Committee, in July of 2003. It was a very interesting meeting, very helpful in raising some issues that related to do we need this test; what is it going to cost patients; what is its predictive value and quality, and so on and so forth. We worked through those issues and at the end of the day this subcommittee recommended including pharmacogenetic information in a revision of the label for thiopurines.

One of the issues that was discussed in July was whether or not this test should be required before receiving drug, or the information put in the label for informational purposes to be used by the physician and the patient in certain circumstances. The recommendation of the committee

was that the test should not be required as a prerequisite for receiving the thiopurines.

[Slide]

So, at today's meeting we are going to shift the discussion of the question of the pharmacogenetics a bit. We are going to focus on what should be done in new drug development for substrates that are metabolites primarily by polymorphic enzymes. We have talked about approved drugs to some degree.

We are going to hear three expert perspectives, an academic, an industry and a clinical view. Discussion will influence recommendations that we are going to be putting in another guidance that is under development. We call it the General Pharmacogenetics Guidance. It is going to be worked on and released probably sometime in the first half of 2004. This topic will be an important part of that guidance. So, we look forward to your input on this issue.

[Slide]

Finally, we had talked about metabolism- and transport-based interactions with just an introduction to the topic at our last meeting. It was intended to be really a foundation for

subsequent discussion which will continue today. So, we wanted to bring to the committee an increased awareness of what we think are some new mechanisms of drug interactions that are becoming, to us at least, clinically important, and what do we do about them during the course of drug development.

Coincident with that, we have a revision of the Drug Interaction Guidance in progress, and many of the discussions and issues that we will discuss in front of this committee will make their way into the revision of that guidance.

[Slide]

So, what are we going to hear today? We are going to hear more specifics on this issue. We are going to be asking what should be done in the consideration of these new drug interactions of emerging importance. We will be hearing different views on the topic and we will be focusing on two metabolic sorts of drug interactions related to 2B6 and 2C8. Again, the discussion will impact future regulatory advice on these issues.

[Slide]

In summary, I have really broken down today's meeting into five separate topics where we

will be asking for your input and advice. I won't go over the specific questions right now. We will introduce those as we get to the specific topic. Again, we are looking forward to today. We are confident, as we have been in other committee meetings, that your input is going to be important to us and we are always trying to refine our thinking about these topics.

So, that is basically an introduction, a framework for today's meeting. Looking at the agenda, I am next on the agenda so maybe I will just slide into my next presentation but that, hopefully, will give you a feeling for what we are going to try to accomplish today.

Proposal for End-of-Phase-2A (EOP2A) Meetings

[Slide]

Let me pause, take a breath and say that we are moving into the first topic of quantitative analysis using exposure response. What I am introducing today really for the first time, or discussing it in a public forum, is a proposal for the end-of-phase-2A two-way meetings. This relates to analyzing exposure response, not at the NDA stage necessarily but at an earlier point in time in drug development.

I am going to walk through this proposal and then that is going to be supplemented by other presentations. Dr. Peter Lee will give an example of some of the issues that will be discussed at this meeting and possible impact, and then will present some case studies and you will have to use your imagination a bit because these are case studies that we drew from our NDA reviews but we want to sort of transpose them in time and have you think about the possibilities and the impact that this analysis might have had, had they occurred at an end-of-phase-2A meeting.

[Slide]

Let me start the story of this proposal with the current situation in new drug development. This is from the FDA strategic plan. What it shows is really an alarming change in the drug development process. There are a couple of things on here but the main point of this slide probably is that very thin white line that you see there, which is the number of NMEs filed with the agency over the last ten years or so.

You can see from a high in 1995 of about 50 NMEs, we are down to 2002 at about 20. It hasn't gotten any better so far in 2003. Recently

I read in the "Pink Sheet" that the number of INDs filed is at a record 11-year low. So, something is going on in the drug development process and many people are looking at this, including the agency, to try to figure out what is going on and how this trend might be improved.

[Slide]

So, the question comes down to what problems need solving in this current situation of drug development. We have seen estimates from Tufts and other places that it costs 800 million dollars to develop a new drug. The agency is concerned about this expense given the return on investment that we have seen in the new drug development process. This figure is high. It includes not only the actual direct cost of developing a drug but also the indirect cost of lost opportunities.

Almost 50 percent of phase 3 trials don't succeed. That is, they fail to show their target evidence of efficacy or safety issues emerge. This figure comes really from the PhRMA FDA website. Throwing figures like this around, I think you realize that this is very much drug dependent. It is higher in certain diseases like depression; it

might be lower in other diseases like antimicrobial drugs.

Only 20 percent of new drugs entering clinical testing are approved. So, four out of five don't make it for various reasons, whether it be safety, efficacy, manufacturing problems, pharmacokinetics. This, in some form or fashion, underpins the situation we have in drug development.

[Slide]

I mentioned that strategic plan that Dr. McClellan released in August of this year. There is a point in that strategic plan that focuses on new drug development and the need for greater productivity. He recommends that steps be taken to reduce the time, cost and uncertainty of developing new drugs and he identified this as an important public health policy.

[Slide]

Well, that brought us around to a specific suggestion that might fall into that goal in the strategic plan which we call the end-of-phase-2A meeting. It is kind of a general term that we have given to this proposal. It isn't intended to exclude the possibility of meetings at other points

prior to the 2A period in drug development. We could have, for example, an end-of-phase-1 meeting but, for convenience, we had to give this a name and we called it the end-of-phase-2A meeting, and I am going to tell you a little bit about it.

The hypothesis for this proposal is that meetings with sponsors early in the drug development process will focus greater attention on the analysis, in particular, of exposure-response information. We think it will improve dose selection and study design for subsequent clinical trials.

We have had prior discussion of this hypothesis with Dr. McClellan, Drs. Woodcock and Jenkins, and you can see how we have begun to sort of get the dialogue going internally at FDA with the Office of New Drug Office Directors, the Division Directors and, most recently, we presented this proposal and some case studies at a CDER all-hands guidance training in which we had several guidances on the agenda, but we talked about the April, 2003 Exposure-Response Guidance and linked that to this particular proposal. So, it has been an evolving concept and what I am presenting today is really a collective input of many of the

internal thought leaders here, at the FDA.

[Slide]

There are a couple of things driving the hypothesis that I mentioned about these early phase meetings. One of them is expressed in this quote by Dr. Temple. This was from a DIA meeting in June. He said there is more to do with regard to dose choice from exposure-response studies and there is much to be gained from better use of biomarkers and more efficient study designs for phase 3 trials.

It is hard to argue with that but the question was where do we have the dialogue on this? Where do we have an interaction with the company? The end-of-phase-2A meetings aren't the place to have this because drug development dose selection phase 3 trials are pretty much set at that point and there is not a lot of time to discuss either biomarkers or dose-response data. So, there was a missing gap.

[Slide]

We have three guidances that drive this hypothesis about early meetings. The most recent one was from April of 2003, exposure-response relationships. We talked a lot about regulatory

applications in study design and data analysis. But we also had behind that two previous guidances on clinical evidence of effectiveness and dose-response information. So, taken together, these are the principles--probably as good as they can get right now I think--of best practices in exposure response. Like a lot of guidances, however, they have to be interpreted and, for interpreting those, having meetings with industry is a good place to do it.

[Slide]

So, as a philosophical point, FDA is interested in good dose-response analyses. There are some data driving this hypothesis as well. We conducted an informal review of exposure-response data in over 100 NDAs submitted between '95 and 2001. The purpose of this review was to try to form a foundation for what this meeting is going to accomplish, where we identified missing data related to the quality of submissions and approval rates. We were looking for the extensiveness of dose-response data, dose selection process, how many studies were conducted, and so on.

We also did a prospective evaluation of over ten NDAs submitted in 2002 and 2003. What we

tried to do here was evaluate the impact of the review, in other words, what happened at the NDA stage with the analysis of exposure-response information. Were problems uncovered? Were doses considered inappropriate? We asked the question of whether or not this type of review--the review at the NDA stage--if it had been carried out earlier in the IND period in conjunction with the sponsor, would it have saved time; would it have saved costs; would it have saved review cycles when it came to the NDA?

[Slide]

Some of the results of exposure-response reanalysis in that collection or cohort of ten studies showed us the following: That we could avoid reanalysis of exposure-response data, potential requests from other disciplines to conduct additional clinical trials. That is, we reanalyzed the exposure-response data. We integrated data across several studies and avoided the need for additional clinical trials.

We found that this reanalysis resulted in the approval of lower doses or different dosage regimens than that proposed by the sponsor for a variety of reasons including safety. We identified

missing data on specific doses or in special populations, including drug-drug interactions that impacted review time. So, these are all significant findings of what a reanalysis at the NDA stage found. Again, can we move this forward into the end-of-phase-2A and achieve the same objective but earlier and result in a higher quality application?

[Slide]

There is an additional goal which we struggled with in terms of resources here at the FDA, and that is efficient and effective use of our resources. We feel that interactions with sponsors early in the drug development process provide not only an opportunity to improve things but to provide advice on development of information of exposure response and other clinical pharmacology issues, rather than waiting until the NDA is in and identifying problems--drug interactions that may not have been conducted; special populations that may have been ignored. Yes, we can deal with those but that involves labeling and very careful labeling. But having these discussions early about the overall clinical pharmacology development plan, exposure-response relationships, dose selection and

dose choices we think is an efficient and effective way to develop drugs.

[Slide]

Now, let me talk a little bit about the timing of the meeting so we are clear on what we are talking about here. What this slide shows basically is the general scheme of things as it currently exists. Typically, sponsors will request--these are all voluntary requests, by the way and they are not required meetings--pre-IND meetings.

The next junction at which FDA and industry has a formal get-together is the end of phase 2. Sometimes there is a pre-NDA meeting. Sometimes there are labeling discussions and then an action letter. So, you can see the wide gap that occurs here between the pre-IND and the end-of-phase-2A.

What we are proposing is a meeting that occurs in between these. We call it the end-of-phase-2A. As I mentioned at the beginning, I don't want to exclude the possibility that we can have a meeting at the end of phase 1. This will be very drug specific, what we know at the time. We are trying to focus on the information that is

available in this time frame of drug development. If you meet too early you have an incomplete data set and the meeting becomes filled with a lot of uncertainty. If you meet too late in this scheme the drug development plans are already cast in stone and it is hard to change them. So, what we are trying to do is find a balance in this drug development scheme, going from preclinical to submission, for where is the optimal time to have the interactions with sponsors for the reasons that I described,

[Slide]

The rationale for the meeting time, end-of-phase-2A, is that we think that it is at this point that there is basically complete information on preclinical pharmacology and exposure response complete in the sense of having healthy volunteer studies, drug dose tolerance studies, things like that. So, we have the safety data in healthy volunteers. We have some efficacy data depending on the drug at that point in time. We have some initial efficacy or proof of concept data from the early phase-2A studies, and we have safety data in patients, albeit a relatively small database.

This is generally, although not always, prior to the so-called conduct of registration of label studies, that is, studies that a sponsor may conduct on special populations, drug interactions, food studies, perhaps some formulation studies. So, taken together, this information represents a fairly rich database for an early meeting with sponsors and an opportunity to analyze exposure response in particular.

What we would also like to add to this, as we talked about in this meeting, is emerging issues. There is a lot of uncertainty about integrating things like pharmacogenetics in the drug development, but we think this would be an ideal place to talk about things like this as well as other topics, such as the use of trial design simulation, and so on. So, this is the rationale for it as to why we picked the end-of-phase-2A.

[Slide]

We also think this is an opportunity to advance the idea that mechanistic and quantitative methods of analysis of exposure response would be beneficial. We envision that this meeting would involve significant modeling and simulation to analyze and integrate exposure-response data across

studies and explore dose choices for both 2B and phase 3 studies.

We think this will be a point at which we can discuss the design of studies using computer-assisted clinical trial simulation, and these are relatively new technologies that we think should be applied in this context. This is a good time for us to talk with the sponsor about the design of PK studies to efficiently identify covariates affecting exposure response in later clinical studies, things like number of patients, sample times, things of that sort.

Also, if you think about all the special populations and drug interaction studies that are conducted, those have to be interpreted as to whether or not a dose adjustment is needed. So, we think this would be a good time to begin to talk about therapeutic equivalence boundaries that would be based upon exposure response or help interpret the outcomes of these special population drug interaction studies as to whether a dose adjustment is appropriate or whether it isn't, and this will help I think near the end of the drug development process with the labeling discussions that we have.

[Slide]

Somebody asked about what is the difference between this meeting and the traditional meeting that we have with sponsors called the end-of-phase-2. Well, I think there are some major differences. For one thing, by the end of phase 2 the sponsor has pretty much made a final decision on the choice of doses or dose ranges for phase 3. Final formulations are developed and it is difficult at that point to change things without affecting significantly the time frame for the drug development program.

The end-of-phase-2 meeting is a formal meeting, very formal. The goal of that meeting is to discuss study design for phase 3; clinical endpoints; heavy emphasis on statistics; and basically leading up to what is the evidence one needs for approval in terms of the adequate and well-controlled trials. Also at the end of phase 2, for most part many, if not all the special populations and drug interaction studies are complete. So, the opportunity to influence the key parts of drug development pretty much have gone by the board at this point.

The end-of-phase-2A meeting, in contrast, will focus on some decision points in the

development program. The meeting will be a bit informal as well. I don't mean informal from the standpoint that we don't take minutes or we don't keep track of the meeting, but I mean informal in the sense that there is a larger degree of uncertainty at the end of phase 2A than at the end of phase 2 because of the lesser amount of information, and we recognize that.

[Slide]

One of the questions we have and would appreciate some comments on is we have limited resources to conduct these meetings. We are going to begin them fairly soon. One of the discussions that we had internally, and that whole list of discussions I mentioned to you, is if we have limited resources where would the impact of these types of meetings be greatest. Would it be a first in class drug or one where there is significant therapeutic advancement where the importance of getting doses is particularly emphatic? Or, in contrast, is it one where we understand the pathophysiology of the disease and the pharmacology so that we can call upon a lot of the experience to enhance the interactions with the sponsor?

We think it would depend on the

completeness of the background package. I will talk a little bit about that. There is another debate about whether this would be for an experienced sponsor or one with less experience in terms of the value of these interactions. So, this is something we are going to have to sort out. We have in our mind a target for these types of meetings to probably have no more than two per month with our current resources and as a way of introducing this as a pilot project.

[Slide]

Let me tell you about the plan for this meeting. We are going to draft a guidance for industry. You have in the package that was sent to you today a concept paper on this meeting which goes into a lot more detail.

The guidance will talk about background objectives, examples of topics, the usual process things for setting up the meeting. These meetings, like many meetings with sponsors, are going to be voluntary, relatively informal and, most important, interdisciplinary. This is not a clinical pharmacology meeting; it is a meeting that will involve resources from ourselves in clin. pharm., but also the medical and biostatisticians in our

review divisions. We would like to evaluate the impact of this meeting after some years of experience. We are trying to think in maybe two or three years we need to look at some metrics for how the impact might be assessed.

[Slide]

So, in summary in introducing this new proposal for an end-of-phase-2A meeting, we think the meeting will serve to decrease uncertainty in further drug development, for example in phase 3. Uncertainty, we think, leads to some of the problems that I mentioned in the beginning in terms of the drug development process today.

We think there is opportunity to do more quantitative analysis of exposure-response data to define better the dose ranging for subsequent clinical trials. We think it is a good time to identify missing information or discuss necessary information prior to submission of the NDA to reduce issues that come up at that point in the process. We think at the end of the day, after some years of experience, we will find this improves the informational quality of NDAs and minimizes the delays in NDA review, for example second and third review cycles that may be related

to dose selection or issues of efficacy and safety.

[Slide]

So, what is it we are looking for today? You are going to hear a story, as I said, about some of the issues we see coming up at this meeting and then some case studies. What we would like is some comment on the goals of this meeting. Do you think they are appropriate? As importantly, what do you see as some obstacles to achieving these goals?

You are going to see some analytic methods employed in these case studies using exposure-response examples from our NDA review. Think about these methodologies, how can they be improved; what should we be thinking about in terms of getting even more from the analyses?

Do you have any thoughts on metrics? What are the metrics that would be used to measure the impact or success of this initiative? That would be important as to whether or not we continue with it beyond the pilot period of a couple of years.

So, that is the end-of-phase-2A meeting. I will turn it back to the chair but we are going to continue discussing this and drill down into some more detail, but if there are any questions I

can answer about the overall concept.

DR. VENITZ: Any comments or questions for Dr. Lesko before we proceed?

[No response]

DR. LESKO: I am going to turn it over to Peter who will continue the discussion and talk about some of the issues that we think will come up.

Issues Proposed to be Discussed at EOP2A and their Impact

DR. LEE: Thank you, Larry.

[Slide]

I think later today we are going to hear several examples that will illustrate a potential benefit of discussing exposure response at an early clinical development stage, specifically at the end-of-phase-2A meetings. But what I would like to do now is go over some of the potential topics that we think will be useful to discuss with the sponsor early on.

[Slide]

As Larry has mentioned, we have informally looked at ten NDAs where the exposure-response information has made significant impact on regulatory decisions. In some of the NDAs the

exposure response was used to approve a lower dose or a different dose than was proposed initially by the sponsor. In some cases the exposure response was used to avoid any additional clinical studies, especially efficacy and safety studies in the submissions. Finally, you saw that exposure-response information has been used to identify the desired missing doses and also special population studies.

[Slide]

So, we thought that if this type of analysis, exposure-response analysis, were done early on during drug development we might definitely save review time and besides it may improve the efficiency of the drug development process. So, one of the general goals for the end-of-phase-2A meeting that we propose is to discuss exposure-response issues. We hope that by this type of discussion we can make impact on the decision-making about the design and analysis or exposure-response study early in the drug development process.

Also, we think that we could discuss the strategy in dose choices and special population studies. We also hope to be able to analyze by

quantitative analysis, for example, modeling simulation and clinical trial simulation so that we can integrate relevant preclinical and clinical exposure-response data and, hopefully, close the gap between what is known at the end-of-phase-2A meeting and what will be applied in designing the phase 2B and phase 3 studies.

[Slide]

So, here are some of the discussion points. A discussion point that we thought would be useful at an end-of-phase-2A meeting--and what I will do in the next few slides is go over each of these discussion points one at a time and also talk about the potential impact of these discussions.

[Slide]

The first topic for the end-of-phase-2A could be the dose range strategy. In the examples that you will be hearing today, in most of those cases a suboptimal dose was selected in the original NDA which would lead to either lack of efficacy of the drug in the phase 3 studies or adverse events. Therefore, I think it would be useful in an end-of-phase-2A meeting to discuss the rationale for dose selections in a planned study, and this can range from the first dose to an

efficacy and safety study. Definitely, this will depend on the preclinical and clearance evidence for the effectiveness and safety of the drugs.

We could also discuss the drug development strategy which could be a sequence of studies that lead to the doses actually in the final efficacy and safety studies. We could also talk about the design of individual exposure-response studies.

[Slide]

The second topic we propose to discuss at an end-of-phase-2A meeting is exposure response to support efficacy and safety. In the Exposure-Response Guidance that was just recently published early this year, we discuss the utility of exposure-response information to support efficacy and safety. Of course, this could be on a case-by-case basis so it would be useful for the sponsor to come in to discuss early on the quantity and quality of exposure-response data that might be used to support efficacy and safety. We will also talk about the potential design of an exposure-response study that may lead to supporting information.

Another useful topic to talk about is the modeling and simulation methodology that may be

used to analyze the exposure-response study and to generate supporting information.

[Slide]

Another topic to talk about at the end-of-phase-2A meeting would be dose adjustment in special populations. Quite often during the NDA review there are quite intensive negotiations regarding labeling language, which usually leads to either a delay of review, NDA review, or in some cases leads to a phase 4 commitment. So, we thought it would be useful, again, to talk about the dose adjustment decision tree early on during the drug development process; and also talk about a required clinical pharmacology study that would support dose adjustment with special populations; also the analysis of exposure response and perhaps also talk about an alternative population PK study design that may replace the traditional intensive clinical pharmacology study supporting special populations and drug-drug interactions.

[Slide]

The next topic that we would talk about is the design of efficacy and safety studies. The objective here is to focus on the likelihood of getting the right doses, and also explore some of

the "what if" scenarios and to look at the study robustness and the study power.

We can look at a variety of study design factors, such as dose range selections, inclusion and exclusion criteria, the inclusion of special populations and PK design, sampling scheme, and so on and so forth.

We could also talk about an alternative study design methodology, such as an adaptive design, a different titration scheme or even a new study design such as a concentration-control study design. Definitely, because of the complexity of the issue, clinical trial simulation could be used to design the efficacy and safety trials.

[Slide]

Another topic we could talk about at an end-of-phase-2A meeting is the population PK/PD study design. At this time, only about 50 percent of the full NDAs contain population PK analysis, however, quite frequently the objective of this analysis was not very clear and a lot of times the population PK studies were not designed prospectively, which will lead to the result becoming non-conclusive. Therefore, it would be useful, again, to discuss the objective of the

population PK study early on and prospectively design a study so that the information can be useful to support labeling regarding special populations as well as drug-drug interactions.

[Slide]

Another important topic that we thought would be useful to discuss is the QT study design. QT has become a very important topic and has attracted a lot of attention recently because of several drugs being withdrawn from the market due to the QT prolongation property. As you know, the issue here is the large variability of circadian variation of QT.

There are other issues such as the baseline correction methods, and so on and so forth. Therefore, it would be helpful, again, to discuss the study design issue early on, perhaps using clinical trial simulation to optimize study design as well. We will be giving several examples later on today to illustrate how the clinical trial simulation can be used to design the studies.

[Slide]

So, today we are going to hear many examples on topic 1. This morning we will be hearing three different cases where exposure

response was used to support dose selection strategy or to support efficacy and safety. Later this afternoon we will be hearing two presentations regarding the use of clinical trial simulation to support PK-QT study design. With that, I will turn it back to Jurgen.

DR. VENITZ: Again, any comments or questions before we proceed to the case studies?

DR. SHEK: I have one.

DR. VENITZ: Go ahead.

DR. SHEK: It is my personal belief and I believe most of the industry will welcome any productive and effective interaction with the agency during the drug development process. But specifically, those ten NDAs that you were looking at in 2002 and 2003, how many of those were successful the first time and went through, you know, the first review, and how many of those failed completely?

DR. LEE: Yes, specifically, we looked at the ten NDAs that either received not approvable or approvable. So, all those ten NDAs did not get approved status in the first round.

DR. SHEK: None of them?

DR. LEE: No.

DR. VENITZ: Larry?

DR. LESKO: I was just going to add on to the answer Peter gave and say that one of the issues that has been talked about is the number of review cycles on NDAs. I believe some information was released by the agency that indicated that the reasons for multiple review cycles are most of the time safety issues. I don't remember the exact percent. The second reason is issues having to do with efficacy. The third reason is CMC issues. It breaks down by percentage in that rank order, although, as I say, I can't remember which is which.

The question we had was were those multiple review cycles related to issues revolving around dose response, and I don't believe we answered that question because it was too complex a question to link to the one issue of dose response. But it is probably multiple issues--risk-benefit considerations, but I think the dose response issues were part of the answer, not the complete answer for those multiple review cycles. But that is one of the ideas of what we would like to actually improve, and maybe it is one of the metrics that we would like to look at in the next

couple of years, in those cases where we have these meetings, has that resulted in approval on the first cycle or reduction in delays to the second and third cycles.

DR. VENITZ: Any other comments?

[No response]

Then, let me introduce Dr. Parekh. Ameeta is going to give us the first case that illustrates the potential use of end-of-phase-2A meetings. Ameeta?

Case Studies

DR. PAREKH: Good morning, everyone. Before I start, I was noting some of the words that Larry had in his presentation. He was talking about moving on with the new technologies. Just on a lighter note, I was working on my slides over the weekend, trying to do some spell checks. It was interesting, I had some British spellings and some American spellings, especially on a word like "learnt" versus "learned." So, I was updating my slides and in my panic I brought in this with the updated slides; this with the updated slides; and just as a security measure I sent myself an e-mail with an attachment. Well, I also just took this because my kids said, "mom, you never know." I

came in today. The network wasn't working so I didn't have my e-mail. I asked John to use this to update the computer. It didn't accept this. For some reason it didn't read this.

[Laughter]

So, you never know what might work. So, I had four and one of them worked, and it was the good old well-tested in the clinical trials technology that did work.

[Slide]

Larry has already laid out the CDER plan for the end-of-phase-2A meetings, the focus being on a more rational approach to utilizing the exposure-response data early on during the drug development, mainly for dose selection, dose optimization and dosage adjustment. As Larry also mentioned, it is an interdisciplinary kind of role that these aspects play. It is not just solely clinical pharmacology and us. So, it is the clinical division and at times even the chemistry reviewers and pharm. tox. as well.

What we are going to do is we are going to share some case studies with you and, as Larry mentioned, these case studies are not really derived from the end-of-phase-2A meetings. These

are derived from the NDA examples, for instance, but the principles and the concepts that will be discussed in these cases do lend themselves very appropriately to the general framework of the end-of-phase-2A.

[Slide]

Larry talked about the different milestones during drug development, the different time frames when we meet with the sponsors to discuss the drug development, with some companies more, with some a little less. It depends on the companies. So, I am not going to really emphasize the milestones, the different stages of drug development too much.

I do want to dwell more on the different stages of the review cycle, the clinical, pharmacology and biopharmaceutics role in the review process, and what the reviewers go through and what questions they ask while they are reviewing the NDA, with special attention to the exposure-response relationships and, of course, exemplified with some case studies and the bottom line upshot of all this, the lessons learned.

[Slide]

Again, I am not going to focus on all the

different stages of drug development but certainly I would like to draw your attention to this region, here, which is basically the NDA submission. The NDA comes in; we look at the NDA, the volumes, and we look for the primary components in order to file the NDA. If those primary components are in the packages that are submitted, the NDA gets filed. Interestingly, at that point how well exposure response is evaluated is not one of the components. So, there are certain things that we look for that makes the NDA reviewable. We file the NDA and then it goes through the review cycle.

[Slide]

Basically, what I am going to focus on is in this circle, here, which is that the NDA gets filed. It is the review and the focus is what goes into the label if it does get approved. Of course, the bottom line is the action letter that goes back to the sponsor.

[Slide]

So, I would like to zoom in on this circle, here, the stages of clinical pharmacology and biopharmaceutics review. I classified the three components into three broad components, the NDA review, the label and the action letter.

[Slide]

Let's zoom in on the NDA review. What are the different stages of the clinical pharmacology and biopharmaceutics reviewer in the trenches? What do they go through? I would acknowledge Dr. Sheiner and one of his earlier papers, the question-based approach. We do take the question-based approach to reviewing an NDA.

Basically, when a reviewer starts the review of an NDA we do ask a series of very logical questions and each one is inter-linked with the other, the bottom line being the big umbrella that Larry talked about earlier, risk assessment, risk management, dosage adjustment.

How was the dose determined? Again, it is interdisciplinary; it is not just us. We do work with the clinical divisions on this. When you think of how the dose was determined, an obvious question that comes up is what is the exposure-response relationship? When you think of exposure-response relationship, you think in terms of both safety and efficacy. What is the most useful thing for determining or getting a good feel for the exposure-response relationship? It is choosing the right dose, the right starting dose in

relation to where the profile is in terms of its efficacy as well as its safety. So, you can't be just blind-sided by let's get the biggest dose on the market so it beats placebo.

There is another downside to it, and that is what are you going to lose; what are you going to give up should there be several doses so that the patients have the option of titrating up or down? Or, another aspect, which is really primarily clinical pharmacology, is extrinsic/intrinsic factors. How will the exposure change? Will the patients have an option for a lower dose given that, for example, they would be taking the drug with, say, ketoconazole and it is a 3A4 substrate? So, things such as that is where we come in.

[Slide]

Once you have a good feel for the exposure-response relationship, both in terms of safety as well as efficacy, the obvious questions asked are what are the effects of extrinsic factors and what are the effects of intrinsic factors? When we consider these things, it is interesting how to us, I guess because of the number of NDAs we see, things just are so obvious or maybe the

hindsight is 20/20. You would think a 3A4 substrate is an important inhibitor study. There are times when the right studies are not done, and that is an example where we can help during the early development so that time is not lost towards the end. Is the dose of the important inhibitor done right, or will that become one of the approvable issues? So, things such as those could be useful and discussed during the end-of-phase-2A meeting. Of course, if you have the option for dose adjustments, is the pharmacokinetic dose proportional? That is where we come in as well.

Peter mentioned earlier cardiac repolarization. The QT effects have taken on a big role in current drug development. These are also safety issues but we also look at the exposure response with the effects on the QT prolongation, and there is going to be an extensive discussion of that later on.

Again, designing the QT studies--we have a concept paper out. It talks about phase 1 studies but even in those are phase 1 studies there are certain aspects that you need to understand very well about a drug. For example, the concept paper talks about super-therapeutic doses. What are the

relevant super-therapeutic doses? You need to know a little bit more about the drug. Again, that is where we can help out. For example, is a positive control used? Is a placebo used? Again, there is going to be more discussion on that later.

Some biopharmaceutics aspects become important towards the end of the review cycle as well. Are appropriate bioequivalence studies done? Minor as it may seem, some QT aspects can become, you know, a little bit of a discussion issue towards the end, as well as the stability out there, things such as that.

[Slide]

Once we get all this information and we understand all this, the relevant information from all these studies and our understanding goes into the label. We try and make all this information in the label in a decipherable form as much as possible. Basically, what it translates to is what doses should be approved? What is the optimal dosing regimen? What is the right patient population? What are the extrinsic and intrinsic variables for which dosage adjustment might be needed? Again, it is interdisciplinary and it is not just clinical pharmacology and

biopharmaceutics. We do interact with the other disciplines extensively to make these decisions at the end.

Again, if intrinsic/extrinsic factors result in exposure changes, how critical are these? Should it go into precautions, warnings or even contraindications for that matter? Again, another aspect that has become quite important lately is the QT prolongation, the cardiac electrophysiology of the drug.

The bottom line for all this is the action letter and it could be approval. If everything falls in place you could write a very good label. It could be approval with some phase 4 if the phase 4 could add value to the label, and the examples that Peter mentioned, approvable or non-approval--that could be very common as well, depending on what is missing from the whole picture.

[Slide]

I will discuss a couple of case studies. Basically they make slight subtle different points, optimizing dose and dosing regimen, case A. Case B, selection and dose adjustment.

[Slide]

Starting with drug A, it is an injection formulation. Interestingly, the dose finding was done by the sponsor. A very nice dose-finding study was conducted. However, it was done on a short-term period, and that was fine. It was done on, say, X days. The efficacy was evaluated over 3X days, and this may be very common. You don't do three-year dose-finding studies. You do some short-term dose-finding studies and then you go into the clinical trial.

Interestingly in this case, the dose finding that was done over an X period of time was done with a dosing regimen that was more frequent than the 3X time. You would think, you know, it would be okay depending on where you are on the exposure response with respect to efficacy. If you are way up, you know, a little change in concentration shouldn't make a difference. However, if you are not, then you need to very carefully evaluate what doses you are studying in this whole long-term period, and the observation was loss of efficacy over time.

[Slide]

We did have some exposure-response data. As this profile shows for drug A, the

concentrations that would provide, say, 90 percent of the patients with efficacy was about 10. Interestingly, 10 was about the concentration that was targeted and it was studied in the phase 2 dose-finding study.

So, if you look at the profile here and if the doses were here you would think that if the frequency of the dosing is not the same as the dose-finding study then, you know, even if it drops from here to here it wouldn't really lose too much. However, you are at the threshold of efficacy here. If you are targeting 90 percent of the patients with efficacy, you don't really have much room to slide. Basically, that is what was observed.

[Slide]

Here are a little more specifics on drug A. The dosing was on day 1, day 15, day 29 and then monthly thereafter. So, if the dose finding was done in this region, here, you would think that efficacy was achieved mainly because of the more frequent administration here. But as time progressed there was loss of efficacy and, as you can see, there were patients that were going below the 10 targeted exposure. The reason you would think again hindsight is 20/20, you would think

they could have done some simulations. But, you know, it is easier said than done I guess at the end of the NDA cycle.

[Slide]

Here is another example where we think we could have maybe helped out with some simulations and some decision-making. When we looked closer at the concentration distribution and if you just focus on the four boxes, right here is the concentration distribution at day 29. This is month 2. This is month 4 and this is month 6. If you look at this X axis with 10 as the target concentration, you can see that all these patients at month 1 were above those concentrations so obviously efficacy was achieved and 90 percent or more of the patients did achieve efficacy. However, as time progressed there were several patients who lost efficacy.

[Slide]

Simulations suggested higher or more frequent doses could achieve and maintain therapeutic drug concentrations based on the exposure-response relationships. Of course, you do want to factor in the side effects. So, of course, factoring that in, higher doses or more frequent

doses could have helped. So, need for appropriate dose and dosing regimen selection could be where we could have contributed early on in the drug development.

[Slide]

Moving on to drug B, I do want to add that drug B is not a particular drug. What I have done here is I have taken several issues from more than one drug. I have combined it into this supposed drug B just to make the point. So, it is a new drug. The critical issues related to exposure response, in this case dose selection and dose adjustment due to intrinsic and extrinsic factors.

[Slide]

This is the dose-response relationship that is available to us based on phase 2/phase 3 data. When you look at this profile you would be tempted to go over the highest possible dose, which is maybe 200. So, the temptation to pursue the highest possible dose has to be balanced off with what you are giving up. If you are going from 100 to 200 you are not really gaining that much in terms of efficacy, but what are you losing? Even if you go down to 50, going from 50 to 100 you are gaining a little bit but at what cost? I would

even go down further. How about this? This may be better than placebo. It is not as good as 50. But, you know, some patients may benefit from that and maybe we need to consider some extrinsic/intrinsic factors where even these strengths here could be approvable.

So, looking at all this in and of itself is not sufficient. Again, as I mentioned earlier, in choosing the doses it is very useful to know the shape. Here you have the shape of the efficacy curve, but you also need to know the location of this curve in relation to the adverse events.

Here is the adverse event profile for different adverse events, several studies, phase 2/phase 3. As you can see, for up to 50 you don't see much difference in terms of adverse events compared to placebo, but as you go higher you do see an increase in adverse events. How do you balance this off? Thinking in terms of the utility function--we don't have that yet but thinking in terms of the utility function, you wonder how severe are these adverse events. Would it be reasonable even to approve this dose? Again, it depends on the utility function or the severity in terms of risk-benefit analysis.

So, again, going from 100 to 200 you do need to factor all this in. It may be prudent to cover lower doses just so that the patients have options. So, there were dose-related adverse events. What if, in this day and age, it is dose-related QT effects? Again, bringing in the utility function, how critical is this 200 dose? What if it is dose-related QT events? Should it even be approved, the 200 mg dose? So, all these aspects were considered in drug B.

At this point, when you have a good feel for the exposure response for efficacy as well as safety, the next obvious question that we asked is what is the effect of extrinsic/intrinsic factors? If there are changes in exposures, big changes in exposures, don't you think there should be more than one strength available to the patients so that patients can start at, say, 25 mg, right here, and have the option of taking it with, say, ketoconazole if it is a 3A4 substrate so that the exposure does give you some room for safety as well as efficacy?

[Slide]

Then you target an exposure profile. That is the exposure profile; you want to keep a balance

of safety and efficacy. You see what happens with intrinsic factors. In this case, say for hepatic impaired patients, the exposure went up. You can have a lower dose in these hepatic patients.

It could be something worse in an intrinsic scenario and in that case you may want to consider a much lower dose, and is that strength available with stability data? I mean, should that come at the end or should that be thought through early on because you don't want a small thing like that to be a show stopper. In this case, for instance, you want to consider not maybe just lowering of a dose but even the dosing interval. So, things such as this did lead to dose adjustment for drug B.

[Slide]

In conclusion for drug B, exposure-response analysis suggested that more than one dose should be considered for optimal balance between safety and efficacy. Based on the changes in exposure due to these factors, dosage adjustment was recommended in the label. And, considering these outcomes early in drug development can help plan appropriate clin. pharm. studies, say for example, the drug-drug interaction studies. We

often go back and say, well, you have done the study with 200 mg ketoconazole; you should do it with 40 mg ketoconazole.

[Slide]

So, things such as that are minor but they can become important issues with respect to safety and labeling at the end. Based on experience for changes due to extrinsic and intrinsic factors, sponsors may consider additional strengths for marketing and have appropriate work done for these lower strengths.

[Slide]

The concluding slide is basically that exposure-response information is at the heart of determination of the optimal drug with respect to good safety and efficacy, and the cases have exemplified that. In conclusion, it is important that carefully and timely consideration be given to these assessments, and that emphasis be laid on exposure-response analysis for both safety and efficacy and also extrinsic/intrinsic factors. Thanks.

DR. VENITZ: Thank you, Ameeta. Any specific questions?

DR. JUSKO: Dr. Parekh, I wasn't clear,

for drug A were you showing us the results of a phase 2A study? It seemed like there was a large number of patients. Are you saying that the manufacturer did not recognize this drop in concentrations and did not deal with it appropriately?

DR. PAREKH: Again going back, we don't have any cases with end-of-phase-2A type of setting. What I presented in those two cases is based on phase 2B and phase 3 data where there was available to us some exposure-response information. Based on that, if at least phase 2 data could be evaluated early on maybe a better assessment could be made on dose selection, dose titration or dosing regimens for example. But the two examples that I gave are definitely not phase 2A because we haven't really implemented phase 2A yet. But certainly end-of-phase-2B is where we can get some of the data. So, there were good dose-finding studies done but the exposure response was not evaluated as well as we think so it could have helped the sponsor as well as us.

DR. LESKO: Bill, I think that point is actually relevant because one of the things we are trying to look at from the NDA is to sort of

sequentially go back and take information from what we know and see if our analysis of earlier data would have led to different conclusions than the sponsor actually did. Because one of the realities of end-of-phase-2A is, yes, you are going to have relatively small studies compared to phase 3 and whether that information, depending on a case-by-case, is going to be enough to do effective analyses of dose response to go forward with or not depends.

We won't always have the extent of information that Ameeta presented from that particular NDA, but our experience in going back and saying let's not look at the phase 3 data; let's look at what we knew--you know, try to mirror a real example, still seems to show that we would come up with some valuable analyses and maybe different recommendations. But that is something we have to learn and get through.

DR. VENITZ: Any further questions?

[No response]

Thanks again, Ameeta. Our next speaker is Hae-Young Ahn. She is going to talk about another example involving a drug that was recently reviewed.

DR. AHN: Hi. This is Hae-Young Ahn.

[Slide]

I will discuss two studies with rosuvastatin. Since rosuvastatin is approved I don't have to blind the drug name. At this moment I would like to discuss the role of exposure-response evaluation in drug development and regulatory decisions using rosuvastatin.

[Slide]

The background of rosuvastatin--it is a synthetic lipid-lowering agent. Its mechanism of action is competitive inhibition of HMG-CoA reductase. Its pharmacokinetics is as follows: Its absolute bioavailability is about 20 percent in the Caucasian population, and food decreases Cmax about 20 percent, however, it does not alter the exposure of AUC. It is not metabolized extensively. However, 10 percent of a radio-labeled dose is recovered as a metabolite. A major metabolite is formed by 2C9. Rosuvastatin is primarily excreted in the feces and the elimination half-life is 19 hours.

[Slide]

Japanese and Chinese ancestry have two-fold AUC that of the Caucasian population;

patients with severe renal impairment have three-fold higher compared to healthy volunteers. And, there were significant drug-drug interactions. Cyclosporine increased the levels of rosuvastatin about seven-fold. Gemfibrozil increased exposure about two-fold.

[Slide]

The original NDA was submitted in June, 2001. The sponsor proposed doses of 10 mg, 20 mg, 40 mg and 80 mg. In May, 2002 an approvable letter was issued to the company by the agency. In the letter it was stated that 80 mg was not approvable because of little added benefit over the 40 mg. This small added benefit does not outweigh the risk of myopathy and renal concerns. The letter stated that 10 mg, 20 mg and 40 mg are approvable.

Before the NDA was approved the following issues should be addressed by the sponsor: The first was additional safety data on 20 mg and 40 mg because the number of patients in clinical trials were not adequate to provide assurance of the safety of either 20 mg or 40 mg. And, the company had to address the renal issues because safety monitoring in clinical trials was not adequate to determine the nature of the renal toxicity.

Finally, the agency believed the clinical data was not adequate to assess optimal dosing. After the sponsor addressed the above issues adequately, in August of 2003 the approval letter was issued to the company. At this time we approved 5 to 40 mg.

[Slide]

How could exposure response or PK/PD modeling guide optimal dosing for rosuvastatin?

[Slide]

This slide shows the LDL cholesterol percent change from baseline. This data is from two clinical trials. This slide clearly shows that lipid lowering is dose related from 1 mg to 80 mg even though the company proposed 10 mg to 80 mg.

[Slide]

This slide clearly shows lower than 10 mg and 1 mg to 5 mg, can have significant LDL lowering effect. For example, 1 mg has 33 percent LDL reduction; 5 mg has 43 percent LDL reduction. The titration from 40 mg to 80 mg does not provide any additional significant benefit. However, the 80 mg dose provides a mean of 2-4 percent of LDL reduction compared to 40 mg. However, the range of responses was very similar to that of 40 mg. So, at this moment I would like to draw your attention

to the lower dose than 10 mg.

[Slide]

The Office of Clinical Pharmacology and Biopharmaceutics did PK/PD modeling. The first column is dose. The second and third column represent observed percent LDL reduction. The fourth column is the mean predicted percent in the reduction at week 6. The last column represents the minimum percent LDL reduction in 85 percent of the populations.

Let's look at the fourth column. Our prediction shows that 1 mg has a mean of 38 percent of LDL reduction; 5 mg can provide 44 percent of LDL reduction; 10 mg can provide 50 percent of LDL reduction.

Let's look at the last column, a 1 mg dose can provide a minimum 26 percent of LDL reduction in 85 percent of the in patients; 5 mg can provide a minimum of 32 percent of LDL reduction in 85 percent of the population.

[Slide]

Since there are so many modeling people, I would like to satisfy you modeling experts. This is LDL percent changes from 1 mg up to 80 mg. The efficacy endpoint was after 6 weeks. This is our

predictive simulated data and these are observed data from two clinical trials. A mean observed in clinical trial data overlaps with the predicted value. So, we can say our model was validated.

[Slide]

At this moment I would like to switch gears from efficacy to safety. This slide shows the incidence of CK elevations in myopathy seen in steady treatment. This summarizes the data from the clinical trial development from Baycol, rosuvastatin and all currently marketed statins. For rosuvastatin, a 40 mg dose lowers the incidence of CK elevation and myopathy within the range of all currently marketed approved statins. However, there is a clear break at 80 mg. The two highest does of Baycol, 0.4 mg and 0.8 mg and rosuvastatin 80 mg have similar frequency of CK elevations of 10-fold of the upper limit or normal and myopathy as you can compare these two values.

[Slide]

This slide shows the percent of patients with proteinuria. Patients include all controlled and uncontrolled clinical trials at any visit. The numbers in parentheses are total number of patients in each group. There is a clear percent of

patients with proteinuria that is kind of dose related. There is a clear visible transition at 80 mg where the peak incidence of proteinuria was 17 percent. However, for all the marketed statins the frequency of proteinuria was less than 4 percent. It is very similar to the incidence of placebo. Actually, there is a typo; it is supposed to be dietary run-in.

[Slide]

This slide shows the steady state concentration of rosuvastatin. The rosuvastatin plasma concentration compared 20 mg, 40 mg and 80 mg, and these values were compared with patients who developed rhabdomyolisis or renal toxicity. There is no overlap in exposure among the patients who received 20 mg and patients with renal toxicities. There is a small overlap in exposure among patients taking 40 mg and patients who developed toxicities. However, one-third of the patients who took 80 mg had steady state plasma concentrations of 15 ng/ml, which is the lowest concentration associated with toxicities. Therefore, this slide suggests that any drug-drug interactions or using special populations may result in steady state plasma concentration

elevations similar to patients with these rhabdo. cases.

[Slide]

This slide shows the percent change in AUC and Cmax. Cyclosporine can increase exposure seven-fold. Gemfibrozil increases exposure two-fold. Japanese ancestry increases the exposure two-fold. Patients with severe renal insufficiency, creatinine clearance less than 30, had increased exposure about three-fold. These increases are considered clinically significant and require special consideration in dosing for patients.

[Slide]

Therefore, the highlighted statement was incorporated in the label under precautions: Pharmacokinetic studies show 2-fold elevation in median exposure in Japanese subjects residing in Japan and in Chinese subjects residing in Singapore compared with Caucasians residing in North American and Europe. These increases should be considered for dosing decisions for Japanese and Chinese ancestry.

[Slide]

Based on the finding of PK/PD modeling,

the following dose and administration was incorporated in the label. For hypercholesterolemia and mixed dyslipidemia, baseline LDL lower than 190, the dose range is 5 mg to 40 mg once daily. Therapy should be individualized and the usual recommended starting dose is 10 mg. However, 5 mg should be considered for less aggressive LDL reduction or predisposing factors for myopathy.

[Slide]

In dosage and administration in the labeling there is a limit for the maximal doses as well. Patients who are taking cyclosporine should not exceed 5 mg. They should use only 5 mg. Patients who are taking gemfibrozil should not exceed a dose of 10 mg. Patients with severe renal impairment should not exceed 10 mg of rosuvastatin.

[Slide]

So, my conclusion is that although the sponsor has proposed doses of 10 mg, 20 mg, 40 mg and 80 mg, the exposure-response relationship clearly shows doses lower than 10 mg have a potential clinical utility. There is apparent relationship between adverse events and plasma concentration of the drug. Therefore, findings

from exposure-response relationships were used in recommendations for dosing adjustments. That is my last slide. Thank you.

DR. VENITZ: Thank you, Hae-Young. Any comments or questions by the committee? Let me make a comment, Hae-Young. If I look at your slide number nine that discusses the dose response of safety and the topic that we are discussing is end-of-phase-2A, here you are making the argument that the incidence of CK elevations goes up quite dramatically after a dose of 80 mg. I don't think that at a 2A stage you would have had that information. This is really looking at, I am assuming, a phase 2 and phase 3 large database in order for you to be able to assess 0.2 and 1.0 percent prevalence of adverse events. Is that true?

DR. AHN: I agree with you because in all the phase 2A trials there is no way you can find CK elevation.

DR. VENITZ: So, as far as the end-of-phase-2A meeting is concerned, the only contribution that exposure response would have been able to contribute is not based on safety because you wouldn't have that safety information at that

stage.

DR. AHN: But there is a possibility you can measure proteinuria in phase 2A.

DR. VENITZ: Okay, and that is at a high incidence so you would have a better chance of seeing it in 2A. Any other comments? Go ahead.

DR. SHEINER: Let me follow-up on that. You have to know the chemistry, the pharmacology and all that, but if you believe that these drugs are sufficiently similar both in mechanisms of efficacy and toxicity, then you could argue from the Baycol experience. So, the question is at what point what are there prudent plans for going beyond phase 2A. You could argue that maybe at that point in time--I don't know where it occurred in the history of this whole story, but it could be argued that it might have been prudent at that point to have a plan to look very closely at the higher dose, both from the point of view of whether it added enough efficacy to be worth it and whether it was toxic. Again, you know, hindsight always gets you there, but you could say that even without toxicity data on the drug itself you might have been able to say something.

DR. AHN: Actually, this is true because

safety is one issue but efficacy is the other issue. When the company titrated from 40 to 80 the LDL reduction was very small. So, that is one issue we can discuss.

DR. VENITZ: Thank you again. Our last case study is going to be presented by Joga Gobburu.

DR. GOBBURU: Dr. Venitz and Committee, I will be presenting a case study, from the same team you have heard so far, on the utility of an interaction between the agency and the sponsor early on. The drug I am going to present is a very simple, straightforward application of quantitative exposure-response analysis. So, the key point I would like to highlight here is not the methodology of quantitative analysis but, rather, the progressive thinking of the agency.

[Slide]

The drug I will be presenting is being developed for symptomatic benefit and is proposed to be given once a day. Clinically it is desired to have a sustained effect over the dosing interval, that is, 24 hours. However, the drug exhibits a short half-life of two hours. In this setting, typically we don't see large clinical

trials. They are relatively smaller clinical trials. However, for this particular drug the sponsor elected a relatively large pivotal trial and the data from those trials were analyzed both using conventional and experimental analysis methods.

[Slide]

Let's briefly look at the development diary. As with any other compound, we had preclinical data and data from early drug development, including proof of concept and the PK/PD information in a small target population. So, there were data available in a target population for the intended effect. Then it was followed by the pivotal trials and regulatory review, which is about ten months.

[Slide]

Let's focus on the regulatory review box. The conventional analysis clearly showed that the treatment beat placebo. The endpoint was change in symptomatic benefit at trough versus baseline. So, by conventional means it met the primary analysis goal.

As I said earlier, the drug is supposed to be a once a day drug. However, the magnitude of

effect was small to modest, if at all. Then, given the fact that the terminal half-life is short, we don't need any modeling to come up with the question to ask whether this drug is really for once a day use.

[Slide]

But we do need the quantitative exposure-response analysis to answer the question in a very definitive manner by first answering several of these questions, such as is the effect in the first place, indeed, concentration-dependent at all? If so, is the concentration-response relationship, indeed, linear or nonlinear? Why that is important we will see in the next slide. If there is a delay between PK and PD, even though the drug is eliminated with a terminal half-life of two hours, the pharmacodynamic effect could persist for a long period of time. Is there tolerance that is being developed over the dosing interval? Importantly, is the toxicity concentration dependent? If we have answers for all of these, then we may have a proposal--if it is not a once a day drug, what are the alternatives?

[Slide]

Let's get the toxicity out of the way. It

was concentration dependent so there are limitations on how high you can push the concentrations beyond what was studied in the drug development. There was a clear concentration-effect relationship and no considerable delay that was estimable between the PK and PD. The relationship was nonlinear, meaning that having higher concentrations would prolong the duration of the effect but will not increase the magnitude of the effect. However, we have to keep in mind that the toxicity was also concentration dependent. So, we can't push the dose any higher.

Now, all this analysis, for all practical purposes, was conducted by the agency and, unlike the conventional analysis which used the trough measurements only, the whole time course of the effect at several locations was used to utilize the data collected in these studies to the maximum.

With respect to the time course of concentrations, the graph you see on the right-hand side has time on the X axis and concentrations on the Y axis, and there is a dotted line with the EC50 estimated using quantitative analysis. As you see, at about six hours, if we agree that EC50 is a reasonable target for the concentrations, the

concentrations go below this level and then sustained effect is compromised. Clearly, modeling demonstrated by answering all the questions posed in the previous slide, the inadequacy of once a day dosing, at least for this formulation.

[Slide]

Quantitative analysis has offered us more, meaning what could be done to ascertain sustained effect over the 24 hours. So, you know, it is a very simple simulation. What if you give the same dose twice a day or thrice a day or, more practically, this graph shows that sustained release may be a reasonable alternative rather than this immediate-release formulation. So, as you see, with the more frequent administration the concentrations lie above the EC50 value and they assure that the effect is sustained over the dosing interval.

[Slide]

Regarding the drug development diary, we identified that the lack of sustained effect across 24 hours was a deficiency and that the sponsor needs to address that in the next round. We also encouraged them to consider more rational dosing strategies. What that has led to is an extension

of the drug development program by probably three to five years. These are numbers that I have made up; I have no clue as to how long it usually takes to redevelop the formulation and recruit patients and conduct the pivotal trials. But the review will again be about six months.

[Slide]

To summarize the exposure-response analysis, first use of all the data collected in the trial, supportive evidence for effec in addition to the conventional analysis. It also aided in judging that once a day dosing is probably suboptimal and eliminated the need for testing higher doses but, rather, to focus on alternative dosing strategies because concentration-dependent toxicity was observed, as well as that the effectiveness was clearly plateau-ing at higher concentrations.

[Slide]

Now, if we rewind the development process and now introduce an end-of-phase-2A meeting somewhere before the total trials are undertaken, since we had the data from the proof of concept and target population earlier on, it would have been possible for us to first comment on the agency's

view about the sustained effect over the dosing interval.

So, early studies, as I said, were available. Of course, the availability of the data--I mean, we have to make sure that they are properly analyzed before such a meeting takes place. It would have been very clearly communicated to the sponsor that the optimal dosing is expected not just a p value of 0.05. That would have led to a considerably smaller study because we don't need to power the study to get the significant p value and need a large trial. Ultimately, probably it would have led to improving the efficiency of drug development.

[Slide]

Finally, I would like to acknowledge our team, DPE-1, Division of Pharmaceutical Evaluation Pharmacometrics Team and the director and deputy director and their support. Thanks.

DR. VENITZ: Thank you, Joga. Any questions for Dr. Gobburu?

DR. SHEINER: I don't question that had they been able to look at what they were aiming for they could have designed a better phase 3 to get that, but I do question, and you admitted that you

made up the numbers--do you think the FDA would have demanded new pivotal studies at the end? I mean, wouldn't it have been enough to show that the new preparation sustained concentrations over that period of time? If you had a good concentration-response relationship, wouldn't that be enough to argue that that was adequate?

DR. GOBBURU: Well, I am going to be very careful in answering this. I thought that somebody from the company would ask me this question. The very fact that there is a concentration-dependent effect and that we are testing new regimens, there is some uncertainty if you take the interdisciplinary team into account.

I have two points to say about that. One is are we in that way supporting poor drug development, meaning it is okay to do a suboptimal study and then, since you have a model, we don't need to do anything else? The second point is that there is definitely a mixture of empiricists and modelers, Bayesian modelers here. So, there has to be empirical evidence. If I have to take a stand I would say that there has to be empirical evidence with the other dosing regimen.

DR. SHEINER: I think we can discuss this

more later but it certainly is true, for example, that drugs have been approved at doses that have never been tested.

DR. GOBBURU: That is true.

DR. SHEINER: Especially if you bracket it with one below and one above and it really looks like the one in the middle, which you didn't test, would really do a better job and you have nice dose response, toxicity and efficacy. So, it sort of sounds like you are giving and taking at the same time and it is really tough. I mean, if you are saying that science is going to be helpful here, then you want to, you know, sort of follow that through.

I think the agency has to think about what its policy is and to what extent it will rely upon good empirical evidence that the drug works, good empirical evidence of what the concentration response is and, therefore, extrapolate or interpolate to a place that says, well, we know what is going to happen if we do this because we know what happens if you give more, if you give less, and so on. I mean, there has to be room for that. You can't just say that everything has to be empirically demonstrated.

DR. GOBBURU: If you are increasing the frequency of dosing and we have never seen any safety information about increased dosing, it is just a black box. We have no clue as to what to expect. So, I would still stick with my stand that we need empirical evidence.

DR. DERENDORF: We don't know what kind of a drug it is and what kind of an indication it is used for but conceptually you use the EC50 as your target. Now, EC50 is the concentration where you have 50 percent of the maximum effect. It doesn't tell you anything about where you stand in terms of therapeutic benefit. Actually, 30 percent concentrations below the EC50 may still have considerable therapeutic benefit. So, I am not sure if that is a given cut-off that you can use.

I think the second part of the question is you said the dosing regimen is not optimal. Does that mean that if you have a suboptimal regimen that you propose that it would be acceptable from the beginning? Again, you could have a suboptimal regimen that is still of great therapeutic benefit.

DR. GOBBURU: Okay, these questions are very hard to answer because you are asking me a question about what the target effect is. I think

the meeting here is to really move from the conventional analysis to bring in more advanced technology in order to optimize the therapy. I do agree to that. But today we do not have--for example, for this indication the target effect that is acceptable, nobody gives us that number. That is why when I presented the curve I said if EC50 is accepted as a reasonable target concentration. If you want to choose 70 percent or you want to choose 20 percent, that is fine but, still, you look at the effect curve over time and it is going back to baseline at about six hours. There is no question about that.

DR. KEARNS: I think that is true but it is important to step back for just a minute. I mean, certainly the technology and the modeling--and all of us can understand when it drops below some threshold number, but what if it was a drug and a disease where the relief of symptoms extended beyond the time when the concentration was below the EC50? Because in that instance it can be argued that the need to push a sponsor into another three to five years worth of study with a new formulation and more pivotal trials may not be wise. In fact, that would be

contrary to the strategic plan of the agency now, which is to effectively collapse drug development.

So, dragging this in early, Larry, as you mentioned with using the medical expertise in addition to the kinetic, dynamic modeling expertise I think is critical because at the end of the day you want to make the best decision for the life of the compound and its development, not necessarily say, well, we have created more questions; now we have to make answers to them.

DR. GOBBURU: If you look at question number three, if there is a delay between PK and PD, if that is true, we would have found it and we systematically tested for that. So, I am not presenting this example saying that we didn't take the time course effect; we did.

DR. VENITZ: Go ahead, Wolfgang.

DR. SADEE: I think one of the critical questions is whether you really have enough information at the 2A step to decide here is your threshold; here is what you titrate for and that is how you go forward in designing the trial and you then come up with a relatively arbitrary sort of threshold, let's say the EC50 or something like that. Or, in the previous case with the statins

you base your decisions on LDL cholesterol which is a very crude measure and, in addition, one that is not forward looking; it doesn't tell you possibly anything about the eventual outcome as to how this should be used. Personally, if I were to be put on this particular statin I may have started out with 2 mg, depending on what the case is, or 1 mg and that could have been just as effective.

So, given the complexity I am just wondering-- you said we want to bring in more technology or more science, that would mean more information. For instance, in the case of the statins I would say, all right, let's look at the different sizes of LDL and HDL and how that is affected by the different dosage levels and get a little bit more information on it. Then it may be worthwhile to come in early. So, I am just raising the question, after hearing the discussion, as to do we know what to recommend at that point?

DR. VENITZ: Can I just make a statement? Let's just focus on the presentation and we may have a general discussion after the break. I think you raise a very important question but I would like that to be discussed after we have done with the individual cases. So, if you want to respond,

feel free.

DR. GOBBURU: Thank you. Dr. Lesko can comment more about this. I don't think the intention of these meetings is to pin-point exactly where to go. As long as we have a range of options the drug development could be tailored accordingly to answer those uncertainties. So, in this case, I agree that we didn't know what would have happened if you had given the doses repeatedly over the day. But we have identified the inadequacy of this once a day dosing so that has definitely opened up new avenues that need to be explored. So, I don't think we will ever have a precise answer at the end of phase 2A but at least we may have a more precise direction to go forward.

DR. SHEK: Just a general question, I wonder whether this example is a good example. First, looking at the drug development diary, it looks like it took ten years to develop it, which maybe is on the high side. Then if the boxes are linear there in the diary, it looks like a long period of time, which I would assume is a phase 2 study. If you just think back, I mean some of those questions should have been answered. So, I think something was going on with this project and

I just wonder whether that is a good or typical example.

DR. GOBBURU: Well, as I said in my presentation, I have no clue about these numbers. I just made reference to the numbers so that we will have a time frame and a ratio of the period that--extra time needed to redevelop the drug when compared to the original drug development time period. So, the ten years--I have no clue how long it took the sponsor to develop it; it could have been five and a half but relatively there is a 20 percent to 30 percent increase in time, I would guess, because they had to go back and revisit the dosing issue. So, it is just a ratio you should be looking at.

DR. LESKO: Yes, I think the three to five years was just a speculative estimate, you know, trying to make the point that whatever analysis occurred at the late stage led to a need to reformulate and some additional trials. Now, what those trials might have been is still open to question. As Dr. Sheiner pointed out, can you use the exposure-response relationship and treat this in essence as a therapeutic equivalence situation and look at comparable blood levels from a revised

formulation, and if there were additional efficacy data needed, what would be the size of that study. So, I think it is an open question there.

I think the point of it though is that this analysis occurred at the end of the game, a ten-year process when the NDA was submitted. It wasn't adequate and the data was available early on. So, I think it was trying to represent the type of information that could be used more optimally earlier in drug development. Yes, you can approve drugs based on doses that are effective and not necessarily optimal. I think one of the goals of this strategy is to try to move from just effective to something more optimal, taking into account the type of issues that we have seen in this case and the prior ones.

DR. VENITZ: Any other questions or comments for Joga's presentation?

[No response]

Thank you, Joga. We are going to get an early break. It is now 10:25. We have a 20-minute break so let's get together at 10:45. So, the committee reconvenes at 10:45 for the discussions.

[Brief recess]

Committee Discussion

DR. VENITZ: To get us started on our discussion I would like for Dr. Lesko to review the three specific questions that you have in your background material that he would like to get some feedback on.

DR. LESKO: These are the questions that we wanted to bring before the committee. Just to summarize this morning's session, what we tried to present is a framework for thinking about improving drug development through a new initiative that would bring the agency and the company together to discuss, in specific terms, the dose response and the rationale for dose selection and dose-range selection as the drug development program moves forward.

As a secondary objective, we also see this as an opportunity to review the overall clinical pharmacology development plan with respect to what the drug interactions are, special populations ara, and any formulation issues to try to come to some sort of agreement or dialogue on what is necessary in a particular case.

So, what we presented today--again, we recognize they weren't the technology underneath what was presented but each of those cases involved

the usual technology of modeling, simulation, predictions and so on. More than the technology, what we really wanted to get some reaction to today was the general plan to move forward. As I mentioned in my introductory comments, this is really the first time we are discussing this publicly and the Center would like us to develop a guidance in this area and make it available to sponsors in the sense that it would lay out the goals and background information, and so on.

So, what we are looking for today in these questions are your thoughts on the proposal that we have put before the committee, the rationale for it, any ideas you might have on how that could be improved, and any obstacles that you would anticipate from your own experience that would limit the success of this program.

The second question--we presented some examples of analysis and there were some comments with each case as it was presented. But, hopefully, it gave you a flavor for the types of things that might be discussed at this meeting, obviously dependent on a case-by-case basis.

Then, the third point is that we have been asked by the Center to develop some measurements

and metrics for measuring the success of this program in the sense of continuing it and adding more resources to it as we move forward.

So, these are really the three broad areas and certainly any comments would be appreciated, or anything else that we haven't thought of in terms of these three questions.

DR. SHEINER: First, let me say that I think it is a good idea but I am not exactly sure why and I think we need to think about that, or at least I do. So, let me just say that we even accept--I mean, there are people who would argue with this but let's accept for the sake of argument that there is insufficient use of prior existing data in the planning of the later stages of drug development, to put it very broadly, and in particular with respect to dose or regimen that is going to be tested in later phases. That prior data consists of, you know, science which generally people agree is known; public domain type data, actual numbers and data that is out there that you could incorporate into your analyses; and then there is proprietary data, the stuff that the manufacturer has been developing in the course of phase 1 and whatever comes before this meeting.

So, let's assume that they are not adequately taking advantage of that, as we see it, in planning what comes later. The question is what is the cause? Because you come up with a remedy in a sense. Without being a little facetious, if the remedy is a meeting in which you help them figure out how to use this data, it means they are not smart enough to do it themselves. That is what you have diagnosed as the cause and I don't think that is true. I think there are a lot of very smart people and obviously you do too.

So, what is the reason that the smart people in the pharmaceutical industry who are perfectly capable of looking at the data when they change hats and go to work for you or change hats and go work in academics, or whatever, why those same people in industry are not doing that, and why could looking at these things, the kinds of examples we saw which are not, you know, rocket science, why is that useful and why does it look like it would have been useful to do that and why didn't they do it?

I have thought about this a lot and a lot of people have thought about this a lot, and I am sure there are as many reasons in our minds as

there are people in the room. So, the question really is will this particular action, which is offering help, aid, guidance--will this help to get over whatever the reason is that they are not doing it themselves? Personally, I think calling attention to the whole issue and making a point of saying it is important, important to the regulatory agencies, will be a help because I think there are institutional reasons why it isn't happening which would, to some extent, be mitigated by doing that.

Remember, I made a suggestion here the last time or the time before where I said, you know, maybe for a while the FDA could try saying you have to give us some reasonable decision analysis-based argument for why we should approve the dose that you are asking to be approved. Show us one efficacy endpoint, one toxicity endpoint and some utility function and a computation and data. Not that that is required for approval; we are not changing the rules but we just need one of those things before--you know, that is part of the dossier.

I was addressing the same issue. I said let's make people think about it and maybe if they have to think about it they will find that it is

useful. Here you are not quite making them think about it. You are offering them the opportunity to think about it with you, and that is a little gentler and maybe it is a good idea. But I do think we should spend a little while thinking about whether this is the most efficient use of your time and effort to overcome that problem which doesn't look like it is because they are too stupid. That is not the issue. There is something else, some other reason why it is not happening.

DR. LESKO: And it is an excellent question, and it is one we have asked during the sort of roll-out of this internally. We talked about the facts that I had on one of the slides about the failure rate of clinical trials. That number comes from the industry; it doesn't come from us. We don't know actually what the underlying reasons for those failures are. I don't think that has been studied in a systematic way.

Some of the observations that we have are, for example, instances where a single dose is chosen for phase 3 trials. We have tried to encourage more dose-response data from phase 3 and continue to look at that, and that was the gist of the quote I had from Dr. Temple from his

presentation at DIA. So, this might be a way to talk about that.

You are right, you did make a point at one of our earlier meetings, and this does actually represent a time at which we might ask what is the rationale for this dose and discuss that collaboratively. I don't think it is an issue of people being too dumb to know what to do. I think it is an issue of a fair amount of uncertainty in the drug development process, for a variety of reasons, and can the agency offer some experience that it has from its NDA review. Most of our time goes to NDA review and, as you know, at that point in time everything is history. You are basically looking at a document and picking out deficiencies or looking at areas where missing data might occur.

So, in terms of using resources efficiently, it seems like the efficient use would be to move the resources forward a bit and not sort of dwell upon--although we have to but not necessarily dwell more than we need to dwell on the shortcomings of an actual submission but try to improve things early on. So, part of it is sharing perspectives on dose response, which is not predictable from a scientific standpoint. When a

company comes in they don't exactly know how the agency is going to react to that assessment of dose response and risk-benefit. So, having the opportunity to talk about that earlier on I think allows one to be a little bit smarter about the way to move forward. But there is uncertainty here.

The alternative ideas for looking at the problem, there aren't very specific suggestions that I can think of. So, we look at this as a pilot study; look at how it goes; and see where there are improvements to be made.

DR. KEARNS: Larry, I think you just said it very much as a cart and a horse issue here. I mean, right now if your shop is brought in at the point of time of NDA review, with all the new technology it is easy to see the gaps. Then, as you go back and interact with the review division or the sponsor and begin to address ways so those gaps could be, or should be, or must be filled, then that has a definite impact on the process.

I think there are a couple of key elements to doing it early and I support the integration of clinical pharmacology early in the process. Number one, when you go into that meeting with the sponsor not only does it have to be, quote, informal--we

know those interactions are never interpreted as informal by a sponsor, but the expectations that might be set out based on the information that is available have to be plastic because we all realize that in the subsequent process of drug development new information is going to come out that may cause us to go back and even make a mid-course correction or change. So, all the parties at the bar have to realize and agree with that and abide by it.

The other thing is that what clinical pharmacology does and what the medical people in the review division do have to be congruent, and it has to be congruent at the beginning of the process not brought into some congruence at the end of the process. I know those are more political than practical--well, maybe they are practical comments but I think it is workable if it is done right.

DR. LESKO: When we discussed this internally with the different units of FDA that was an important principle, that this would be a collaborative meeting and there has to be congruence in order to make this work.

We have had some experience with the informal meeting and I imagine this meeting would be similar to, say, meetings that we have had as

informal meetings on the integration of genetics into drug development. This is an area of sort of evolving science as is, in some ways, the analysis of exposure response and modeling and simulation evolving. The meetings have been I think successful for everyone concerned, but it does have a little more of an acknowledgement that benefit-risk is a changing thing as you move through drug development. I think the informal meeting recognizes that. The atmosphere is different in those meetings, as I think it would be in this meeting as well.

DR. SADEE: I want to reflect a little bit on what Lew said. The question is what is the purpose? If the purpose were to avoid error being made, that is easily picked up and that may not be the purpose because, as you said, there are lots of smart people out there who can look at this rather reasonably.

But I think what you said that if an early stage a strategy is being devised to look at dose-response curves, and so on, and dose effect relationships, and that strategy could be viewed and kind of agreed upon--but that may be dangerous too because it could lock the agency into

something--well, you agreed to this and this is the way we are going to go forward, and it turns out to be wrong. So, I think a way has to be found to say that the purpose of the meeting is to just give you this and, just like you said, to indicate that this strategy might be a good way to finding what the real relationships are and what one has to look at and do this in a quick way. That would make sense to me.

DR. LESKO: One of the things that frequently characterizes the other type of meeting, a formal end-of-phase-2 meeting are specific discussions of study design, endpoints, statistics and so on, and I can imagine a meeting of the type we are talking about that would actually not necessarily be question based. It could be discussion based or exploratory based or informational based where people might discuss alternatives based on analysis of data, and there might be a sharing of experience between a sponsor and ourselves. It would be informal in that context. I think that would probably be characteristic of this meeting.

DR. VENITZ: First of all, I am very much in favor of having this at least as an option and

as something that we want to review on a regular basis to see whether it actually has an impact. But I look at this more as an evidentiary hearing, if you like, where you are not necessarily reviewing the evidence based on the merit but what are the rules of evidence.

What do you think down the road in five, six years, would be evidence that is necessary to support an optimal dose? Are you going to at least be willing to consider biomarkers, something that I didn't see in your discussion? I think this, to me, is a key point in terms of assessing potentially biomarkers. Obviously, this should have been discussed pre-IND but at least at that stage you have some experience. You have some proof of concept possibly for biomarkers on efficacy. You may have some at least potential biomarkers of toxicity. All those are things that I think should be discussed not necessarily in terms of how they pick the right dose, but what kind of evidence would ultimately be needed for biomarkers from exposure-response modeling to support an optimal dose and to, hopefully, speed up the process of getting to approval.

DR. LESKO: I agree with you. I mean, I

think at this point in time there is usually a fair amount of biomarker data available, if not clinical endpoint data. One of the ideas of having this meeting is to look at things a little more mechanistically and integrate this information in a way that actually isn't being done very much at least by ourselves at the NDA stage where we tend to look at clinical endpoints.

So, I think the idea is to look at this in a quantitative mechanistic way and integrate information perhaps in a way we haven't done before as part of the interactions with sponsors, and doing it in a sense of trying to improve things as opposed to being an obstacle, I suppose.

DR. VENITZ: I think part of the discussion has to be what is the payoff. If certain things turn out the way you expect them at that stage, which is obviously affected by some degree of uncertainty, what is the payoff? What is the improvement on your side as well as on the sponsor side? Otherwise, while we are doing those studies, we still have to do a formal study to prove whatever needs to be proven. That is what I am concerned about.

DR. FLOCKHART: I guess to put it bluntly,

to me, it is a tradeoff between whether this would really make drug development better, as you point out, versus would it just be another piece of red tape, another hurdle that people would have to jump through.

So, my question would be what are the alternatives. If you look at it historically, presumably in the old system we are saying, you know, we are very worried about this because the number of submissions is going down, and all the rest of it, but we had this system in place when they were going up as well before 1996.

So, I guess an alternative might be to look at that from a distance. Okay, so why don't we just issue some good guidances, like you have done, in the interim period before the end-of-phase-2. These would include the kinds of things you have done on drug interactions, in vitro and in vivo and on PK/PD and a large number of other things. So, a way of thinking about this might be whether you consider those guidances to have been ineffective and whether they are not having the desired effect in terms of improving--I mean improving, not speeding necessarily but improving drug development, and what effort--this

is kind of like an alternative resolution on the floor--would effort put in the area of more consolidated or more effective guidances be as good as having a meeting like this?

DR. LESKO: I don't know whether that was a question or not.

DR. FLOCKHART: I am really speaking to the wisdom or lack of wisdom of having meetings like this. I think the question I am posing really is are there better alternatives and what do you think about them?

DR. LESKO: Well, we think, and industry really can better speak to that--we think the guidances have helped drug development and helped clarify regulatory thinking. We see a guidance as helpful in this initiative as well to lay out the goals and objectives. As I mentioned in my introductory remarks, this is a voluntary type of meeting, as are the other meetings, and we have sort of talked to companies about this as part of our interaction with them in the normal day-to-day business and the reaction has been positive in terms of the counterparts in industry to the clinical pharmacology group here, at FDA. Whether that positive feeling is pervasive through the

regulatory affairs and clinical departments we don't know. But the initial reaction has been very positive.

But I think the way forward is to put the guidance out as a draft guidance; get some experience with this type of meeting, and we think it will be at least two or three years out before we have enough examples of this to determine whether this has been helpful or not. But we need to get feedback from each individual company that would come in for a meeting like this and look at how that impacts the subsequent NDA that we had meetings on. I think we can look at this somewhat systematically and see what impact it might have.

DR. SHEK: I agree with the guidance, that it is helpful, as well as the meeting. I look at that from the industry perspective. It is more setting up expectations as you go through. Guidances are fine but, you know, they are still open to interpretation and a specific case might be unique. It is also an opportunity for the FDA maybe to see some of the data that has been developed. So, I see benefits there.

But, still, we have to look at the bigger picture and that was my question earlier, how many

of those cases--we are saying 50 percent of, let's say, programs in phase 3 are failing. I know from my own experience that the target is, you know, once you go into phase 3 studies you want to be pretty sure that you know it will be a success. So, out of that 50 percent, what are the reasons for failing from a regulatory view? I would assume some of them are failing even by the company itself. Once they have the data, they say, well, we don't have the product here and they don't even submit an NDA. Or, the scope doesn't fit when they will try to position it into the market so it takes longer. But then if you take those out, how many of those are failing because the dose was the wrong dose and how many of those are failing for other reasons?

So, I would assume the FDA is in the same position as the industry. If you have the resources and they are limited, where do you spend them and when do you spend them? So, I think here it would be interesting to go into that and maybe this two-year pilot will bring us some of the information.

Saying that, basically I believe it picks up from the FDA strategic plan, whether this

specific proposal will improve or to make innovative medical product development sooner and then, the other part, also developing safe and effective medical products. As I understand the proposal, it looks like let's tackle drugs that we know how they work and how they are effective. I wonder whether that is the target of drugs that you would like to look at or, rather, look at those maybe new breakthroughs where we really don't have a therapy this year. Maybe those should have more time spent looking at the system.

DR. LEE: I just want to clarify that the guidance that Larry just mentioned is a procedural guidance, which is a guidance to industry regarding how the sponsor can request a meeting, not a guidance to discuss drug development.

Secondly, to answer that question regarding the reason for failed NDAs, in the ten NDAs we looked at one of the most common reasons for failing is that the dose chosen was not optimal which led to lack of efficacy or safety problems. But I agree that it would be useful to look at not only the failed NDAs which have already been submitted, but also look at the failed phase 3 studies and see what the reasons are for the failed

phase 3 studies.

DR. HUANG: I was going to comment on guidance. I guess you said there are alternatives to communicate and we do have a lot of guidance documents. So, those may be helpful instead of additional ones. That is what I take from one of your comments. The guidance is a living document. For example, the Drug Interaction Guidance may not be updated and we have new information that we may have just learned from reviewing certain NDAs or company meetings where we know some other factors need to be considered.

For example, Ameeta has shown an example where QT prolongation, if not evaluated properly, could be a cause for approvable instead of a first cycle approval. We did have quite a few examples. To communicate this information, this could happen when we have this type of information. I mean, some of the examples show that information comes in later and we might have communicated at end-of-phase-2 or pre-NDA. However, if you can do it earlier we probably can share the information early on with the sponsors with the current information or different interpretation based on the science which may not be covered in various

documents already in place.

Larry has mentioned about pharmacogenetics. With the information that we have right now, how do we learn about the information that industry has or how do they know what we will see as issues? This type of information, even if we have quite a few informal meetings, that is not exactly end-of-phase-2A but I think they have provided an opportunity for us to learn what are the issues that a company is facing. I think what we heard is valuable on what questions we would have when we see certain data that may not have been submitted early on.

So, I think this offers an opportunity not only, hopefully, I think to be beneficial for the sponsor but also very helpful for us. Once we learn this information, we can also communicate it to the other sponsors.

DR. MCCLEOD: I think it is a good idea but I am not sure why. I didn't find any of the three cases especially compelling. The reason why, as I thought about it, is you can't retrospectively reconstruct the data if you want to really answer whether this is a good thing to do or not. As you look back, there was great data that at the end you

could have looked back and made a better choice, but not at the end-of-phase-2A. At the very end of the study you could have.

I think maybe, if nothing else, going through this two-year pilot, whatever the time is, will at least allow you to construct the data and to come back and say that this is something worth doing or that this is really no more insightful than we have now. We really don't have enough data to say this is a good thing to do. It seems like a good thing to do. It should be a good thing to do but the examples that are out there don't say, yes, this is definitely something that is going to really improve the development of these drugs.

DR. SHEINER: Again, putting the best possible light on it, let's imagine that, first of all, the basic hypothesis is true, that there is more information to be gathered from early drug development that is relevant to later drug development than is being fully exploited. Let's grant that and then let's also grant that the pharmaceutical industry in general and companies are trying to find a way to better exploit that data and that they might find this kind of a meeting useful. Even given those two things, you

know, you sort of can't do any harm except for the cost in time and effort on the part of the FDA and that is a finite resource, and it is not holding anybody's feet to the fire and it is not making new rules, or anything like that, which is something that, you know, obviously would cause a much bigger shakeup.

You know, I am just sort of trying to get to Larry's third question. I have no idea then, if that is the case, what you would use for a benchmark other than customer satisfaction. I can't think of how you would try to actually quantitatively measure the influence because, as I think you just pointed out, it is likely to show up in the quality of the data that is gotten after that meeting and it is very hard to say, well, it would have been otherwise or wouldn't have been otherwise. It is the same problem going forwards in a sense as going backwards and saying, you know, make believe I didn't know the end result now what would I have done back then if I had been faced with those data? It is just almost impossible to do.

So, I don't think you can measure it. I do think that it can be seen as a positive

endorsement of the idea of better exploiting all these data in a quantitative way that takes account of all uncertainties and tries to allow decisions to be made. I think in that sense it is a public service, but I don't know if you are going to be able to measure the impact.

DR. MCCLEOD: You could do a randomized study of offering end-of-phase-2A consultation or not and see whether the doses are picked correctly.

DR. JUSKO: I see this as a good idea from the viewpoint that it offers the companies a chance to interact with the FDA probably for problem situations. I kind of view 2A studies as proof of concept and none of the examples that we saw were really phase 2A situations with the great uncertainties that frequently exist.

I was a little bit concerned by what Larry said early, that oftentimes at the end-of-phase-2 meetings the companies are already wedded to an array of plans for phase 3 studies and may have difficulties making adjustments in those plans. The examples that we saw were more of that ilk. So, this kind of proposal could offer opportunities to influence what would be happening in making plans for phase 3 studies earlier in the whole

progression of things. So, in that context it seems like it could be very beneficial in certain situations.

DR. LESKO: It has been interesting, in discussing this individually with companies, whether or not this is even an early enough meeting to discuss the issues we proposed to discuss in this meeting. Dosing strategies are set individually by different companies in many different ways but this seems to be a fair balance.

The other thought we had on this, and we have begun to explore this, is the introduction of some discussion of disease progression models as part of this meeting, and determination of whether or not this might have some impact on the way exposure response is assessed and if that would have a positive impact on clinical trials in specific disease state areas.

We are doing some ongoing research in certain diseases with disease progression models, and we have used it before in our analyses in selected cases but we think there is some potential to look at this more fully in the context of these meetings, again, with the collaboration and agreement of the company to do this.

DR. VENITZ: Are there any more comments for question one because I think you got a lot of feedback from the committee? So, any more comments about the general objectives of this end-of-phase-2A program?

[No response]

Then let's see if we can focus on the second question. That is a more methodological question. What approaches can be used in order to maximize the efficacy, I guess, of those end-of-phase-2 meetings? Any comments by the committee to question number two?

DR. SHEINER: Just to beat the same horse as before, obviously they are going to want to do the analyses in a sense. I mean, you are going to sort of help them out and make suggestions. But I do think that some attention to some kind of value function--call it utility, whatever it is--where you say, you know, there is something we are trying to learn here in particular; we have some measure of what we are trying to learn, rather than everything there is to know about concentration response and all possible responses. I am sure you would never say that but some formal attention, some agreement that one of the things you are going

to talk about--not formal because it is an informal meeting, but some agreement that one of the things you are going to talk about is how you are going to measure the value of what you are going to learn.

DR. VENITZ: I would echo that. I think a lot of the things we have seen were retrospective data analysis and I think one of the objectives of this end-of-phase-2 meeting may be to decide or at least give guidance on which issues need to be studied in a prospective manner as part of a prospective study, be it a clinical or preclinical study. On the other hand, which other issues which may be playing for lower stakes can be dealt with retrospectively as part of some kind of a population PK approach.

Again, just give guidance to the industry for what the stakes are for the different issues that are going to come up down the road, and what is the potential payoff if they improve on the way the analysis is being done.

DR. SADEE: So, what you are saying is identifying the problem issues as far as they can become apparent so that there is already a foundation that would save maybe energy later for the FDA because the issue is already at hand.

There may be new issues emerging, but I would imagine that at that point one would know what the key questions are. That would be very helpful.

DR. VENITZ: And one component that didn't really get any discussion time today is to incorporate enough preclinical information, both in vitro as well as animal pharmacology, safety and toxicology information that may be quite relevant at that early stage. How would that impact not only on endpoints that may need to be monitored but also in terms of dose selection, including using qualitative methods?

Any more comments to question number two?

[No response]

Then let's look at question number three. We already heard Dr. Sheiner's recommendation that customer satisfaction might be the only measurable outcome. Any other recommendations or suggestions by the committee?

DR. DERENDORF: Well, it is actually under strategic planning. It is steps to reduce the time, cost and uncertainty of developing new drugs. So, that is the goal and I think that can be measured. You said that in your examples there were a lot of components that were dropped because

of the wrong dose. That number should come down.

DR. LESKO: That is true, and there is another conceivable metric one might look at, and that is the dose changes post-approval. There is published literature on that recently by Jamie Cross and colleagues, looking at dose reductions post-approval in terms of the time following approval, what percent reductions were downwards, and so on. That also might be over time another metric that could be looked at I think.

DR. SHEK: Yes, the only issue there is that in two years you wouldn't come out with the metrics I think. You would need a longer time than two years.

DR. LESKO: Yes, I agree. I think we have said two or three years. It is hard to say, depending on the frequency of having these types of interactions.

DR. FLOCKHART: I don't think it is actually very difficult. I think a simple catalog of decisions made by sponsors in itself would be very instructive. I mean, it goes everywhere from killing a drug--I mean, how many drugs got killed and what kind of decisions sponsors made in response to those meetings. You could easily have

an analysis to ask them, well, what did you do as a result of this that you wouldn't have done otherwise? Change your clinical trial design? Add a surrogate? Build in a toxicity monitor? Monitoring based on animal data or preclinical data that you hadn't done before? I mean, there are lots of potentially valuable things you could talk about that would be persuasive, simple broad statements.

DR. HUANG: I was just going to say since initially the end-of-phase-2A meeting will be limited so we will only have a few cases--this is like an open trial so we look at these cases and, like, a customer satisfaction survey including whether the sponsor changed a development plan based on the FDA input or based on this meeting. So, even though we don't have a randomized control, we do have the set of sponsors that went through the end-of-phase-2A meeting.

DR. VENITZ: Can we maybe add a fourth question? I think you alluded to that, Larry, and that is, can we as a committee identify specific scenarios where the end-of-phase-2A may be most helpful? The new drug in class or first drug in this particular class or should it be a drug where

we know a lot about the class? What does the committee think?

DR. SHEINER: But the problem is that the answer to that depends very heavily on the first question we never answered, which is why is inadequate attention being paid to the information? But my guess is that the newer the drug in the class, the receptor and all that, the less advantage you can take of prior information because there isn't any. So, you are in a more empirical mode and we know that the pharmaceutical manufacturers do a reasonably good job of being empirical.

So, my guess is that you might be most helpful in the case where there is a fair amount of knowledge and where the company maybe feels that, for some reason, it can't use that and they can be encouraged to do so for whatever is the problem that this is solving. It would seem to me it has to be most applicable in the case where there really are things that should be brought into the thought process that are not being brought in.

DR. VENITZ: I would concur with that and add that I think it might be worthwhile particularly for drugs that treat symptomatic

conditions. Again, the payoff might be earlier than for drugs to treat chronic conditions, depending on how much we know about the disease per se regardless of the pharmacology of the drug. So, actually acute indications might be the ones to focus on early on to see if it does any good.

DR. KEARNS: Larry, I think one of the things is thinking about drugs that may be useful in children and other special populations. The end-of-phase-2A meeting could be a very important point for the agency to begin to discuss with the sponsor really what kind of studies need to be done; what do we need to think about; what are the endpoints that might be appropriate. As it goes now, those questions are often asked very, very late in the game when not a lot of synthetic thinking can be brought to the bar.

DR. MCCLEOD: I was just going to ask, Peter, was there any central theme to the ten drugs where you could have predicted dose alterations? That failed because of incorrect dose? Were these all first time in class or were they all fourth time in class? Is there anything that could guide where you should be focusing this work?

DR. LEE: I am not sure. I think at least

they all have good exposure-response relationships, which means the endpoint is either a shortened endpoint or a surrogate endpoint that is easy to measure and connect to the exposure. But I think it was the clinical endpoint being used but it was a shortened clinical endpoint. Again, I think the central thing would be a good exposure-response relationship being established based on the early studies.

DR. HUANG: If I remember correctly, the majority of them is not first in the class. Was that one of your questions?

DR. MCCLEOD: Maybe what I am trying to get at is what drugs you should focus on to try to make this work or not work.

DR. HUANG: Many of those are fast follow-ups but a lot of information developed later on. So, some of the information we may not have well elaborated or well recognized when they first come up. So, some of the examples you have seen, they are the fourth or the fifth on the market.

DR. MCCLEOD: And certainly those are less interesting but might be a good place to start just because you might actually be able to intervene and see whether intervention improves things.

DR. HUANG: Yes, I think it was in Larry's slide, either that we know a lot more now than when it was first introduced, or some of them may be novel so we want to help with the development. But in a lot of cases they are fourth or fifth in the class.

DR. VENITZ: Any further comments to any of those questions? If not, Larry, I want to give you an opportunity to wrap things up before we take a break, if you choose to do so.

DR. LESKO: I don't need to take much time but we presented this morning a concept for a new initiative and I think appropriately received some excellent input from this committee. We are going to continue to move this forward and maybe share with the committee at some point in time some experiences we have with this initiative.

I believe our next step will be to develop a draft guidance for industry on this concept, taking into account what was said today, and put it out really for comments so people can raise issues, identify important aspects of it and continue to move forward.

DR. VENITZ: Thank you. That brings us to our lunch break. We will have a break from 11:30

to 12:30. Just for everybody's information, we do not have any open public speakers so we will start with the official program at 12:30. So, I would hope that all presenters will be ready at 12:30 to present on the QTc prolongation modeling. Thank you.

[Whereupon, at 11:30 a.m., the proceedings were recessed for lunch, to reconvene at 12:30 p.m.]

- - -

A F T E R N O O N P R O C E E D I N G S

DR. VENITZ: Welcome back for the afternoon session. We are continuing with the general topic of exposure response, and our second topic for today is the use of PK/PD modeling in the context of QTc prolongation. I would like to ask Peter Lee to give us an introduction of the topic. Peter?

PK/PD (QT) Study Design: Points to Consider

DR. LEE: The next topic we are going to talk about is the PK-QT study design.

[Slide]

Specifically we will be talking about using the clinical trial simulation, which is a simulation methodology for designing a PK-QT study. I want to start by saying that there has been increasing regulatory interest regarding the QT prolongation. As a result, a number of drugs have been withdrawn from the market due to the QT prolongation property. Most recently we published a concept paper regarding the QT study design. I believe there is also an ICH E14 guidance that is under preparation.

[Slide]

There can be several different objectives

for a PK-QT study design. The first may be to use the study to determine if there is a drug effect on QT. Secondly, the objective could be to estimate the extent and the time course of the QT effect. Finally, to determine the PK-QT relationship so that a relationship can be used for dose adjustment if intrinsic or extrinsic factors may influence exposure of the drugs. So, the regulatory utility of a PK-QT study could be to evaluate the safety of the drugs; to determine the dose selection in the patient; or use information for dose adjustment.

[Slide]

Therefore, there are actually many different issues relating to the PK-QT study design. One of the most significant ones could be the large and unpredictable within- and between-subject variabilities, including inter-day variability as well as within sampling window variations which can cause a decrease of the study power to identify a small change of QT due to the drug effect.

There is also a different way of selecting the baseline, sometimes one sample being selected pre-dose; sometimes 24 hours as a baseline. The sampling schedule is also an important factor that

may influence the study power and other additional issues, such as the selection of meaningful and sensitive QT metrics and the variability associated with PK and PK/PD relationship.

[Slide]

Additional issues are dose-ranging studies. Whether a placebo control or active control is included as a comparison and different types, such as crossover or sequential designs.

[Slide]

So, when we see a study report where there is an X millisecond change in QT due to a drug effect, then we have to ask the question what is the correction method being used to correct the QT regarding the R interval? What is the QT parameter we are talking about? Is it the maximum QT effect, or the average QT effect, or just randomly selected drug dosing interval? We also have to ask what this QT change is from? Are we comparing to the placebo group? And, also ask the question at what doses has QT effect been observed? Once we have answered all these questions, the most important question we have to ask is how sure are we about this X millisecond change in QT.

[Slide]

I will just give you an example. This is just an informal survey of QT studies of terfenadine that have been published in the past. I have a list of ten different studies and their study designs. The dose regimen in those ten studies ranged from a single dose, 120 mg for most of them, to 60 mg BID.

The general study design could be a sequential crossover, parallel, and the number of subjects could from 6 to over 60. The baseline is sometimes one sample; sometimes 12 hour. The sample of treatment is even more variable. It could be one sample, 6 hours, 12 hours or 24 hours. The metric of QT is sometimes point-by-point comparison with the baseline, sometimes the maximal, sometimes one sample.

[Slide]

These are the study results from these ten literature studies. Seven out of the ten studies show no effect, no QT effect of terfenadine against either baseline or control depending on whether it is a sequential study design, crossover or parallel design. If we exclude the first two studies, the single dose studies, then five out of the eight studies actually show no effect against baseline or

control.

Although this survey is really informal and may not be conclusive, we really had to ask the question whether the inconsistent results are only by chance due to inter-study variability or is it a study design issue. I think we believe it is the latter because of the variety of study designs involving these ten different literature studies.

[Slide]

So, we proposed the use of clinical trial simulations for designing a PK-QT study to address the complexity of the study design issues because it was deemed that there is no one-size-fits-all PK-QT study design. Each study has to be designed for its own specific objective. You have to consider the variability of PK/PD. We can use clinical trial simulation to explore a variety of study designs and integrate the effects of all study design factors into the considerations. The trial simulation can be used to estimate the study power to achieve the specific study objective and it also can be used to address "what-if" scenarios under different possibilities.

[Slide]

So, today we will have two different

presentations. The first presentation will be given by Dr. peter Bonate, from ILEX. He will be talking about the use of clinical trial simulation for PK/PD QT studies. The second presentation will be given by Dr. Leslie Kenna and she will be talking about the QT evaluation studies from some regulatory experience. With that, I will give it back to the chair.

DR. VENITZ: Thank you, Peter. Are there any questions for Peter? If not, let's proceed to the first presentation. Dr. Peter Bonate is going to tell us about clinical trial simulation and QTc. Peter?

Use of Clinical Trial Simulation (CTS) for PK/PD QT Studies

DR. BONATE: I would like to thank you for inviting me to speak. I am very honored; a little intimidated.

I am going to talk a little bit today about using simulation to address QT issues. I first got involved in this a couple of years ago, right at the time when Seldane--you know, the QT issues about it were starting to come to light. So, I have been doing this now for a couple of years. I have had the opportunity, some might say

misfortune, to work on about half a dozen of these compounds now, doing these analyses. They are very stressful. They are not like a regular exposure-response analysis. I think the stakes are a little big greater. The pressure on the kineticist are a little bit more because for a drug that has warts, this could kill it. So, it is a pretty stressful analysis.

[Slide]

What I am going to talk about today are some of my experiences with modeling and simulation of this type of data; how we have used simulation to address and interpret some of the results from these analyses.

[Slide]

Just to make sure everybody is on the same page, I am going to briefly address some of the issues regarding QTc so that we all have the same background, and I am going to talk about some placebo analyses that I did because in order to do clinical trial simulation you have to understand what the placebo response is before you can adequately model what your drug effect response is going to be. In doing the placebo analysis, some interesting results came to light and so I will

talk a little bit about the pitfalls that might come from just naively modeling QTc data. Again, I am going to focus on using Monte Carlo simulation to help interpret our results.

[Slide]

There is a variety of different metrics to analyze this type of data. The guidance talks about different varieties of them. One is looking at mean QTc interval. This is probably the least sensitive metric because it basically dilutes the drug effect from ECGs that have no drug effect.

Another one is maximal QTc interval. This one is relatively insensitive too because there is a lot of variability whenever you start talking about maximums.

Another one is area under the QTc interval-time profile. This one is starting to gain more--

DR. SHEINER: Excuse me, Peter--

DR. BONATE: Yes?

DR. SHEINER: Could you just say a word about the design? This is the mean of intervals, for example, across time beat-to-beat or is this moment-to-moment? Because not everybody here is exactly clear on what the design is.

DR. BONATE: Well, let's say you collect ECGs at zero, 0.5, 1, 2, 3, 4, 6, 8 hours after dosing, the mean QTc interval is just the mean of all those measurements. I didn't want to talk about how do you actually measure QTc. That is more of a cardiology issue. But when I talk about mean QTc, it is just the mean across different time intervals. I am going to assume at this point that the QTc interval data that you have has been over-read by a cardiologist and that it is a real number.

Another one that is just starting to appear, although it has been recommended for a number of years, is area under the curve. The problem with this approach is that the units are difficult to interpret. You get numbers like 10,000 millisecond times hour and nobody knows what that means. So, it is difficult to interpret.

Then you have maximal change from baseline. When you are talking about baselines you are controlling a little bit for within-subject variability. These tend to be more sensitive metrics.

Another one related to that is maximal QTc with baseline as a covariate. This is an ENCOVA

approach. They tend to be more powerful than just simple ANOVA approaches which are what the other approaches use.

Lastly, there is area under the QTc interval with baselines as a covariate. When I did some simulations a few years ago this was probably the most sensitive metric at detecting QT effects. But, again, you are confounded with difficult to interpret units and such. But these are basically the metrics that we have available to us and pretty much change from baseline and maximal QTc are the ones that people focus on.

[Slide]

I am sure everybody knows these, but the guidelines for what is "prolonged" are 450 msec in males; 470 msec in females, or 60 msec change from baseline. Then there is an absolute QTc greater than 500 msec. These are all considered clinically significant QTc values.

When looking at mean change from baseline, there really are no agreed upon guidelines for what constitutes prolonged. Generally we took 5-7 msec as prolonged because, using terfenadine as the yardstick at the doses that were given clinically, that tended to produce a 6 msec increase in QTc and

since that was pulled from the market for QT problems that is our yardstick that we have used. Hence, we now have the 5 msec change in QT as being a yardstick for what is prolonged. And, there are no guidelines on the AUC-based metrics at this point for what is significant.

[Slide]

I have found that companies tend to go through three stages when they are dealing with QT problems. One is--remember the guy from Mad magazine where he says, "what? Me worry?" There is the what QTc effect? It is the head in the sand approach--we don't have a QT problem; we are not going to worry about it. That is a dangerous attitude to have.

Then there is the, "okay, yeah, we've got a QT problem but we're not any worse than any other drugs on the market so we're going to take this approach and since they're approved, we're going to get approved." Then there is the, "yeah, we've got a QTc effect. We're going to characterize it and, hopefully, we'll be okay at the end of the day."

I think more companies are coming around to this third approach of we are going to characterize it and we are going to understand what

are the intrinsic and extrinsic variables that affect it so that we can make some rational decisions for whether this drug is safe or not.

[Slide]

So, I would like to move back to a study we did actually back in 1998 and 1999. Seldane has just got pulled off the market. We just had Allegra approved. At the time we were extremely sensitive to QT issues and so we had a new drug that was in development and we were concerned about QT issues, obviously. We felt that because we were Hoechst Marion Rousel, we would be looked at for QT problems a little more closely than maybe other companies at the time.

So, we went and we did what was probably a cutting-edge study at the time; it seems fairly straightforward now. We wanted to characterize the QTc response relationship for our drug. This was a single-center, randomized, double-blind, placebo-controlled, 4-way crossover where we took 20 males and we took 20 females, with standard phase 1 exclusion criteria.

[Slide]

We gave them three doses, 20 mg, 30 mg and 60 mg once a day for seven days, the fourth arm

being a placebo arm. Within each period we also had a placebo day on day minus-one. There was a week washout between periods. And, we gave meals one hour post-dose in the morning, lunch, dinner and snack. Interestingly, at the time we felt that our case report forms were getting too big so we were looking for ways to cut down on how we could make them a little bit smaller and one of the things we thought at the time was let's get rid of the mealtimes. We don't really need that. You know, it is a phase 1 study. The food effect for QT wasn't known at the time so in hindsight we kind of wish we had kept that data. It would have made interpreting some of the food effects a little better. All ECGs were taken prior to meals if they were scheduled at the same time. So, in hindsight, this seems like a pretty straightforward design but it was probably one of the first of its kind.

The results of this analysis were published last year in a book by Kimko and Duffull and I am going to talk just very briefly about it.

[Slide]

We did ECG analyses on 0, 1.5, 3, 5, 9, 12 and 24 hours on day 1, day minus-1 and day 8. So, we did it after the first dose of active drug and

then at steady state, and also on the placebo lead-in day. We also did it at trough on days 4, 5, 6 and 7. All the ECGs were over-read by cardiologists blinded to treatment, dose and period. They calculated Bazett's QTc for each chest lead and the largest one was taken as the QTc at that time interval.

[Slide]

We had a number of issues arising from this data set. First of all, what is the baseline? Is it the pre-dose at time zero on the day of dosing? At the time, much of what I am going to be talking about we really didn't know at the time. For instance, the circadian rhythm, we didn't really know that that was really such a big issue. I am not really sure that it is a circadian rhythm; I think it is more food effect that gives it a circadian nature. We also took only one ECG at each time point. I wish, you know in hindsight, we had collected multiple ECGs to lower inter-subject variability.

We could have used the mean of the placebo date, day minus-one. It is more robust. It is going to be based on many measurements. But it too fails to correct for any circadian food effects

that happen on the day of dosing. If were to take this forward into phase 3, you know, such a design couldn't be useful for phase 2 or phase 3. Lastly, there is point to point with placebo administration. For instance, we could take the 1.5 hour on day 1 with the 1.5 hour on day minus-1 and that would be the baseline. But then the question becomes, well, should the baseline be day minus-one or should the baseline be the placebo period?

So, there are a lot of different ways to analyze this data. The proposed guidance talks a lot about these things and I think one of the things that it could do a little bit better is to more fully delineate what should be the preferred baseline when doing these analyses.

[Slide]

We decided to build a placebo model because you need the placebo model to really understand what is going on with drug. We had a number of covariates available. We had period, day and time. We had chest lead; time of the last meal. We didn't know exactly what the last meal was but we could guess probably within five or ten minutes what it was. The sex; the race; what was

their baseline calcium and potassium at the beginning of each period; body surface area; and stress. When I say stress, the way they do these studies is that on days one, seven and day eight there are a lot of ECGs being taken so it is a pretty hectic day around the clinic. Everybody is running around so stress tends to be a little bit higher. So, we thought that might be an interesting covariate to look at.

[Slide]

We did the modeling using NONMEM. I will show you a little bit later why I used NONMEM instead of mixed, but all models were developed using LRT, standard model building techniques. The factors were entered into the model linearly and random effects were treated as normally distributed, which seems reasonable for QT data.

[Slide]

Just for the placebo period we had 769 ECGs from 40 subjects. That was a 449 msec

variance. So there was 5 percent variability across all the ECGs that were collected.

Interestingly, the placebo data showed a trend over time, over day of administration and the QTc intervals tended to go up from day minus-one to

day eight. The way I interpret that is that these phase 1 studies--we call them healthy normal volunteers but they are not exactly healthy normal volunteers; they are marginally healthy normal volunteers. Some of these guys go out bringing a couple of days before they enter the clinic. They get sobered up and they come in and they dry out enough to pass the screens and then they are in the clinic. What they are doing is while they are in the clinic they are getting healthy. They are getting three square meals a day. They are showering. You know, they are starting to get healthy. So, that is kind of how I interpret this trend effect over time. You know, they are getting better is what is going on.

We also found that chest lead was important. Lead IV tended to be about a 9 msec greater than other chest leads. Now, if you look at other papers in the literature, chest lead II tends to pop out more often but chest lead is an important covariate that needs to be controlled for.

This was probably the first time where we actually quantified the food effect. We found that breakfasts increased QTc and that lunch increased

QTc and dinner increased QTc, and each one of these increased them a little bit more. You know, each one of these meals tends to be a little more fatty than the one before it and fat tends to prolong the QTc interval, which raises an interesting question. Because of the food effect, it is going to make analyzing QTc data a little more problematic and I will show you that in a minute.

There was a stress effect. On the days that there were a lot of ECGs being taken the QTc intervals tended to be a little bit higher, and females were greater than males. You know, I did this about four years ago and now it seems really straightforward but back then this was cool stuff.

[Slide]

You don't have to worry about it but if anyone is interested, here are the quantifiable numbers for the model. The reason that NONMEM was used to do this analysis is that to model the food effect what I did was I just assumed that the QT effect declines exponentially since the last meal. I could have done this using a linear model and treated meal as just a fixed effect but, because I included the exponential term in there, I had to use a nonlinear mixed effect model. In doing so, I

probably could have increased the time it took to do this by about 100-fold.

[Slide]

Here is a fit for what the day 1 data looked like. If you look at where breakfast, lunch and dinner is you can see that after every meal QT intervals tend to be a little bit higher than the interval before it. The spike out at 16 hours were there is no time point, that is where they got their snack just before bedtime.

[Slide]

Here are the results over eight days of treatment. I won't show you all the goodness of fit plots but the results fit pretty well so we were pretty confident in the model that we had.

[Slide]

It raised some interesting observations. One was that there was a relatively large variability and when you broke it down to within-subject and between-subject variability we found that within-subject variability was more than between-subject variability, which is not something you see every day. Within-subject variability was about four percent but between-subject variability was only about three percent. So, it is kind of an

unusual finding.

Keep in mind that within-subject variability also includes measurement error and model misspecification. So, that may be the reason why we have such large within-subject variability and had we done replicate ECGs at each time point, we could have been able to separate the variance components maybe into a measurement error and into something else. At the time I was trying to convince people to include dummy ECGs to the cardiologist so that we could get a better ideal for what his reliability was but that was a can of worms that nobody wanted to open. Every time I proposed that, that is a very difficult sell.

Interestingly, when inter-occasion variability was added to the model, it accounted for very little of the variability, less than 10 msec2 so it was not included in the model. I have seen other papers where they have looked at this and they have pretty much come to the same conclusion, that if you look at individual corrected QT intervals over different days that tends to remain fairly constant across days, which is kind of surprising.

[Slide]

I am just going to take a step aside and do my sell for the AUC corrected QTc. I think more effort should be spend in identifying this as a variable measurer instead of change from baseline or maximal QTc. AUC is an integrated measurement over the drug effect and it tends to be more sensitive than any of the other metrics that we are looking at. When you look at maximal change from baseline you are only looking at one time point and you are ignoring all your other observations, which is a loss of information. So, when you look at AUC, it tends to be more sensitive. As I said before, if you use just raw AUC the numbers are like 10,000 so it is difficult to interpret.

But if you divide by the interval in which the AUC was measured, now you get a weighted average QTc which is interpretable with the weights proportional to the time difference between measurements and the numbers are right in accord with what you would expect. So, when I did the placebo model for the AUC many of the covariates that were important before no longer become important.

Here is my methodology In this case I just did linear mixed effect models. You can see

my covariates. But in this case none of the covariates were statistically significant. The day effect was gone. So, it is something that we need to consider. More people need to do research on this so that we can get a better feel for how it performs as a metric.

This time the between-subject variability is greater than the within-subject variability, which is what you would like to see. Interestingly, the sex effect that you normally see with QTc was not observed with the AUC metric. I don't know whether this was a power issue or what.

[Slide]

Now that you have a model--you know, just having a model isn't of any value unless you do something with it and that is where simulation plays a role because simulation is really just applied modeling. It is a tool that can help you understand the behavior of your system. It can help you assist in discovery and formulating new hypotheses; where you need to go next. Of course, it can be used for prediction. That is probably what it is most often used for. Sometimes you can use it to substitute for humans, like with expert systems. You can use it for training and, of

course, you can use it for entertainment, not just for the modelers but for the people that use it.

[Slide]

If you want to simulate QTc trials, what is it that you need to know? Well, you need to define your metrics. What is going to be your primary metric? What is your goal at the end and what is the metric that you are going to use? Once you know your metric you need to know the variability of that metric, both within a patient, across patients, measurement error, that kind of thing, and how it is distributed. Is it normal distribution? Is it log normally distributed? QTc intervals tend to be normally distributed. I have yet to see a log normal QTc distribution. If you have an estimate of variability, does that estimate of variability pertain to the population that you are interested in studying?

What I showed you was done in healthy normal volunteers. The question then becomes are those variance components applicable to the population of interest? Probably not because patients tend to be more heterogeneous than healthy normal volunteers. So, the question then becomes, well, how useful are the results of your simulation

if your variance components might not be valid?

Of course, you need a PK/PD model. You need to know what the variability is in those estimates. Then, what is the experimental design? How are you going to actually dose the drug?

[Slide]

One of the things that came out of the placebo analysis, as I said, was the food effect. Well, surprisingly, if you just do a QTc analysis you can get food effects that mask drug effects, that act like drug effects. Think about this, on days when we were doing intensive sampling we had patients fast for 14 hours. Then they get their meals and then they go on to the next day. Well, QT is prolonged after a meal. So, right away we are increasing QTc from baseline, regardless of whether the drug has any effect or not, simply because of the timing at which the samples were taken.

So, I did an experiment. I simulated 100 subjects after oral administration of the drug--the same time points as in the last study. Concentration and QTc were totally independent. There was no drug effect in the simulation. Then I analyzed the data using pop mixed and used a random

effects model. I treated concentration as a covariate in the model.

[Slide]

Here is the simulated QTc data. There is nothing unusual about it. It looks exactly like what you would expect when you look at population QTc data.

[Slide]

Here is the PK data. It is actually pretty tight. There is nothing big there.

[Slide]

Then, when you look at the concentration QTc effect relationship, it doesn't look like much but it is statistically significant. The p was less than 0.0001. What it said was when you look at the solution to those fixed effects is that for every 100 ng/ml increase in concentration QTc is going to go up 2.2 msec. If you look at where Cmax is on the previous curve, 400 ng/ml, QTc in this study is going to go up 8 msec. That is not a drug effect. That is a total artifact. So, you have to be careful.

So, I said, okay, what if I control for baseline? As my baseline I am going to use my pre-dose sample. This is a real common way of

analyzing retrospective phase 1 QTc data because these studies are often done where the patients come into the clinic; they get their ECG; and then they are dosed with the drug and then they get an ECG maybe at Cmax and then again off-study. The question then becomes, you know, is there a QTc effect? Well, the only baseline you got is the one at time zero. So, when you do that you get the same results. I mean, you are just subtracting out a constant. You get exactly the same effect.

So, this is the pitfall of using a time zero baseline and doing your QT analysis. You can get a total artifact and be totally fooled by it. The only way to avoid this is to do a point-by-point baseline correction.

[Slide]

Here is another simulation that I did. It is a very simple one. What is the false-positive rate of these metrics that we are using, that the EMEA put forth in their guideline? This was done a couple of years ago as well.

A percent of subjects will have a QT more than 470 msec in females. This is after placebo administration. What percent will have a change from baseline of 30 msec to 60 msec of greater than

60 msec?

So, I sampled 5,000 subjects and I serially sampled the ECG values and calculated the percentages for each of these. What it shows is that these metrics do have a false-positive rate. For instance, for a 450 msec change in males the baseline false error rate is 1.5 percent. So, under these metrics you are going to have a QT effect in your analyses. The question is, is it real and is it important?

So, by using simulation in your study you can help interpret the results from your analysis so you can show, well, if concentration is independent from QT, then this would be my false-error rate. This is what we showed with the drug. So, now we can interpret the relevance of these percentages.

[Slide]

This goes back to a different drug. We did a pop PK analysis on it. We did a QTc analysis of it. We saw that there was a QT effect with this drug. We were convinced it was real. We found out that body surface area was an important covariate. The idea was that we would do the PK/PD analysis for identifying the important covariates and then

use simulation to determine the impact of those covariates on the QT and with or not we needed to do any studies in special populations, like maybe obese versus anoretic patients.

It turned out that once we did the pop PK analysis we only found one covariate, which was BSA. It was on intercompartmental clearance which, if you think about it, is probably not going to lead to anything but we continued the exercise anyway and I will just go through the motions for you because it is an informative exercise.

[Slide]

The question was is BSA and important covariate? This was our change from baseline model. We showed that there was a 2.94 msec increase for every 10 ng/ml with the drug. This kind of plot--and I show it to clinicians who are unfamiliar with population data or with ECG data, they look at this and they go, how in the world? I mean, this is all over the place. You can't fit a model to this. So, you had better have a good answer for that question when it becomes time.

[Slide]

What I did, I simulated the placebo lead-in day and then concentration-time profile for

150 subjects at steady state. We took the worst-case scenario. We dosed from 10 mg to 60 mg once daily and we varied the body surface area from 1.2 m 2 to 2.2 m2. We simulated the placebo data and then we added on the drug effect. From that we calculated the standard metrics for assessing QT prolongation and we computed the means by dose and weight, and we fitted a response surface to this. Now, there was more to this analysis. We looked at the percent of subjects having values more than 45, etc., etc. but I will just show you the mean profiles.

[Slide]

When we got through at the end of the day, we saw that there was a linear relationship with dose. That is the axis, over towards the right. But BSA, as you might expect, had no effect on QT interval so we felt there was no need to do any further studies with weight as a special population. We saw that the 5 msec point was at the 60 mg dose. Clinically, we were planning on going to phase 3 studies with 10 mg and 20 mg. So, we felt we were at a pretty good place on the concentration-effect curve.

[Slide]

Here are the males. It is the same thing, just a little shifted. So, at this point we felt that there was no further need to do any special population studies with weight as a covariate.

[Slide]

The last application I want to show you is using simulation to test the power of a phase 2 study where now you are given a study design and you want to know what is the probability of detecting a true QTc effect-response relationship in that population.

This is what the project manager gave me. He said, look, we are going to do 10 mg, 20 mg, and 40 mg in a three-arm study. They are going to get dosed every day for 8 weeks. I want to collect ECGs on screening, week 4, week 8, at zero and 8 hours post-dose. We will collect 4 hours post-dose because we know that is around where Tmax is. We are not sure of the sample size; we are flexible on that. You can help us on that, but 30 to 120, that is kind of what we are leaning towards.

So, a varied the sample in 30 to 120 by 10, and I just analyzed the results using mixed effect models, using sex, day, time within day, concentration at baseline as the fixed effects and

intercept and concentration as random effects between subjects. I repeated the simulation 250 times.

There are two ways you can analyze this data. You can treat concentration as a continuous random variable. you can treat dose as a continuous random variable or you can treat dose as a categorical variable. I think in the last meeting that we had here there was a discussion on categorizing continuous variables and its effect on power.

[Slide]

Here is an example of what could happen. The solid circle is when concentration is used in the model. The squares are when dose is either continuous or dose is categorical. You can see that when you categorize dose the power becomes a little bit smaller, but by far the most powerful metric was concentration. But even with 120 subjects we only had a 60 percent chance of detecting a true QTc effect. So, I told them if you really want to power the study to find something, you are going to have to go back and either increase the sample size or come up with a better design.

[Slide]

But there are a lot of unresolved issues in this. There are a number of issues that the guidance does not address and I just want to raise those. One is the choice of the covariance matrix. A lot of studies have shown, particularly in the linear mixed effect model literature, that the choice of the covariance matrix can have a profound effect on whether you detect fixed effects. So, how you go about choosing that covariance matrix, which one to use, has not been addressed yet. Should it be simple? Should you treat the intercept and concentration as independent? Should you allow them to be unstructured? You know, how should you do this?

And, what about within-subject variability? These observations are probably correlated. Every analysis that I have seen so far has treated the within-subject variability as independent, which is probably incorrect.

[Slide]

When I did the lagged residuals on an analysis from a couple of years ago, this plot is a lag 1 correlation plot. So, this is the residual against the observation next to it. Here is lag 2

which is the correlation between two observations later. You can see that the correlation tends to dissipate as time goes on. So, treating within-subject variability as a simple covariance matrix is probably not entirely appropriate. It may be an AR1 or Toeplitz is probably more appropriate for this kind of data.

[Slide]

The other issue is whether we should use maximum likelihood or REML estimation. This applies if you are going to use a linear mixed effect approach. You have two options, particularly within SAS, REML being the default. But in order to these simulations you need to know what the variance components are, and whether you use maximum likelihood or REML you are going to get different variance components.

I think it was shown about 20 years ago that the within-subject variability is more than between-subject variability but you probably want to use maximum likelihood, whereas most people would probably just use REML and be done with it. So, you know, which estimation method is best hasn't really been examined.

The other is what is the best model

selection criteria? Everybody uses likelihood ratio test, particularly when using NONMEM, but when you use SAS you get AIC, you get BIC, corrected AIC, and which of these metrics is most relevant to model selection I don't know.

[Slide]

In summary, I think there are a couple of points I want to point out. One is that using a time zero baseline just pre-dose is probably the worst baseline you can use. It leads to a lot of artifacts in the data, the food effect in particular, and you just want to avoid it as much as possible.

Whatever metric you are going to use, there is going to be a false-positive error rate and the question is what can we live with. You know, if placebo data has a three percent false-positive rate, is it five percent that you should be concerned with? Is it six percent? You know, if you get ten percent of your subject meeting the criteria? When it is important and what are we willing to live with?

Simulation can be a powerful tool to help answer some of these questions, not only with the agency but internally it can help you make

decisions on where to proceed next.

[Slide]

Lastly, this is my opinion and I am probably going to take a little bit of heat for this but I think we are spending a lot of time on QT and I am not quite sure exactly, totally why. I mean, QT is really no different than any other laboratory parameter. We need to decide how to measure it. We need to decide what if important, what is clinically significant. I have a theory. This is my snowball theory. We started to get a little sensitized to QT because of a couple of drugs that might have shown it. Not everybody that has a prolonged QT develops Torsade. We need to more fully understand what are the issues relating QT to Torsade and sudden death before we start throwing the baby out with the bath water. If the NIH needs to get involved, so be it. Let's have a prospective study to really examine is this an issue because all of these analyses are retrospective and whenever you do a retrospective analysis you have the benefit of hindsight. So, we may be missing something here. We may be making a lot out of nothing.

I think that a couple of years ago when

this first started being an issue a couple of conferences were held and maybe a QT topic was held within those things. Then somebody else said we need to have a whole meeting on QTc and the next thing you know, we are at the FDA. Let's put some perspective on QT and let's do this right. Let's not just say that a drug that has prolonged QT is the death knell for the drug. Let's be reasonable about it. Let's understand what is the science behind this and how it relates to patient safety.

I want to thank you for letting me speak here today. I would like to thank Tania Russell and Quintiles and Danny Howard at Adventis for helping me bounce some of these ideas around. Thank you.

DR. VENITZ: Thank you, Peter. Any questions for Dr. Bonate?

DR. SHEINER: I will start with questions and do comments in another round. I had a question but I think you answered it, which is that this artifact that you think will happen is with the meal so if you did, in fact, prevent people from eating then maybe the zero time baseline correction might be okay. Is that what you were saying?

DR. BONATE: You know, I think a more

appropriate study design would be one where patients get low fat meals at every meal and maybe just small meals throughout the day. I don't think you can reasonably prevent them from eating throughout the day.

DR. SHEINER: No, but it is the confounding of the time effect which you believe is due to a meal--

DR. BONATE: Correct.

DR. SHEINER: --with the drug effect that is the problem. So, however you might get rid of that time effect, whether it is changing the type of meal, not getting a meal or whatever, that was the issue, that confounding.

DR. BONATE: Yes.

DR. SHEINER: Because you didn't have the placebo, so to speak, curve over time to compare to.

DR. BONATE: Yes.

DR. SHEINER: That is the usual design. The other question I had was I didn't understand what your point was about the false positives. You said 1.5. Was it that 1.5 percent of males, for example, would show a QT prolongation greater--

DR. BONATE: Yes.

DR. SHEINER: Okay, but that doesn't mean your study would show a QT effect.

DR. BONATE: No.

DR. SHEINER: No.

DR. BONATE: That is just the placebo baseline.

DR. SHEINER: Yes, but that is individuals. What you are saying is that you have a threshold that says it is abnormal to be above the following thing. Typically in laboratory tests when there is no biology to tell you, you take five percent. So, actually, that is pretty good, 1.5 percent--

DR. BONATE: Yes.

DR. SHEINER: --false positives is actually a pretty specific laboratory test.

DR. BONATE: Yes, but in some of the metrics, like the 30 msec to 60 msec, the number was 50 percent.

DR. SHEINER: Oh, I agree. That is very non-specific. I just didn't understand. You weren't talking about studies at that point.

DR. BONATE: No, I was not.

DR. DERENDORF: The QT intervals are a classic biomarker. We are not interested in them

as such but we are interested in them to maybe make them surrogates for other events, as you mentioned. You said that right now the cut-off is sort of a 5 msec change where people get worried. If I look at the effect that you get from your dinner, that is 10 msec. So, there is something that I don't understand. If that biomarker is effective for something as trivial as a dinner, then that is not a biomarker.

DR. BONATE: Well, the 5 msec is based on a mean. So, it is based on the average across all the observations within the day. It is completely taking out the time course of it. When you talk about the food effect at dinner, that is a particular point in time. So, they are kind of apples and oranges comparisons.

DR. DERENDORF: The question that comes up then is what is the mechanism of these changes? What does the food do that causes the prolongation and what does the drug do? Are they the same mechanism? Are they additive or are they two completely different events that are manifested in the same change?

DR. BONATE: I imagine that would be drug dependent. I mean, not all drugs prolong QT by the

same mechanism and why food does I don't know.

DR. DERENDORF Coming back to the original goal of this whole thing, it is that we want to measure something that tells us something, something else that we are really interested in. That should be as specific as possible and that doesn't seem to be the case.

DR. BONATE: No, I don't think it is.

DR. VENITZ: Peter?

DR. LEE: I was just wondering how conclusive we can be regarding the food effect. Would it be just some sort of variation during the day that just happened to coincide with the food? Would a study comparing different foods on QT be more conclusive, say, giving low fat food compared to high fat food? If, indeed, there is a food effect, would including a placebo arm in the study take care of the food effect, which means that if you see a food effect in the placebo arm you can subtract that from your drug effect?

DR. BONATE: Going to your first question about quantifying the food effect--I know I skipped through the slide very quickly, but I did quantify the food effect in this analysis and for breakfast it was 10.6 msec; lunch, 12.5 msec; and dinner was

14.7 msec. I don't know if it is a volume effect or if it is a fat effect.

DR. FLOCKHART: But is that an average of an area or single time point? What is that number?

DR. BONATE: It is a fixed effect. It is more of a shift from the baseline. So, the baseline is 389. So, if you had breakfast it would be 399. Do you see what I am saying?

DR. FLOCKHART: Yes.

DR. BONATE: If you think of it like an analysis of variance, that is kind of what it is. So, if you included the placebo--I think if you did the point-to-point correction you would control the food effect, provided the same meal was given on both days.

DR. VENITZ: Let me give you a possible mechanism for the food effect.

DR. BONATE: Sure, please.

DR. VENITZ: Did you look at your heart rates at all? Because you are looking at Bazett-corrected QT intervals.

DR. BONATE: Oh yes, I didn't even want to go there. Right.

DR. VENITZ: But my point is you might well look at secondary effects to the heart rate

because every time you eat your heart rate will go up, as most of us who have just had lunch can experience. So, it might be an artifact in your correction. It may well be that you have sympathetic activation that somehow affects repolarization as well. So, I think it is not unexplainable that you see food effects on something as esoteric as the QTc interval.

DR. BONATE: No, you are absolutely right. I left this on my slide but I wasn't going to talk about it, but I will now, and I want to say our "Slavic" devotion to Bazett's--I mean, why can't we dump this dog and go to something that is a little less sensitive to heart rate? I have heard this argument that with Bazett's we have historical data to compare it to. Well, if your historical data is wrong what is the point of making the comparison? Let's just say in the guidance no Bazett's. Why can't we say that? I don't know. Let's go to Fridericia's or something.

DR. SHEINER: Fridericia's doesn't work any better either.

DR. BONATE: Well, it is better than Bazett's.

DR. SHEINER: Maybe, but not much. It is

an interesting point. First of all, I have to correct your English there. There is nothing about the Slavs that--

[Laughter]

--it is "slavish." You know, I think it is interesting. It is an artifact that I think is very similar to sketcher plots and stuff like that. There was a time when you could only make a scattergram so if you had two factors that were affecting what you were interested in, heart rate and, let's say, drug or something else, you had to get rid of one of them. So, what you did was divide it by its square root, cube root or whatever it is, and then it just sort of persists like body surface area, and we know that formula is not the formula for body surface area. In 1919 it was--well, I won't go off on that.

In any event, what you want to do is heart rate as a covariate. You may find that you can find some kind of parametric formula and you may find that you can't. It doesn't much matter, but you can correct for it and I think that some of this sort of stuff, you know, may go away. So, I think the general principle is we have measurements, like interval, ECG and heart rate,

and keep them separate because now we don't have the problem that we can only look at one variable at a time.

DR. BONATE: Well, I think an ideal situation--I mean, I think there is a lot of value to individual corrections, which I think is where you are going with that. The problem with that is that you need a lot of data for an individual to be able to make that correction. If you have one ECG on a person it is difficult to say what is the correction that you use for that subject.

DR. SHEINER: I am not saying that. I am saying we could analyze lots of data and find what the heart rate correction in general was. It might not be any particular simple formula that allows us to then take that "corrected" thing and plot it against something else. It might be more complicated. The point is we have plenty of data.

DR. BONATE: Yes.

DR. HUANG: A quick question. You mentioned that the area under the QT time curve has potential but is not really investigated. I wonder, with the several applications that you listed, have you tried to use that? For example, in the food effect you said if you do a

point-by-point in the placebo phase you might be able to correct it if they are taking the same food, but we know that is probably not reality. So, if it is the other measure would it provide a method to decrease the sensitivity of this circadian or food effect? You have shown that using AUC a lot of other measures become insensitive--the differences that you would ordinarily see that you don't see anymore.

DR. BONATE: Well, I think it depends on what your baseline is. If you use a time zero baseline the AUC metric will exacerbate the food effect.

DR. HUANG: I am talking about if you do have a placebo. The concept paper recommends using a placebo.

DR. BONATE: Yes, if you have a time-time, then AUC I think would still be more sensitive and you wouldn't have to worry about the food effect.

DR. HUANG: More sensitive or less sensitive?

DR. BONATE: It should be more sensitive. I think you have to have the point-to-point correction to really do this.

DR. HUANG: That is what is recommended.

DR. BONATE: Yes.

DR. HUANG: By the way, I think Bazett's being mentioned partly because a lot of devices right now are calibrated with Bazett's.

DR. BONATE: You know, in 1920 they could probably only do the square root on a slide rule. I don't know; that is all I was thinking.

DR. VENITZ: Wolfgang?

DR. SADEE: Just a comment on the food effect. If you test chemicals, drugs maybe ten percent have a chance of causing QT prolongation. With a meal you take in about 10,000 compounds. So, I think it is a chemical effect.

DR. BONATE: Maybe.

DR. VENITZ: Any further comments or questions?

[No response]

Thank you, Peter.

DR. BONATE: Thank you.

DR. SHEINER; Let me just say one thing. It is a biomarker and the problem is that it is probably the heterogeneity of repolarization that is the problem in Torsade so the average goes up if it is a real food effect. My guess is it is also a heart rate effect. But if it were a real effect,

it might be that it is a general effect with, let's say, a vagal effect and sympathetic effect and it is going to happen everywhere. It is not increasing the heterogeneity. Unfortunately, we haven't got a measure of the heterogeneity or repolarization so we take the average as a poor measure of it. So, for drug it is one thing; for food it is another thing. That is entirely reasonable, you know, to have two different causes of the same biomarker and one of them you consider dangerous and one you don't.

DR. DERENDORF: Oh, I completely agree. It just becomes a design issue. I fully agree with your approach that the point-to-point comparison would be the way to go. But looking at your curve here, you need a lot of data points to get that sensitivity to detect the difference there. That is going to be the issue.

DR. BONATE: Especially if you were comparing, say, day 8 because then you would need a day 8 point-to-point to really make a proper comparison. Yes.

DR. SADEE: I have one more quick comment. You mentioned 30-50 subjects or so. Their polymorphisms in the candidate genes are associated

possibly causatively, in a causative way, that have a frequency maybe much less than that. Since the real danger is 1/1,000 it is not quite clear to me whether 30 or 50 subjects would do. So, if you have polymorphism as one percent that sensitizes a particular individual to a particular chemical, you will not detect it.

DR. BONATE: You are talking about the link between the biomarker and the outcome. I think, you know, 30-50 subjects is more than adequate to determine the change in biomarker. Making the next step, you are absolutely right.

DR. VENITZ: Thank you again. Our next speaker is Dr. Leslie Kenna. She is going to give us the second part of this case study on QTc.

Case Studies

DR. KENNA: It is a great privilege to be able to present to this committee. I have to say though that if Peter, with his years of experience felt intimidated, I am going to try not to act like a deer in headlights up here. This is a very wonderful opportunity.

[Slide]

My presentation has four parts. First, I will present the question of interest. Then, I

will present data from the trenches to illustrate some of the challenges we face. Next, I will present the clinical trial simulation methodology under consideration to address those issues. Finally, I will present some very preliminary results. As you listen keep in mind that this is a work in progress. We are assembling a QT database and developing tools to analyze those data. We are soliciting your advice today on an effective approach.

[Slide]

In the interest of safety, we would like to know the effect of drug on QT interval in the worst-case scenario. That is, to know what response might occur in the case of increased drug exposure due to, say, drug-drug interactions.

[Slide]

As Peter said, a major challenge is that there is tremendous variation in observed QT response, greater than the response of interest.

[Slide]

There is wide variability in measured QT interval in a given subject at a given time in a given day.

[Slide]

Just to give you a sense of that, this is a plot of Fridericia-corrected QT data collected in one subject on one particular day before any drug was dosed. So, that is baseline, before--you can't see that? At each point ten measures were taken at one-minute intervals. Just by looking at the data, you can see, for example, that at that nine-hour time point measures taken one minute apart had a range of 15 msec. Maybe you can't see it but this cloud of points is shifting over the course of a day.

[Slide]

So, not only is this response shifting over the course of a day but a given subject may have different QT response patterns at baseline, one observed on different days and now we actually have a black line connecting basically the average between the ten points on a given day in a subject. You can see that the lines don't overlap from one day to another.

[Slide]

We just looked at data from one subject but if you compare subjects you can see that different subjects have different QT response patterns over time.

[Slide]

This slide provides a side-by-side comparison of the QT measurements taken over four baseline days in two different subjects. We looked at subject I but now subject K's data exhibits the same overall characteristics but the pattern of change appears out of sync with subject I. You see all the points going down when the other subject's points are going up.

Given that we may want to detect a change in QT interval of about 5-10 msec, if there can be about a 15 msec change in response over measurements taken one minute apart before any drug is even given, in some ways we are trying to find a needle in a haystack. That response is not impossible to find but it becomes very important to design QT evaluation studies effectively.

[Slide]

For this reason, we set out to review the study designs used in several recent submissions. A review of several recent submissions to the FDA revealed that different study designs have been used, for example, in terms of the duration time.

[Slide]

To illustrate this point consider the

definition of baseline in six recent submissions. Here you see that baseline was defined as anything from a single measure taken 14 days before the start of a QT evaluation study to over 100 EKGs taken during two pre-dosing days.

[Slide]

Another observation is that in different studies a different response has been observed to the same drug at the same dose. 400 mg of moxifloxacin is recommended to be tested in subjects to evaluate whether a trial is sensitive enough to detect a change in QT interval. The moxifloxacin label says that it cases a 6 msec increase in QT interval at that dose. In one study we reviewed, however, 400 mg of moxifloxacin was associated with an 8 msec change in Fridericia corrected QT interval. In another it was associated with a 13 msec change.

[Slide]

Just to show you some key features of those two studies, you can see from these confidence intervals that case one yielded a much more precise estimate of drug effect than case two. There were some subtle differences in terms of the number of baseline measures and the number of

replicate EKGs.

So given that study design is something we can control if it becomes important to identify how much of this difference between effects estimated depends on the study design, especially if you consider or if you imagine that moxifloxacin was actually your drug of interest because, depending on the indication and effect of 8 msec, might have been considered clinically insignificant while an effect of 13 msec might have raised concern.

[Slide]

Just getting back to observed trends, we have also been presented with incidences where the observed response was sensitive to the data analysis method.

[Slide]

For example, consider the following difference with regard to mean versus outlier analysis, drug X was associated with a 4 msec increase in Fridericia corrected QT interval at Tmax. The positive control in that study was associated with a 9 msec change. This suggested that the drug had less of a QT liability than the positive control.

[Slide]

The outlier analysis, however, suggested that the drug and positive control yielded a similar effect on QT interval and that this effect was greater than that on placebo. So, this raised the question of what data analysis method we should trust.

[Slide]

Then consider the following example of how the estimated risk depended on the definition of baseline. In one analysis of a particular data set baseline was defined as measures taken during a treatment-free period plus measures taken on placebo.

[Slide]

In that case a five-fold increase in exposure was associated with a two-fold increase in the number of outlying QT measurements. The appearance of a shallow dose-response relationship suggested that increased drug exposure would have little effect on QT interval or that the drug was relatively safe.

[Slide]

However, when the same data set was analyzed having baseline defined as measures taken during the treatment-free period only, it appeared

that a five-fold increase in exposure was associated with a four-fold increase in the number of outliers. This suggested that the response was proportional to dose and could potentially increase with greater exposure.

[Slide]

Given these challenges, our goal is to learn from available data to aid in the prospective design of QT studies.

[Slide]

The specific aims are to assemble a QT database from data in submissions, then resample from those data and use clinical trial simulation to evaluate the clinical trial designs and data analysis methods.

[Slide]

I will now shift and give you an overview of our proposed approach and then go into greater detail illustrating each step.

[Slide]

To evaluate the success of a study design we need to know the true underlying effect of the drug. So, the first step is to simulate your data. The proposal is to use baseline QT data that we have, much like the data I presented earlier, so we

don't have to assume a shape of the distribution. We will choose a study design and models for the drug's PK and PD profile. We will then add baseline response to the simulated response to treatment.

In any real study one only gets to sample the QT responses according to the study design. The next step then is to sample from the true data according to the chosen study design. Then response will be estimated by the methods of analysis of interest. We can explore those proposed in the concept paper and those used in recent submissions. In order to get a sense of how a particular study design performs it has to be repeated many times. Finally, performance will be quantified after all the repetitions are carried out. One possible way to do this is by computing power.

[Slide]

Now just to show you our plan in greater detail, we start by randomly drawing baseline data for each subject in the trial from the database. In the data I showed earlier we had four baseline days of measurements. If we only need baseline observations from one day, then a particular day

will be selected at random from these data. Here you see ten observations for time as collected on a given day.

[Slide]

Next, depending on the study design under investigation, N measurements will be sampled at random at each time point in a given individual from the day of baseline measures selected. Here you can see that three measures were randomly selected at each time point from the original data set.

[Slide]

Given a study design where we evaluate two doses--two doses because one recommend in the concept paper is that you would use a therapeutic dose and a super-therapeutic dose that covers drug-drug interactions or whatever that worst-case scenario is for your drug--two doses of drug, and using both placebo and active controls we would like to investigate the impact of the following parameters, whether you have a crossover or parallel design; single dose versus steady state design; the number of subjects; timing number and duration of EKG measures; the PK/PD model for the drug, for example, whether maximal response occurs

at the time of maximum drug concentration or whether there is a delayed effect and, along those lines, one mechanism for effect delay that we can simulate is if the drug and the metabolite both affect QT interval. Then, the PK model for the drug would also be varied. For example, we could explore the effect of the clearance of the parent and, say, an active metabolite.

[Slide]

After we have randomly chosen a baseline profile for a subject before and while receiving drug and before and while receiving placebo--so here is baseline before drug; baseline before receiving placebo--we are going to add the baseline to the simulated true response to a given treatment. For drug the treatment effect over time might be as follows, QTc might increase with time and decrease just due to the fact that it is driven by drug concentration which is also rising and falling. Then, for placebo there might be a slight increase in QT that has no dependency on time.

[Slide]

Then one adds the sample baseline to the true underlying treatment effect to get treatment resistant pathogen observed in a subject. The

responses that are shown here are just what you get when you add each of the baseline points to the true drug or placebo effect at that time. Here, for placebo you see a trend that just simply reflects the baseline variability in QT.

[Slide]

In the previous slide I showed you how to simulate true underlying response, as shown here, but in clinical trials, as you know, you only get to observe the response according to the study design. From that true response, if one chooses to sample one QTc value at a given time, then you might see this response to drug and this response to placebo. Likewise, for baseline.

[Slide]

If you sample three QTc values, for instance, as baseline just before starting treatment, then your sample baseline might look something like this.

[Slide]

Then to estimate response we performed some operation on the collected data to evaluate the difference in response to the treatment after baseline effect is accounted for. That is just symbolized here as a minus sign. One example of an

approach that you might use to do this is, for example, you might take the mean sampled response on treatment minus the mean response on baseline. Some others are listed here and this is certainly not an exhaustive list.

[Slide]

These are not supposed to be question marks. They are supposed to be arrows. This process of randomly sampling baseline data, simulating response to treatment and then estimating response will be repeated many times because, due to all the sources of variability including baseline QT variability, although we have fixed the drug effect within a given simulation study, different trials will enroll different subjects causing the estimated effect to vary, as I just show here.

Since we set the drug effect parameters when we design the simulation study we know the true underlying response that we are trying to detect, so we can just compare the estimates across all those replications to compute performance.

[Slide]

One way to evaluate how study designs and data analysis methods perform is to compute power.

That is, given a particular study design, we can tally up what fraction of simulations allow you to detect the drug effect on QT interval when there really is such an effect.

[Slide]

I will now show you some very preliminary results of our investigations.

[Slide]

As I pointed out earlier, we need baseline data to conduct our simulation studies. The source of the baseline data presented here are 72-hour baseline profiles in 45 subjects. The simulation conditions were as follows, the trial was a randomized, parallel design with two arms, treatment and placebo. There was a 24-hour placebo run-in and 24 hours on treatment. QT sampling was hourly from 1-24 hours post-dose. We varied the number of subjects.

Treatments were administered orally at a dose of 100 mg. The drug exhibited one compartment PK. PK/PD was a linear effect added to the baseline variation, and there was no effect delay.

Analysis methods included taking the difference in maximum QTc on treatment and maximum QTc at baseline, taking the difference in the mean

QTc on treatment and mean QTc at baseline. These are things that may have either been seen in submissions or in the concept paper.

[Slide]

This slide illustrates how PK/PD data in 40 subjects looked for a trial under the parameters just presented. As you can see, we presumed that response was directly related to concentration so both of them peaked at the same time, and that maximum response was about 16 msec.

[Slide]

This slide shows the power of the data analysis methods to find that the drug caused a significant change in QT interval relative to placebo as a function of the number of subjects in the study. Each line represents a different way of analyzing the data. Power ranges from zero to 100 percent where 100 percent means the method correctly identified a significant difference every time it was used. Recall that the difference really was significant; it was about 16 msec.

[Slide]

As you would expect, all methods have more power as the number of subjects is increased. For a given study size you see that the methods of

analysis influence how often you can expect to correctly identify drug response. For example, when we subtracted the man QT value at baseline from the mean response after taking drug, which is the black square at the highest point on the plot, 85 percent of the time we were able to identify that the drug prolonged QT interval if 80 subjects were in that trial.

In that same trial if you, instead, subtracted the maximum QT value at baseline from the maximum QT value on drug, the correct response was instead identified 55 percent of the time. Keep in mind that the data didn't change, just the way they were analyzed.

[Slide]

So, we slightly altered the study design so that instead of collecting several measures at baseline only one sample was collected at baseline which, as Peter has already pointed out, is a horrible way to design your study.

We examined the result in the top panel on the previous slide where baseline included measures taken hourly over 24 hours. The bottom panel shows the results under the same conditions except that, as I said, one baseline measure was taken. You can

see that power is greatly reduced. If you estimate response by subtracting the single baseline value from the mean response on drug you only identify significant difference between drug and placebo seven percent of the time if the study has 75 subjects. You also see that the metrics actually flip around in terms of which was more powerful and now taking the maximum is a little more powerful than taking the mean.

[Slide]

As you can tell, this is definitely a work in progress and we would greatly appreciate the committee's feedback on the following questions. These questions could just guide the discussion but we are certainly eager to hear what you have to say. Thank you.

DR. VENITZ: Thank you, Leslie. Before we get into the specific questions, are there any comments or questions about Leslie's presentation?

DR. SHEINER: Leslie, did you sample the QTc in you baseline, your 72 hours? Was that the QTc or the QT?

DR. KENNA: That was the QTc.

DR. SHEINER: So, apropos of the last discussion, it might be interesting to sample both

the QT and the heart rate since they are both available, and then see, making this particular correction you are using, whether it is Bazett's, Fridericia's or whatever you are using, whether there is a better way to do it with respect to that as well. You have the potential to do it. You are investing a lot of effort and that would be a small addition that might have a payoff in showing what the price is of using this standard correction, which we all know isn't very good.

DR. FLOCKHART: What surprised me about Leslie's data was that one of the things that has been a kind of unquestioned assumption is that when we do circadian rhythm once in a person, that will be the same if we did it ten times, but it is not. I think that is a really important message in what you are saying.

I think the thing I am most worried about in this approaches, and this comes somewhat from history, if you like, the history of quinidine to terfenadine to, in our case, pimozide. The thing with quinidine was--we did this in the same study where we gave people intravenous quinidine--we wouldn't be allowed to do it now--to see if there was a gender between men and women, and if you had

analyzed that study using an averaging effect, if you had done a circadian rhythm before on one day and then you had done an averaging effect after, you would have missed a humongous change because we were sampling for two days. If you had actually done an average, the average would have diluted it. Point-to-point comparisons would have done the same thing, you would have missed this thing that lasted no longer than about an hour, even though you are giving a drug that prolongs the QT 30 msec, 40 msec, 50 msec, because of the very short time interval.

I actually don't know a drug--and I would be interested if there are other members of the committee who do--where you don't see this cardiac reaction to the prolongation of QT. In three of the drugs that I have studied, pimozide, haloperidol and ziprasidone, you see an actual reverse, a negative QT interval change. It is like the heart knows somehow that it is being prolonged and it protects itself in a kind of rebound way. Again, that can dilute the effect that you see. So, timing here is important because, again, if you are doing averages or you are doing point-to-point comparison with circadian rhythms you miss that

effect completely.

The other thing, you build it into your model but I think you did the absolute best thing to do, you built in a model where the time effect was immediate. In other words, you see it right away. Obviously, you can't do that always. It is hard for a sponsor in advance to know what that thing is going to be, whether it is going to be four hours. Imagine you have a situation where you have a drug whose concentration Cmax is at two hours, the Tmax is at four hours and then it is gone, and you are looking for that within--you know, you have a relatively short period of time in which the thing is prolonged.

Now having said all of that, if you look at quinidine itself which is a drug, you know, known to cause Torsade. The Torsade seem to occur in the early phases of when the drug is given, shortly after change in dose or shortly after a rapid infusion. It is debatable whether a decrease might do that as well. But it is very possible that averages are not the biological parameter we care about anyway; that a high number in general simply reflects the fact that at some time points you are much higher than that, or you are changing

quickly.

So, I think the models you need to put in, in terms of delay--I think the metabolites are a totally appropriate model and it could actually be that a delay in a metabolite would simulate that perfectly well, I think. The models that you need to build in need sometimes to be models that can that can pick up something that happens over a relatively short period of time during the dosing interval.

DR. SHEINER: So, what you are saying, and I think it is a good idea, is that you consider other models for the drug effect. You add that one that was perfectly proportional to concentration. I am fascinated by the adding one that goes up and then has a rebound and then comes back to baseline because that, you know, with the averaging, would really create havoc for anybody to detect it. You can do all this stuff with simulation. I think it is a nice opportunity.

DR. VENITZ: I would also suggest, as Lew already said, not only to look at heart rate as a covariate to explain your QT, but look at drugs that change heart rate and QT at the same time. We are going to hear about sotalol in a minute which

does exactly that.

DR. KENNA: Okay.

DR. VENITZ: So, can you differentiate the primary effect of heart rate on QT versus the intrinsic effect that the drug has on prolonged repolarization? That might be a significant issue.

DR. SHEINER: This is a quick question. What do you have, 48 patients that you are resampling from?

DR. KENNA: When we resampled there were 45 I believe.

DR. SHEINER: Is there any thought on whether--it is a funny thing, it is 5,000 simulations but 48 distributions. You kind of wonder how you should trade those things off.

DR. DAVIDIAN: Yes, I was wondering that myself. I am not sure; I am not sure exactly what I think. That is what you have available, right?

DR. KENNA: Yes. Well, we have other data so we are up to about 100 subjects having four baseline days. Peter had an approach to address that issue, and it was if you assume that there is no diurnal variation he would pick different points on the time axis and shift it that way so that you were getting a difference. Peter?

DR. LEE: Yes, if you have a continuous measurement and you don't assume that there is a circadian variation that doesn't repeat itself, later if, for example, you want to simulate to baseline you could pick, say, a 12-hour baseline here and then pick another 12-hour baseline even over the original 12 hours. With that approach you could literally get hundreds, thousands of simulated baselines with 50 subjects or even 100 subjects.

DR. DAVIDIAN: I just have a question. Did you simulate a case where there was no treatment effect and see what the power is?

DR. KENNA: This is Peter's call.

DR. LEE: Yes, there is a placebo arm and there is a treatment arm. So, there is comparison between placebo and treatment.

DR. DAVIDIAN: So, when there is no treatment effect at all--you had that hump, right?

DR. KENNA: Yes.

DR. DAVIDIAN: So, what if you just had the same?

DR. KENNA: Yes, there is a placebo arm without any effect.

DR. DAVIDIAN: Suppose there really were

no treatment effect, you are doing it at 95 percent--

DR. KENNA: Yes, I guess we are revealing our regulatory spin, which is looking for the false negative--

DR. DAVIDIAN: Sure. I was just wondering because some of these powers that are higher than others might be the fact that at no treatment effect it is, you know, not consistent there. So, that could possibly carry over to where there was a treatment effect.

DR. SHEINER: Let me ask you about that because they are doing pretty standard statistical tests. I mean, once they have their statistics they are doing a pretty standard test on it. So, do you really think it isn't operating at the right--

DR. DAVIDIAN: I would expect it were but just for completeness I would do it, just to be sure, just in case there was something strange going on, you know, working with these maximums, or whatever. I don't know. I would think it would be fine, but just to be sure.

DR. KEARNS: Leslie, I am going to ask you a question that is theoretical and probably a

little unfair but it is after lunch, so. I am sitting here, listening to all this and looking at your excellent presentation and thinking, well, the approach is evolving on how to examine QT data. So, sometime we are going to come up with something that is going to be predicated from a lot of adult studies, and I am thinking about the pediatric world where--and I should publish this--we observed in a study of cisapride what I have called the pacifier effect on QT. If I have a baby and I am doing an ECG, getting a reasonable QT and the baby is crying, and I measure it and I put the pacifier in the mouth of the baby it changes. It changes very quickly, which has nothing to do with diurnal anything. So, how do we take this and apply factors in another population that may drive this whole thing in a much different way?

DR. KENNA: Then, the other thing to consider is that both of us have looked at baseline variability, and Peter looked at placebo variability, I don't know if the drug effect on top of that is somehow an interacting component or if that is just additive on top of that. So, that is another thing to consider.

DR. JUSKO: I have a question that kind of

relates to the underlying mechanism. Dr. Lee pointed out that most of the studies that he found most believable with terfenadine were multiple dose studies. Dr. Bonate did simulations based on the multiple dose regimen. Most of what you presented, although you proposed doing steady state experiments, is based on a single dose exposure. Is it known with these drugs whether the duration of exposure is a factor in changing QTc intervals?

DR. FLOCKHART: That is partly what I was trying to get at. I think it goes beyond that. I think the actual risk you are incurring might be different for different drugs. So, in the case of Seldane, you know, the studies that Peter Honig did were steady state studies in which he did see a real increase. That is where the 6 msec comes from. He could see a real increase when he measured the QT before the dose in that kind of trial design.

Lots of other people did sampling in other ways and missed that effect. But if you look at the real time effect in Peter's studies there was absolutely no debate that in a short period of time--we did a similar thing with pimozide. There was a short period of time when it was

unquestionably prolonged and then it goes away. The problem is, and the thing I am trying to figure out how to do in terms of statistics, if you have the possibility--if you have a data set there and it is possible that out of a 24-hour time interval you have 3 hours during which it is prolonged, and you don't know when that is. It might be immediate; it might be 8 hours later. How do you do a statistical test that allows all the multiple comparison testing, and all the other things you guys do, to pick that up? Does that really hurt your power or can you design it in such a way that you are able to simulate it well enough to pick it up?

DR. SHEINER: That is a little bit like what the maximum does. I don't like the maximum as a statistic. You just pick the longest QT you saw all day long. In a way, it is saying let's find the worst point, and you can do statistics on anything. So, the nice thing about this kind of simulation thing is you could add in an effect which was essentially a spike at six hours, even though the dose was given at time zero and the concentration didn't spike then, and analyze that. What is the kind of design, what is the kind of

analysis that, under the constraint that it have the proper operating characteristics under the null, gives you the greatest power? The greatest theoretician could tell us but otherwise you could just grind away and find a reasonable one.

DR. LESKO: I don't know if you had mentioned this or not, but in the six studies on that one slide--six drugs, I should say, which represent six studies, what was the range of subject numbers across those studies? What was the sort of range between subject variability given the different baseline methodologies? It was slide number 12. What was the range of subjects in those cases?

DR. KENNA: In terms of the numbers?

DR. LESKO: Number of subjects, yes.

DR. KENNA: They were fairly similar. I would say anywhere from about 40 to about 60 subjects seems to be what we are seeing.

DR. LESKO: And how about the variability within each case given the way the baselines were varied? For example, which one had the highest and lowest variability?

DR. KENNA: Between confidence intervals? I would have to go back and take a look at that.

DR. LESKO: I was wondering did the studies control for diet or food effects at all? How much attention is paid to that in the study design?

DR. KENNA: Well, I know they pay a lot of attention to when they are going to sample blood. They definitely lay out that they don't want to poke somebody and then do a QT interval. I haven't seen so much in the way of food till more recently.

DR. LESKO: Yes. Is it controlled, do you know, from placebo to drug?

DR. KENNA: I think the meals were the same for all arms of the studies, but in only two of these six I believe were meals really paid attention to.

DR. VENITZ: Any additional comments or questions for Leslie? Yes, go ahead.

DR. MCCLEOD: One thing you may want to start thinking about including in your model in the future is going from the QT interval to Torsade de pointes because that is what is cared about. You can now model in either allele frequency for the high risk genotypes or preclinical data on sensitivity of HERG, whatever other channel to the drug. I know it is premature to include it now

because you are generating the front end, but that way you get to a point where it might get to what Peter talked about at the end of his talk where you can stop using to kill drugs and start using it to better select drugs in an earlier setting.

DR. KENNA: That is a great idea. Thanks you. Thank you very much.

Committee Discussion

DR. VENITZ: Thank you, Leslie. If you don't mind, can you post the questions so we can kind of go through them one at a time? I think the first one is asking for the committee's input on additional study design points for the analysis. Any additional comments on study design?

[No response]

Then what about question number two?

DR. FLOCKHART: Lew and I were talking over here. I think the thing about the maximum--it is so easy to critique but often it actually represents the most important thing you are going after and it is what, in my experience, is very often the most valuable thing. The problem is that to determine whether the maximum that you actually determine is not just a random fluctuation.

So, in study designs it would be possible

to figure out how many patients you needed to study to figure out where the maximum is basically in a pre-study and then, subsequently, to intensely sample around that. That would get around the issue of what we are really doing all the time; we are testing for some long period of time in the hope that during that period of time you are going to pick something up. It is not really a time-directed thing. So, the right way to do it or a reasonable way to do it, if you are not dealing with something that stays up for days, weeks and months and then comes down but usually you are dealing with something that does this, is to determine where the time is first and then intensely sample right there, and Leslie's model would be great to test that in. You could basically figure out how many patients you needed to get power to do that for a given change.

DR. SADEE: It is not quite clear to me, since this is such a major issue for the industry and can cost extraordinary amounts of money one would like to ask what would be the best way of studying this. The way I would go about it, and there is a lot of literature, if we agree that polymorphisms do play a role in whether or not a

person responds more or less, a company would go ahead and sample, let's say, a 1,000 patients and genotype those 1,000 patients to get a fair representation--or let's say 2,000 and select 50 patients that are representative of the major phenotypes, in which case one would have much greater assurance of seeing unusual reactions that one would have to then treat very carefully, maybe with lower doses, because one is probing exactly where one should be probing.

So, I am not sure. That wouldn't be such a big expense to actually find these people because apparently it is done with every single new drug. So, that would be my suggestion.

DR. FLOCKHART: Are you saying, Wolfgang, to simply collect the DNA and keep it? I mean, I would totally endorse that, but actually finding it right now would be--I mean you would have to take a trip to Stockholm to be able to do that right now.

DR. SADEE: Well, there are a lot of polymorphisms known and the five candidate genes so you and you just then would sample a population for these 15 main polymorphisms and select your study population of 50 people.

DR. FLOCKHART: Well, I think there are a

number of issues there. One is I think we have registered that the five candidate genes only explain only about at third or, at most, a half of the total deal. So, we would be missing a half to two-thirds by doing that. I would never argue against collecting the DNA; I wouldn't do that. I think right now though it would be incredibly hard to do. You have so many variants and so many genes. I mean, there are more than 500 you would actually have to put in the pattern. You might mathematically be able to do that but at the moment it would be extremely challenging I think.

DR. SADEE: It would be challenging but considering the amount of money that goes into studying this and the failures, and if you really would catch half of the problem I think it would be worthwhile.

DR. SHEINER: You are not talking about simulation now. You are talking about an enrichment design where you have a bunch of people and you keep on having them come back every time you have a new drug and say you are a panel. I think that is a kind of futuristic vision and I think it is a good idea, although the safety issue would be something that people--but I guess you

would watch them very carefully and I suppose you could do it.

DR. VENITZ: Just a more general comment along the same lines, I am not sure how much longer it will be ethically justifiable to actually expose individuals, without having genotyped them, to positive controls. You would obviously emphasize the need or at least the possible need for positive controls to rule out baseline changes. What that means is that you know a healthy volunteer, who is not going to benefit other than the stipend that you pay him, is going to be exposed to a risk.

DR. FLOCKHART: But we are doing that. We are doing moxifloxacin in positive controls all over the place.

DR. VENITZ: And I am saying wait until the IRBs get full understanding of what we are testing for and it may not be permissible any longer. That is what I am basically telling you.

DR. HUANG: Jut to clarify, you are suggesting that maybe certain subjects with certain genotypes, that we actually recruit them to the study. A lot of times our study protocol will pre-specify subjects with certain prolonged QTs are not qualified. So, in a way, you are saying we

want to modify the protocols purposely to include subjects with baselines that are higher than normal, than the usual limit that we have set up.

DR. SHEINER: I think it kind of goes against--how can I say this?--the current philosophy which would say let's find the biomarker like the QT, bad as it is, that regular people can demonstrate without danger, which we believe is an indicator that the people who have a high propensity will get into trouble, and that will occasionally knock out drugs that weren't going to bet anybody into trouble and it will occasionally miss things. But I think that is more sort of in the philosophy. What you are suggesting is a very empirical approach, which is let's get the people who are in trouble and try it on them, under conditions we can control, so we will know for sure. I think the whole philosophy, if you will, of clinical trial simulation is that you are doing all this kind of stuff with the data to see how we ought to best test this is more in the direction of trying to see what we can do without actually exposing people who could get hurt.

DR. VENITZ: Any other comments about question number two? Other methods? We talked

about genotyping, preslecting.

DR. SHEINER: I just wanted to add I think it is a very powerful tool and I love the idea of sampling from real data. I mean, that at least gets you away from having to make a bunch of assumptions that you can't justify about distributions, and if you have lots of data--that is one of the things I have always thought, that the FDA is in a wonderful position. They have all this data that is handed to them in a more or less machine-readable form and they can do these kinds of simulations. They are limited only then by the kinds of subject matter imagination, like the sort of thing David was suggesting, that those models for drug effect be varied across a much wider range than just proportional to concentration. I think you may well find that there are some designs that are, you know, much better than others and that is at least a place to start.

DR. SADEE: If there are limits as to what the QT interval would be and those individuals who are truly at risk would be excluded, then I do see a problem with it. So, maybe one should rethink that because you could then say, well, these individuals should be exposed to maybe one-tenth of

the dose so that the risk is reduced because eventually, if you don't test these individuals, you will hit them with any new drug coming on the market and it will cause fatalities. So, there must be something about how can we prevent this type of risk by tests that are more forward looking and more realistic, and at the same time not put people at risk.

Alteratively, I don't know whether one can study cardiomyocytes directly electrophysiology but I suggest that to companies that deal with stem cells. They could turn them into cardiomyocytes and genotype them and have a panel and that would be another methodology to look into in vitro.

DR. VENITZ: Let's move to the last question, question number three, clinical design elements to identify meaningful change in QT.

DR. KEARNS: One of the comments that Leslie made at the beginning of her talk was about the attitude perhaps of the agency for looking at this with some kind of idea of wanting worst-case, especially for drug-drug interactions. I think something that is critical in an interaction study is understanding the potential of both drugs to have an effect on QT, which has not been done

uniformly. There are a lot of assumptions in the 3A4 interaction arena that if you give an inhibitor and you increase the AUC of the drug that can alter QT that you will automatically increase the risk, only to find out that the inhibitor also has an effect. That wasn't in all cases assessed independently. So, I think it is critical to think about that before making generalizations because the implications of a pharmacodynamic interaction here may be far greater than a pharmacokinetic interaction.

DR. VENITZ: I don't have a comment but I have a question. What is a meaningful change in QT that you are trying to identify? Obviously that drives your own measurement mechanisms. So, what is considered to be meaningful so that you have a decent target that you can shoot for, because I don't know what it is?

DR. FLOCKHART: It is Seldane right now; it is terfenadine right now. That is what it is. If it is like terfenadine it is meaningful.

DR. VENITZ: I guess I am trying to point out that, as much as I understand what you are trying to accomplish in terms of trying to find very small differences and correcting for as many

of the unknown variances as possible, that doesn't give you a meaningful change. That just gives you a change that you are able to detect with lots of sophisticated methods. I am personally not convinced that a 6 msec change in whatever the mean QTc is a meaningful change.

DR. FLOCKHART: Well, let me just expand a little bit. Obviously the 6 msec only looks at one side of the equation. It is a risk/benefit analysis. Seldane is kind of easy to beat on because the efficacy of treating a bit of a stuffy nose is not considered sufficient benefit for a lot of women to die. But in many, many, many situations we are not talking about that; we are talking about drugs that add real benefit for people. So, it is 6 msec weighed against something that we really have to deal with most of the time. So, I think 6 msec for Seldane is really the outside end of it. It is the most extreme situation where you have relatively little benefit and a very significant harm relative to that.

We haven't talked about how we are weighing, but I think the answer to that question, what is clinically significant, actually varies a lot depending on what benefit. It is not like

drugs are bad or drugs are good. I mean, these are parameters, unfortunately, of benefit versus risk.

DR. LEE: I also have a question. That 6 msec or 10 msec change, are we talking about change from pre-dose or change from the average over 24 hours?

DR. FLOCKHART: The way it was used with Seldane; the way it was used with terfenadine, which is the change I believe from the average of one day versus the average of a steady state treatment day.

DR. BONATE: I have a comment. We talk about terfenadine as the gold standard but let's not forget how many millions of people took terfenadine when it was the number one selling antihistamine on the market for years, and years, and years, and how many cases of Torsade were reported. Is there any reasonable expectation that in a phase 3 study we are going to be able to detect a QT change of significance for Torsade or are we fooling ourselves? I mean, is this a postmarketing thing that we should be considering?

DR. FLOCKHART: Well, no one would suggest that we actually want to power it to detect Torsade, I hope.

DR. BONATE: I think it is just a matter of perspective.

DR. HUANG: And I would add that knowing terfenadine and its metabolic pathway, with our current recommendation we really want to push the exposure up. I mean, the terfenadine itself may not really pose a significant problem, it is when it is used with an enzyme inhibitor which greatly increases exposure where you can actually see plasma levels with the contemporary detection method. It is really the maximum exposure that would have QT effect. If this drug is not metabolized, has no interactions, it is not really a big concern and it would not be a gold standard.

DR. VENITZ: Any further comments or questions?

[No response]

Thank you. Then, we are going to move to our next topic for today, and that is a pediatric topic. Here we are going to review the pediatric decision tree that we heard about in both of the previous meetings. Again, I am going to ask Dr. Lesko to give us an introduction to the topic.

Pediatric Bridging: Pediatric Decision Tree Introduction

DR. LESKO: We are going to switch gears on you again and cover, as Dr. Venitz said, further discussions with the pediatric bridging area and the pediatric decision tree. I will be up here relatively briefly to introduce the topic before I turn it over to some of the others.

[Slide]

This is the pediatric decision tree that was posted as an addendum to our Exposure-Response Guidance, and it is really a general framework that we have been dealing with in assessing pediatric approvals and extrapolations of efficacy from adult databases.

In the decision tree I have highlighted with underlines a few things, as you can see--similar disease progression; similar response to intervention; and similar concentration-response relationships; and then down below, on the right-hand side, similar levels to adults. So, similarity comes into play in practical applications of this decision tree and part of what we want to look at today is what does that exactly mean, what does that similarity mean both conceptually and what does it mean quantitatively.

[Slide]

The background in pediatric bridging refers to the extrapolation of efficacy. It doesn't refer to the extrapolation of safety. Safety and dosing must both be determined in the pediatric population. We also have some conclusions that we have to make from that pediatric decision tree, similar disease progression, similar responsive to therapy and also similar exposure-response relationships.

Many factors come into play in applying this decision tree in a regulatory decision framework. Some of those factors include the bullets on this slide--prior experience with the classic drug, whether it is first in class or one from a well-known class; what data might be available from older children; age-defined subgroup differences and efficacy that we might be aware of; the prevalence of the disease in various age groups and we are talking about a host disease or a disease that involves a host and either microbes or viruses. So, all of these factors come into play on a case-by-case basis to interpret the decision tree.

[Slide]

There are some clinical pharmacology

issues in here. PK and safety may provide enough data to extrapolate the adult efficacy and define the pediatric dose, but that really leads to two questions. When may the concentration-response relationship differ between adults and pediatrics? What is it we know about that? Secondly, how should the similarity or differences between exposure-response relationships be determined? So, these are pivotal questions that we are going to focus on today.

[Slide]

The way we are going to do that is to look at two case studies. These are examples of different approaches to the pediatric extrapolation and dosing. They illustrate different principles. Then the case studies will lead to a general approach that will look at comparing PK to relationships between two populations. Finally, we will close out this session with some input from research experience with Dr. Kearns in the use of the pediatric decision tree in conducting trials, and the regulatory experience from Dr. Bill Rodriguez in terms of applying the pediatric decision tree in regulatory decision-making.

Now, the questions for this session, which

we will get back to at the end but just to lead into them, would be basically to provide a case study perspective; provide some feedback on the current use of the pediatric decision tree in the framework of the case studies that will be presented. We are looking for some input on the methodology that will be presented to determine similarity of exposure-response relationships and then, finally, maybe some discussion around the assumptions that are inherent in terms of adjusting dose and exposure, and under what circumstances the assumption of similar exposure response might deviate what we think it to be.

So, with that in mind, I will transition to the first presentation.

DR. VENITZ: Our first speaker is Dr. Peter Hinderling. He is with the Office of Clinical Pharmacology and Biopharmaceutics. Peter?

Case Studies

DR. HINDERLING: Thank you.

[Slide]

It is a particularly interesting situation I find myself in because I will discuss with you the data, now as a regulator, that I previously obtained together with my colleagues in the

pharmaceutical industry. Also, I would like to point out that the data that were obtained were obtained in 1999, which is four years ago.

[Slide]

So, sotalol pediatric decision tree and exposure-response relationship: First of all, I would like to talk about the indication of sotalol in adults and briefly summarize the important pharmacokinetic and pharmacodynamic characteristics of sotalol. Sotalol in adults is indicated for life-threatening ventricular tachycardia and ventricle fibrillation, and a little bit later also an indication for maintenance of sinus rhythm in symptomatic atrial fibrillation and flutter.

The PK of sotalol in adults is linear. There is high bioavailability. The drug is largely excreted unchanged and the half-life is about 12 hours. The PK/PD is linear with respect to Class III antiarrhythmic activity as well as for beta-blocking activity.

I also would like to point out that the pharmacokinetics of sotalol are non stereo-specific, however, the pharmacodynamics are in that the beta-blocking activity is basically due to the L-sotalol moiety, whereas the Class III

antiarrhythmic activity is shared by both the DL and Tl form.

[Slide]

What was the knowledge of sotalol PK and PD-wise in pediatrics when we started the studies? There were a few published however uncontrolled studies in children that used the adult doses which were adjusted for body surface area or body weight and used the dosage interval which is used in adults, namely 12 hours. However, looking more carefully at those studies, it became apparent that at the end of the dosing interval of 12 hours there were some breakthrough arrhythmias.

[Slide]

Study demonstration of efficacy and safety of an antiarrhythmic in the pediatric population is a particular challenge. If you think about suppression of the arrhythmias as well as demonstration, for instance, of Torsade de pointes in children, this is clearly a challenge which cannot be surmounted.

Basically, Lipicky--and I would like to cite his paradigm--proposed the following: Do what is feasible in children, see what can be extracted and use it. In the case of antiarrhythmics where

the demonstration of efficacy even in adults is shaky, it is not reasonable to ask for efficacy in children.

[Slide]

Basically, we had to determine biomarkers instead of real clinical endpoints. The biomarkers that one can use are the Class III probes for activity, antiarrhythmic activity, as well as safety, the QTc interval, and then the resting RR interval to check out, again, efficacy and safety of the Class II activity of the compound.

[Slide]

Here is the pediatric decision tree which you just saw before. In the case of sotalol, based on some of the published data, it was reasonable to assume that there was a similar disease progression as well as a similar response so we could say here to both yes.

The next question, is it reasonable to assume a similar concentration-response in pediatrics and adults? The answer here is we don't really know. So, we say no.

Is there a PD measurement that can be used to predict efficacy? Yes, as we just saw. Therefore, conduct PK/PD studies to get the

concentration response for the PD measurement. Conduct a PK study to achieve target concentration based on concentration-response relationship and conduct safety trials.

[Slide]

The written request that we obtained stipulated the following studies? First of all, a PK study, an open-label, single-dose study, one dose level with extensive sampling, at least six neonates, at least ten infants, and least ten preschool children and at least ten school children.

A second study, a PK/PD study, similarly open-label but a multiple ascending dose study using three dose levels, with sparse sampling. This study should be done in at least either eight neonates or eight infants.

[Slide]

The study protocols--the PK study used a single dose of 30 mg/m
2. This label extrapolates from adult data. The PK samples, 12, were taken over a period of 36 hours after administration. The PK/PD study was executed at three dose levels, 10 mg/m
2, 30 mg/m2, and 70 mg/m2. The 10 mg was not effective, we knew that; 30 was and 70 was the

uppermost dose that could be tolerated that was considered safe. We used, as you can see here, an 8-hour interval because of the breakthrough arrhythmias that were demonstrated in the published but uncontrolled studies. The sampling mechanism for both PK and PD was sparse sampling. We added for PK about 4-5 samples. Similarly we took about 4-5 samples for PD. We took very careful measurements over the entire dose interval at the same time of the day during baseline.

[Slide]

A brief summary of the methodology that was used--the formulation was a syrup and extemporaneous compounding procedure was used. A very sensitive assay, LC/MS/MS that required 0.4 ml of blood. The ECG, the same type of machine was used in all sites. Baseline values ruing the 8-hour dosing interval were taken. There was a blinded cardiologist. Measurement was manually using a digitizing pad. The QT heart rate correction was according to Fridericia or Bazett. Data analysis used the traditional and population approaches. PK used a linear two-compartment model. There was also a non-compartment model method used, and the PK/PD used a non-compartment

model dependent methodology using either linear and/or Emax models.

[Slide]

We enrolled 24 sites for the PK study and 21 sites for the PK/PD study. Totally, there were 59 patients enrolled and the database included 58 patients with analyzable PK data and 22 patients with analyzable PD data.

[Slide]

Here are the results. We looked first at semi-log plots in four representative individuals in all four age categories. Patient 1 was a neonate; patient 6 was an infant; patient 11 was a preschool child; and patient 21 was a school child. You see that the half-life is very similar in all four age categories. That tells us basically that the volume of distribution and clearance relationship ought to be constant and independent of age, weight or body surface area.

[Slide]

Here we see plots of the apparent total clearance against the body surface area. On the right-hand side you see that these data can be fitted by linear curves with small intercepts.

[Slide]

On the next plot we see all data of the entire population, 58 pediatric patients, and added to them 40 adults. You see on the Y axis area under the curve normalized for dose and body surface area against the body surface area. What becomes quite clear from this plot is that basically down to about 0.3 m^2, children that had body surfaces larger than that particular critical value behaved like adults. They are basically on one line. Below 0.3 m^2, which corresponds to an age of about two years, just about the end of the infant stage, you see that there is decidedly larger exposure.

[Slide]

Here is the dose-response relationship. In red you see the beta blocking effect; in blue, the effect on QTc. On the left-hand side you see the observed Emax. Again, these are point-to-point baseline corrected values. On the right-hand side you see the average value basically, represented by the area under the curve at steady state of the effect. You can see that increasing dose both affect increase, but it is clear that the beta-blocking effect, like in adults, is greater than the QTc effect.

[Slide]

On this slide we see the impact of body surface area on the PK. Red now means basically the young children, the infants and the neonates, and the blue represents the older children. You can clearly see, with respect to Cmax and AUC at steady state, that the young children, the infants and neonates, have a larger exposure than the older children.

[Slide]

This has an impact on the PD. Basically, the increased effects in the PD in the neonates compared to the older children are simply a consequence of the increased exposure in terms of the concentrations that we observed in the previous slide.

[Slide]

Here are some representative plots of the QTc intervals against the predicted sotalol concentrations in four individuals representative of the four age groups. You see that QTc was linearly correlated with the concentrations. There is some variability, as you clearly can see.

[Slide]

The same thing can be said for the plots

of RR against the plasma concentrations. There seems to be a linear relationship, quite a bit of variability.

[Slide]

In summary, we can say that the pharmacokinetics are basically linear and dose proportionate in children. The half-life, like in adults, is about 10 hours and is independent of body surface area. The clearance and the volume of the central compartment are linearly dependent on the BSA, and BSA clearly is the most important covariate. It is also clear that the smallest children, infants and neonates, have greater exposure and, therefore, need an additional dose adjustment.

[Slide]

You see that in this plot on the Y axis you have the age factor and on the X axis the age in months. So, we are talking about a person that has an age of two years and the factor will be 1. So, up to this point we would just normalize based on body surface area. However, if we go to smaller children this age factor would decrease to 0.5, 0.3 and we would have to multiply that factor into the dose equation.

[Slide]

With respect to PK/PD, the doses were tolerated well. The responses, as you have seen, increased dose dependently. Pharmacological important effects were obtained for Class III at the highest dose only for beta-blocking at the 30 mg/m2 and 70 mg/m2 dose. There was a trend for greater effects in smaller children entirely due to pharmacokinetics, and the effects were linearly correlated with the concentration. Interestingly, it was also noticeable that the beta-blocking effect increased with body surface area. Not only are the heart rates, of course, a function of age but also the beta-blocking effect has an age dependency to it. Thank you.

DR. VENITZ: Thank you. Any questions or comments?

DR. JUSKO: I have two questions for clarification. You were administering the racemic form and probably analyzing for both the DNL and combination.

DR. HINDERLING: No.

DR. JUSKO: What form of the drug did you administer?

DR. HINDERLING: We administered the

racemic drug.

DR. JUSKO: And you analyzed for both forms?

DR. HINDERLING: We didn't analyze for both forms. Preliminary data showed that there was no stereo specificity in terms of the kinetics, as in adults.

DR. JUSKO: And you are sure of that in young children also?

DR. HINDERLING: Yes.

DR. JUSKO: Secondly, when you measured the beta-blocking effects, I don't imagine you gave a stress test to the different--

DR. HINDERLING: No, it was the resting heart rate.

DR. JUSKO: No, just the resting heart rate?

DR. HINDERLING: You know, when you deal with neonates and infants--

DR. JUSKO: That is why I was wondering.

DR. HINDERLING: --there are some limitations. But, of course, all the kids were pacified.

[Laughter]

DR. LESKO: Peter, just one clarifying

question on the dose-response relationship that compared the beta-blocking effect on RR, the one that compared the percent delta Emax and percent delta area under effect as a function of dose at 10, 30 and 70--yes, that one. These are both relationships in children. Right?

DR. HINDERLING: Yes.

DR. LESKO: Did you have relationships of this sort in adults?

DR. HINDERLING: Yes.

DR. LESKO: And how were they when you compared them side-by-side? What was the shape?

DR. HINDERLING: It was basically very similar. The order of magnitude in adults was similar to that of the children. Therefore, one could really deduce that the concentration-effect relationship is really the same. The only difference is really due to the fact that the exposure in the youngest children is larger which can be, and has to be compensated by the appropriate dose adjustment.

DR. DERENDORF: Could you explain this AUE steady state?

DR. HINDERLING: AUE is basically the area under the effect curve taken over the entire zero

to eight-hour interval.

DR. DERENDORF: So, how many points?

DR. HINDERLING: Five.

DR. KEARNS: I think it was very fortunate for you in your previous life and your company that Dr. Lipicky said what he said.

DR. HINDERLING: Yes.

DR. KEARNS: And the bar for you to do these studies and to ultimately get approval and exclusivity was not raised but it was lowered a bit because I can tell you that if this were an antihistamine drug and there were patients that had more than a 500 msec QTc, it would have died a horrible, swift death. The trials would have been stopped and there would have been much worry. But here we have a pediatric study, a small number of patients and, of course, a drug that we expect to have some cardiac effects and the end result is quite different. So, that is not so much a question as a bit of commentary.

DR. HINDERING: I agree.

DR. VENITZ: Any other questions or commentaries?

[No response]

Thank you again, Peter. Our next case

study will be presented by Albert Chen and he is with OCPB as well. Albert?

DR. CHEN: Good afternoon.

[Slide]

This case study is from Merck's montelukast tablet. The brand name is Singulair.

[Slide]

Montelukast is a leukotriene receptor antagonist. It is indicated for prophylaxis and chronic treatment of asthma. Two original NDAs were approved simultaneously in 1998. One is for a 10 mg film-coated tablet for adults and adolescents greater than 15 years old. The other one is for a 5 mg chewable tablet for children 6-14 years old. The dosing regimen is one tablet QD given in the evening. Unlike the previous case study for sotalol, the 5 mg chewable tablet wasn't approved until the original request based on the previously approved NDA. Therefore, this case study is to show you the sponsor's rationale and thinking during the clinical development for the pediatric program prior to the NDA approval.

[Slide]

This is the decision tree. I am going to use this to explain this company's thinking and

rationale and I will use the same decision tree to summarize at the end.

[Slide]

I will go over adult PK dose-ranging studies; adult clinical efficacy and safety trials and then move to pediatrics in sequence. Adult PK was obtained in healthy volunteers. The basic PK information is shown here. A mean absolute bioavailability was about 70 percent. It was about 65 percent from the film-coated tablet and for the chewable tablet it was a little bit higher, 73 percent. It is extensively metabolized, greater than 86 percent of an oral dose of about 100 mg C14, the montelukast was excreted in the bile and through the feces. Only less than 0.2 percent was found in the urine after five days. The parent drug is predominant in the systemic circulation. We are presenting about 98 percent of the total radioactivity over the initial ten hours post-dosing. The T half-life is about 4-5 hours.

[Slide]

The first PK study is a dose comparison study. This is the pivotal study because it provided the head-to-head comparison between the 10 mg film-coated tablet and the 10 mg chewable

tablet. It also provided the dose proportionality information regarding the chewable tablet.

The objective of this study was two-fold It allows for conversion of the AUC from the 10 mg film-coated tablet to a 10 mg chewable tablet, after taking into consideration the difference in the absolute bioavailability, 73 percent versus 65 percent. It also allowed for scaling down the AUC of a 10 mg chewable tablet to a smaller pediatric chewable tablet dose in order to obtain similar AUC as adults receiving the 10 mg film-coated tablet.

[Slide]

The adult dose-ranging information was obtained from the subgroups of earlier phase 2 trials. the dose range studied from 10 mg QD up to 200 mg QD plus placebo. In the parentheses are the patients who participated.

The results of the study showed that the active treatments were all significantly different from the placebo, and no differences were found among the active treatments.

[Slide]

So, based on the above observations, the proposed dose selection for adult patients was one 10 mg dose QD given in the evening.

[Slide]

Two adult clinical efficacy and safety trials were conducted. Similarly, they were 12-week studies in patients with mild to moderate persistent asthma at baseline. The primary endpoint was changes in FEV1, forced expiratory volume in one second, and the daytime asthma symptom score.

[Slide]

These are the results obtained from clinical trial 01 during the four visits every three months regarding the mean percent change in FEV1 from baseline. The montelukast was significantly different from placebo at each visit. The overall mean of the four visits was 12.8 percent for montelukast and 4.1 percent for placebo. Regarding the mean percent change in the daytime asthma symptom score from baseline, montelukast was also significantly different from placebo.

[Slide]

Results from clinical trial 02--the same results were obtained.

[Slide]

Also safety profiles between active

treatments and placebo were found to be similar. So, the proposed dosing regimen was confirmed by adult clinical efficacy and safety studies.

[Slide]

Now we move to pediatric studies. Since montelukast is a new molecular entity and a new class of drug without previous pediatric data, the sponsor's answer to the above two questions is no and this is for the case of 6-14 years old. So, the sponsor conducted PK studies and also safety and efficacy trials.

[Slide]

Pediatric PK was obtained in pediatric patients only. Study 02 is a single-dose PK in early pubertal adolescents 9-14 years old. Two dose levels were tested, 6 and 10, using the film-coated tablet. Study 03 was a single-dose montelukast PK in pediatric patients 6-8 years old using the 5 mg chewable tablet.

[Slide]

Table 1 shows the mean PK data obtained from the pediatric PK study 02 and also compares with the adult historical data. Pediatric patients not greater than 45 kg received the 6 mg dose and pediatrics greater than 45 kg received the 10 mg

dose. This is the adult historical data using the 10 mg dose. For this age group the systemic exposure in terms of AUC is about 2,900. It is very close to the adults receiving 10 mg film-coated tablets, about 2,700. Actually, this value is within the mean adult AUC plus/minus two standard deviations. For this age group the AUC is too high.

[Slide]

Table 2 shows the mean PK data obtained from another pediatric study. For this age group the 5 mg chewable tablet dose was given. As you can see, the AUC is about 2,900, very close to the adult AUC 10 mg film-coated tablet. So, based on the dose normalization in AUC, it was concluded from table 1 after converting a 6 mg film-coated tablet, a 5 mg chewable tablet given QD to children 9-14 years old is expected to provide similar systemic exposure as adults receiving the 10 mg film-coated tablet. From table 2, similar AUC in 6-8 year old patients was obtained.

[Slide]

So, the 5 mg chewable tablet was chosen for the pediatric efficacy and safety trials. Since montelukast was a new class of drug, this

study was conducted to confirm the dose selection and also to prove some concept and assumption which I will explain later. I put a note here that since the adolescents, 15 years and older, had similar plasma profiles compared with adults, they were included in the adult phase 3 trials.

[Slide]

So, for this age group of 6-14 years old no pediatric dose-ranging trials were conducted. What are the assumptions? Similar disease progression in asthma between pediatric and adult patients and comparable efficacy is associated with similar systemic exposure in terms of AUC.

[Slide]

So, this pediatric clinical efficacy and safety trial was an 8-week treatment study in more than 300 pediatric patients. The mean percent change in FEV1 from baseline was 8.7 percent for montelukast and 4.2 percent for placebo, and the difference is statistically significant. So, the original NDA for the 5 mg chewable tablet was approved for 6-14 years old.

[Slide]

Now we move to younger pediatric patients, 2-5 years old. Based on the previous successful

experience in dose selection, the same principle with similar mean AUC, a smaller 4 mg chewable dose was selected. This dose was tested in a PK study employing sparse sampling technique using a pop PK approach. The mean AUC estimated was about 2,700, again very close to adult AUC for the 10 mg film-coated tablet.

[Slide]

Since efficacy has been demonstrated in children 6-14 years old, and the assessment of FEV1 in the children smaller than 6 years old will be problematic, it is decided that only a safety trial is needed. So, the sponsor conducted a 12-week clinical safety trial in greater than 600 patients. There was no dose-ranging study conducted, nor formal clinical efficacy trial conducted. This study actually supported the safety of the 4 mg chewable tablet in this age group and also confirmed the efficacy in this age group. So, the 4 mg chewable tablet was approved later for the children 2-5 years old. It is under internal request based on the approved NDA.

[Slide]

After the sponsor learned more and more from the previous case, 6-14 years old, and they

are willing to answer yes to the above two questions, and to assume a similar concentration response in pediatric patients, and this is the case for 2-5 years old, the sponsor only conducted PK studies and safety. The safety trial actually included a secondary efficacy assessment, and they proved that efficacy is okay in this age group.

[Slide]

I would like to thank my previous medical colleague Dr. Bob Meyer, Peter Honig, Anne Trontell and also my supervisor, Dr. Larry Lesko and Shiew-Mei Huang.

DR. VENITZ: Thank you, Albert. Any questions?

DR. DERENDORF: Yes, in the decision tree it says that it is reasonable to assume similar exposure response in pediatrics and adults. If you look at the data that you have in adults, first of all, you really don't have a good exposure-response relationship. You have a placebo and then you have a range of doses that all do the same thing.

DR. CHEN: Well, that is the phase 2 trial. Because the safety profiles looked very clean the company actually precluded the dose-response study. But with the development of

the guidance, we will probably ask the company to conduct it but at that time they did not conduct a dose-response study.

DR. DERENDORF: Right, but what you did, conceptually, you took one of these doses and you reproduced the same exposure in terms of AUC--

DR. CHEN: Right.

DR. DERENDORF: --in children and they also were different from placebo, but that is different than having the same exposure-response relationship.

DR. CHEN: That is true but this is a special case and they selected the smallest dose.

DR. DERENDORF: We don't know if it is the smallest.

DR. CHEN: The company reported the effective dose could be as low as 2 mg but they submitted the report for review.

DR. LESKO: Just to follow-up and make sure I understand the point that Hartmut was making, the early decision was that there was no information basically to assume that disease progression response to therapy would be the same. So, there was a PK study. It was sort of a hypothesis in the first age group that exposure

response was similar. Once it was demonstrated for an older age group, you sort of went back to that top box and said now I have some data that sort of underpins the notion that I can answer yes to both of those, and then subsequent age groups went down a different path.

But I think the efficacy in the pediatric older children, 9-14 or whatever it was, had a similar change in clinical endpoints as the adults had for similar exposure. So, that was pretty confirmatory at that point that the answer would be yes to the first two. I think the percent change in FEV1 was 9 versus 12, or something very close, so that exposure response was similar.

That gets to your point because if that is the case, then what you said wasn't clear to me, the point you were trying to make.

DR. DERENDORF: The point I was trying to make is that if you don't have any data on the lower end of the children, which I don't think you have or at least it is not in here, it would be possible that there is a different concentration or exposure-response relationship that you just don't pick up. In children maybe a lower dose would do the job.

DR. LESKO: Okay, so targeting the same exposure--

DR. DERENDORF: Oh, it wouldn't be the same exposure. If the exposure response would be different, you wouldn't know.

DR. LESKO: Yes, we don't know the shape of that relationship basically.

DR. SHEINER: Similarity at one point doesn't necessarily mean similarity elsewhere.

DR. VENITZ: Any other comments for Albert?

DR. SHEINER: Let me pursue that point because it is interesting. Remember, we are in a pediatric situation and we are trying to do something reasonable. So, if you had good safety and you had similar response which is acceptable at one point of the dose-response curve, wouldn't that, in the pediatric case, be enough to say, well, okay, go ahead and do that? Even if it is possible conceptually that you could have exactly the same response in children, nonetheless, it is giving you good response, similar to adults; it has adequate safety and, you know, maybe it is okay.

DR. LESKO: Yes, it is almost like the dose selection was based on PK but the real trump

card, if you will, was the evidence of efficacy and safety in that clinical trial. Yes, the open question is could those results have been achieved at a lower dose maybe? But the dose that was achieved, it wasn't bad.

DR. VENITZ: Thank you again, Albert. Our next presenter is Dr. Stella Machado, and she is going to introduce a method to compare exposure-response relationships and see if they are similar or not.

Methods for Determining Similarity of Exposure Response Between Pediatric and Adult Populations

DR. MACHADO: This is a great privilege, to be here, speaking with you this afternoon.

[Slide]

I will be talking about methods for determining similarity of exposure response between pediatric and adult populations. I am with the Office of Biostatistics in CDER, and we are working together with the team from OCPB in a real situation, pediatric bridging situation.

[Slide]

I would like to acknowledge substantial contributions from my colleague, Meiyu Shen, who is also in statistics. We gleaned ideas from many

colleagues, both from within the agency and outside, and also even from the Internet.

[Slide]

This is not complicated statistics. It is more of a way of looking at things. I am just going to talk really in generality about a method for comparing two response curves with the pediatric population and adult population. This could be equally well applied to, for instance, comparing between ethnic regions or comparing response curves for gender and so on. I am presuming that the exposure metric could be dose, it could be area under the curve, it could be Cmin, whatever. The response metric could be a biomarker or could be a clinical endpoint.

[Slide]

The goal in bridging is to evaluate the similarity in PK/PD relationship between adults and pediatrics where we have plenty of the adult data, the original population, and the pediatric population is the new one. The conclusions we can come out with could be that we conclude similarity. Or, we could conclude similarity of shape of the dose-response curves but with some dose regimen modification needed. Or, we also could conclude at

the end of this a lack of similarity.

When we started working on this there really was an absence of precise guidance as to how we should proceed. What I am going to recommend is that really we are in an exploratory activity at the minute, not confirmatory hard and fast statistical testing situation.

[Slide]

Now, we did work with a real drug situation but for the purposes of this talk we invented drug X and heavily disguised it so that you can't guess what it was, the real situation. For drug X there were about 240 patients in the adults and 120 in pediatrics. Those are numbers close to the original. About 40 percent of each of the groups took placebo.

[Slide]

Here is our plot. Here is drug X. The triangles are the new population, the pediatrics; the squares are the original, the adults. How do we compare? How do we say this is similar or not? It is just, gosh, what a mess!

[Slide]

A little bit of notations but I am not going to go heavily into the statistics, we have a

different number of adult patients, generally a smaller number of pediatric patients. Y is our response measure and C is the concentration metric. I will call it concentration but, as I said, it could have been area under the curve or Cmin. Generally, the concentration measurements are all different unless you got data from a concentration-control trial. For drug X, you saw that the concentrations were all over the place.

[Slide]

To establish similarity we need to compare the average shapes of the response curves, taking into account variability of the measurements. The response curve depends on the exposure measure and some various unknown parameters. The adults and the children may have similar response curves but they may have different parameters.

[Slide]

As a first step, looking a little bit further at the data, these are lowest fits, local regression lines plotted onto the data and here we see for the first time that there seems to be a bit of a separation between those two curves. The upper curve is for the pediatric patients and, with increasing concentration, does seem to drift up

away from the adults. So, the suggestion is that there is some difference here but the big question is how much of a difference.

[Slide]

In terms of thinking about it, what we should be doing is assessing similarity between the responses at all the concentrations that are likely to be encountered. So, we are not interested in postulating response curves out into the very, very high doses. That is not realistic. We are interested in the distance between the curves, like the average behavior for the population and accounting for the variability of the response. We suggest an equivalence type approach rather than hypothesis tests, trying to test that the response is not significantly different.

[Slide]

So, where do we start? Well, the hypothetical situation is to focus on what we would do at a single exposure measure? One single concentration, what would we do? Well, this would reduce to the usual equivalence-type analysis and there are various ways to analyze this, different response metrics. We could look at comparing the average response between pediatrics and adults at

every exposure or a combination of average and variance metrics, for instance a population bioequivalence approach or Kullback-Liebler distance metric, or we could actually compare the whole statistical distribution, Kolmogorov-Smirnov type generalization. But we chose to look at the simplest of these, which is comparing the average response.

[Slide]

Again continuing, we are only talking about one concentration. We defined similarity to be the requirement that the average responses in the two populations, for the same concentration, are closely similar. We choose goalposts, for instance, the 80 percent or 125 percent which are familiar, and calculate a 95 percent confidence interval for the ratio of the average responses.

[Slide]

If the 95 percent confidence interval at this ratio falls entirely within our goalposts, then we say that the null hypothesis of lack of equivalence is rejected, therefore, we are accepting the fact that we have similarity here. This is the usual simultaneous two one-sided test procedure. So, our proposal is to use confidence

intervals to measure similarity, to quantify similarity, quantifying what was actually determined from the data we have in the two populations.

[Slide]

Just a note on getting the confidence intervals for this ratio, there is a bit of work required. There are some methods in the literature based on normal distributions. If you are not willing to make that assumption you could use the bootstrap method or computer simulation. My opinion is that it is easier to use the actual data. Then we end up with useful statements. For instance, we are able to say that the average response at this concentration, level C, among pediatrics is 93 percent of that in the original population, and we are 95 percent sure that the ratio of these averages lies between 83 percent and 105 percent. That is possibly a summary statement that we can deal with and make decisions from.

[Slide]

Moving away from one single concentration to the real situation where we have response curves over a whole range, the easiest thing to do is to categorize the concentration axis into

intervals--we chose five or six here--and for each interval estimate the 95 percent confidence interval for the ratio and interpret. A useful way to interpret is to use graphs.

[Slide]

Here is our drug X. That is the range of concentrations. There are quite a number of patients receiving zero dose of this drug. It is sort of interesting that the placebo dose actually falls below the 0.8 lower bound with no drug. I am not sure what that is about. But then there is a tendency for the confidence intervals to drift upwards, outside of the 80 percent to 125 percent, and definitely for the highest concentration range, 80 and above, and that is where we have the least amount of data so the confidence intervals are quite wide out there.

[Slide]

I summarized that. The ratios trend upwards and the upper limits exceed 1.25 for all of the exposures, all the positive exposures.

[Slide]

A second way of doing it is to actually fit a model to the data and estimate the unknown parameters; use the fitted model to simulate the

ratios for each different concentration and estimate the 95 percent confidence intervals, which we went ahead and did.

[Slide]

For fitting the models we actually found that the square root of the response stabilized the variance. The linear models were fitted separately. In the simulation we used 5,000 pairs of studies to estimate different estimates of the ratio and percentiles.

[Slide]

Here we have a smoothed plot of the confidence intervals for the ratio of the two means, again showing a drift upwards. I should say that these particular concentrations I chose for the graph were the mid-points of the intervals that I chose for the categorized concentrations. Because of the model fitting, this picture is quite smooth but we do see a great tendency for the ratios to climb, much bigger than 1, and we really see that for these higher concentrations this new population, the pediatric population, is substantially different from the adults.

[Slide]

Here is the graph of the two methods

compared. The first is the pairs from the simple, straightforward method of categorizing the concentrations, and the second is the model fit. They are kind of similar as we would expect; it is the same database.

[Slide]

In comparing the two approaches, I really feel that both are useful, the rough and ready one, but then the model-based method--well, you have to make some assumptions like actually fitting the model and what is the best shape for it but it is less influenced by outliers and generally has greater precision, not a huge amount, I must say, from this example. But I would say that both of the methods are useful. So, it is not particularly complicated but it will show you whether there are trends in the differences in the two population responses.

[Slide]

In terms of designing a study among the pediatric population, or another situation we looked at, if you are going from one country to another and you want to do a bridging study in the new country, the design should be based on parameter estimates from the data you already have

in the original population, the adult population, and any prior information that you have from the pediatric population.

Make sure to include doses that are likely to produce these concentration metrics in the whole range of interest. Then, perform simulations to determine the required number of patients needed in the new population. You can assess robustness to the model assumptions, and so on, your variance estimates, to see what would happen

[Slide]

I apologize for the spelling mistake here. This general approach can work for response curves for efficacy and for safety. What we are doing is proposing a method to quantify the similarity between the adult and the pediatric populations over the whole range of concentrations. Rather than trying to test that adults and children are different, we are trying to test how close they are and where they are close. This can be applied easily to data from trials with different designs. Then, as a final thought, I put up the usual goalposts such as 0.8 to 1.25, but that may well not be meaningful for this particular drug, depending on therapeutic range, or the disease of

interest. So, interpretation of how much similarity is acceptable, of course, requires medical input. Thank you.

DR. VENITZ: Thank you, Stella. Any questions or comments for her? Greg?

DR. KEARNS: I am glad to see your last point because I was troubled until you put this slide up. I think most of us would agree that the demonstration of statistical difference and clinical difference is not always the same. I mean, not knowing what drug X is, one could argue that that difference, in terms of a clinical context of drug effect, would be not meaningful despite its significance.

My question to you and really to anybody from FDA is what are the implications of finding a difference, especially when you are looking in a retrospective way? I mean, the data that you shared with us ostensibly would come out of the review of an NDA when all the pediatric stuff had been done, the adult stuff had been done and the company has performed now the pediatric studies with consultation from the agency, perhaps it is being done under the Best Pharmaceuticals Act so there is some hope of exclusivity; maybe some hope

of labeling. Then it goes to your Office and, voila, there is a difference. So, what are the implications for the agency to go back to the sponsor and say, well, it was a good try, boys and girls, but no exclusivity for you today because there is a difference between adults and children that we can't resolve from your data?

DR. MACHADO: Thank you, that is a very insightful question. I don't have a nice selection of slides of the pediatric decision tree, but there is one element on the pediatric decision tree that asks the question can we consider that the response curves for pediatrics and adults are similar enough. So, what I am addressing is part of the whole pie that goes into deciding whether to approve a drug for pediatric use. Larry, would you like to comment on that?

DR. LESKO: I guess it goes back to a case-by-case interpretation of the differences that you would observer in that case. Then, I think you would have to draw in some of the clinical efficacy data that were available and try to interpret that. I think the soft spot in this approach is what those boundary conditions are going to be. When you get to the end the 80 to 125 is a default that

we have borrowed from some other areas, but the problem with that is we have tried to apply it in other similar situations, like drug interactions or renal disease versus normals, and the number of subjects needed to meet that boundary condition, given the variability, is unrealistic.

So, the next question then is what are those boundary conditions that we be appropriate to declare similarity and it seems you go down two paths. One would be what do I know about the exposure-response relationship, and what are the boundaries I might draw from the shape of that relationship in adults, with the assumption that PK/PD is similar?

I guess the other question would be kind of a joint medical-artistic sort of approach, well, what difference would be clinically important if you were to think about it in an empirical way? But you have to somehow set some boundaries I think.

DR. VENITZ: The boundaries that we are talking about here are not boundaries on concentrations. We are talking about boundaries in the response--

DR. LESKO: They would have to be wider.

Obviously, the variability is going to be more than concentrations.

DR. LEE: I think my other question to the committee is should we also not only look at the mean value or the difference between the two mean curves, but also looking at the whole distribution of the PK/PD relationship because what we are really concerned about is not the typical patient but the patient who may be exposed to a very high concentration or very low concentration? So, do we really want to make sure that the distribution of the response is similar between adult and pediatric populations?

DR. SHEINER: You are going in a little different direction but we started talking about something that I think is pretty clear, that is to say, two different issues: How do you measure a difference between these two curves, let's say, and then what do you use as regulatory guidelines with respect to that measurement? So, the measurement has to be adequate to the task of ultimately making a decision. That decision issue is always going to be trickier than the measurement one I think. So, I would like to focus a little bit on the measurement one.

I just wanted to say that I noticed in one of your slides, Stella, that you had the statement--you know, we can make statements like we are 95 percent sure that the range is something or other. That kind of almost smacks of a Bayesian statement so I am going to take that as permission because you opened the door--it seems to me what we are really talking about is the posterior distribution, estimating the posterior distribution on some feature of these doser-response curves that talk about a difference. So, if it is in the log world it is a ratio. So, that might be what we are interested in or, as Peter just sort said, we might be interested in some other aspect of the curves than the difference in the means. We might be interested in the difference in the fraction lying outside of a certain range, or something like that.

So, we have to decide, it seems to me, what those things are and they are just qualitative issues of value, not quantitative which is the tough one. The tough question is the second question, where is the cut-off? But the qualitative issues of value, what kinds of things are we interested in, what are things that are relevant, I think we can probably agree on those.

I would say that, you know, personally I would just like to see us talk about posterior distribution of a difference of some kind between the two. Then I would make the point about that that when you get to regulating--even though I don't know how to resolve that--you do really have to be quite careful about saying that because there is a significant amount of the probability mass that lies outside of some acceptable boundary, though there isn't very much evidence that it is there. It just means you don't know very much. It is the same kind of story as, you know, accepting the null hypothesis in the opposite situation. So, I the hard questions are the questions about what regulations you make and how you regulate it.

I think the thing you finally drew there with those confidence intervals, they are not too different than a posterior distribution on the ratio, and you can computationally get it more or less the same way and I do think that is the right way to look at it, but I would say for those of us who tend to sort of enjoy being kind of the technical heads here, let's stop at making the picture that shows the differences and then let the regulators worry about where to cut off the lines.

DR. MACHADO: Thank you.

DR. VENITZ: Any further comments or questions? If not, thank you again, Stella. I suggest we take our break. We will take a 15-minute break and reconvene at 3:45.

[Brief recess]

DR. VENITZ: We are still continuing on our topic on pediatrics, pediatric decision tree, and our next presenter is our very own Dr. Greg Kearns. He is going to give us an academic perspective in using the pediatric decision tree. Greg?

Research Experience in the Use of Pediatric Decision Tree

DR. KEARNS: Thank you very much.

Larry gave me kind of a complex task here today. He said I want you to talk about the decision tree but I also want you to review some of the basic stuff on pediatrics and why are children different. So, if this is a little bit of a hodge-podge, forgive me; I am just executing my orders.

[Slide]

This is one of my favorite all-time quotes from the man who is considered to be the father of

American pediatrics. I like it because in 1889 Dr. Jacobi recognized that the issue of dose being different was of paramount importance.

[Slide]

One of the differences from what we have heard today about empaneling a group of professional subjects who go out for a bender, clean up and come in, is that few of our children that we have in clinical trials do that, maybe some of the adolescents but certainly not the younger ones, and there are many, many differences between adults and children and we tend to think of pediatrics as a continuum.

[Slide]

Certainly there is a physiological continuum. There is a behavioral continuum, all of which must be considered in the context of a clinical trial. We know that children are different. They have different body composition, as illustrated by these data. This impacts the pharmacokinetics, especially with respect to drug distribution.

[Slide]

If you look at their renal function as a function of age for pre-term and term babies over

the first two weeks of life, there are dramatic increases which, if you look at the kinetics of a drug like famotidine, translate directly into changes in the behavior, changes in the concentration-response relationship which are predictable when one simply looks at the pattern of development and its impact on GFR in this case.

[Slide]

As summarized by Alcorn and McNamara in a recent paper in Clinical Pharmacokinetics, if we look at many of the drug metabolizing enzymes and we express their activity relative to the activity in adults, look at them over age, in this case about 160 days, we see some patterns. It is the patterns that are so important for those of you involved in the modeling business because a pattern, to me, means prediction. Prediction is, as we have heard time and time again today, critical for understanding the behavior of something being studied or what might we expect in the context of clinical use.

[Slide]

In the case of something like cisapride--since we are talking about QTc I couldn't help but include one of my favorite drugs

in here--we are not going to talk about QTc but just the kinetics of this CYP 3A4 substrate very nicely go along with the delay in maturation for the enzyme.

[Slide]

If you take a group of very small babies that are not very mature and, in fact, have low surface areas because they are tiny, the clearance of this drug is markedly impaired, which is something you would expect to see. It is not only the enzymes in the liver, as we are finding out--Trevor Johnson and his colleagues, in 2001, looked at 3A activity in the gut and the same type of maturation pattern is evident. This, of course, has implications for bioavailability of drugs that are given to kids that are 3A substrates.

[Slide]

Phase 2 enzymes as well show a developmental pattern. These are some data from Martin Behm, one of our fellows. They were presented at the CPNT meetings in 2003. This is a plot of glucuronide to sulfate ratio of acetaminophen in urine, done in a group of healthy children and looked at, in this case, over nine months of time. Sulfotransferase activity comes on

very quick, as most of you know. UGT activity has a delay. So, if you look over time you see this ratio increase until about six to nine months when it seems to level off--again, another developmental pattern.

I would be remiss to not put the bars on here that indicate that there are outliers. Even at every developmental stage the inter-individual variability in the activity of drug metabolizing enzymes is very, very large. That is important because as we look at some of these pediatric studies with six neonates and the conclusions that are being drawn, it is--at least for me, anyway--a little statistically worrisome at times.

[Slide]

Then there are drugs like linezolid--and we were privileged to do this work several years ago--that are not metabolized by cytochrome P45; not substrates for UGTs. If you look at the impact of age on clearance, you see dramatic increases that suggest that something important, something interesting for this compound goes on in the first week of life but, again, a predictable pattern.

[Slide]

So, clinical pharmacology facts--kids are

not small adults. They have different PK for sure. In some cases the PD is different. Despite our advances, we are still in an age where about 80 percent of all drugs on the market are not labeled for kids. With rare exception, pediatric patients are still thought about late in the game of drug development, something we need to fix. The biggest issue far and away is what is the dose. What is the proper dose that will make the exposure that has the greatest chance of being effective and safe?

[Slide]

Previously, historically there were some challenges to pediatric drug development and most of these have been taken care of in 2003. Analytical issues, we heard so sotalol a method that required 0.4 ml of blood. PK/PD approaches abound. Some of the other scientific issues, the incorporation of pharmacogenetics; logistical issues, we have come up with ways to study children; designs; we have even dealt with the lawyers in some measure. Lawyers who used to say it is very risky to do studies in children; it was dangerous; it was expensive, therefore, we shouldn't do them; have now changed their tune

after the course of a few lawsuits. Ethical considerations have been largely taken out of the equation. Programmatic things, we have networks in our country now to study drugs in children. Even the FDA has gotten pretty sharp about this and have included children in their plans, hence the decision tree.

[Slide]

There are some remaining challenges, for sure. I think these are important, and these are things that have not yet been lit, to use a Missouri word. First, relevant extrapolation of adult data and animal data. There are times to do it and there are times not to do it. But, certainly, the adult data can still be critical.

Study designs--much of what we have talked about today, study designs that are optimal; scientifically robust so they don't make sacrifices beyond belief; study designs that are synergized by adding relevant science; and capable in as many cases as we can of truly addressing drug effect.

Then we need dosing approaches that control the exposure; that we can verify; and that, most importantly, are age appropriate. This even gets into the arena of formulation just a bit.

[Slide]

Here is the decision tree, and you have seen this a lot today. I am going to talk about this not in the context of examples--we have heard some excellent examples, but in the context of where it might be working and where it might be tweaking.

[Slide]

I want to do it by a general example. I am not going to call this drug X but let's call it an acid-modifying drug. The goal that we had to study this drug was to look at it in children 1-12 months of age. The question is how would you do it or how would most people do it? Well, we would look at what is available and then we would make a stab at several things.

First we might select otherwise healthy infants who are being treated with acid-modifying drugs, children who are not severely handicapped, who don't have renal failure or hepatic compromise but kids who are getting these medicines anyway. We would use known PK and PD properties of the drug plus evidence that demonstrates the impact of ontogeny on the clearance pathways or drug metabolizing enzymes and in some cases even the

effect, much as we heard for the montelukast story. There was a pretty good relationship in the adults between the improvement in FEV1 and the exposure. We would use robust, minimal sampling techniques when appropriate. We would assess the pharmacologic effect of the drug if possible; design effect studies with a target exposure-response approach to drive the selection of dose as we looked at effect; and then assess the effect of the drug as a molecule as well a treatment effect and tolerability in an age appropriate manner.

To get back to the montelukast story for just a minute, I think it is incredible that approval and labeling for that drug was done based upon changes in FEV1 that many of us would sneeze at as being important. But the fact is when it is given to children with asthma and you look at its anti-inflammatory effect and you look at long-term outcome, it is a medicine that works. In that case we made a good leap of faith and it is possible to do that.

[Slide]

Those of you at the agency, please don't take this personally. I am going to share some of

the things that were recommended for study our acid-modifying drug from the agency, and we all know that the FDA is a big, big organization and certainly none of the people associated with Dr. Lesko would ever recommend what I am going to show you today.

I put a little asterisk here because I have to give the disclaimer, and rightfully so, that the recommendations that are coming out from the FDA about how to do these studies are an evolving work in progress. But let's look at a few things that were recommended.

First, the primary disease endpoints. To assess the efficacy of this drug in infants, we were told to look at its effect on obstructive apnea. Some of you have a somewhat confused look on your face. I still have one on mine.

Secondary endpoints, to look at pH of the stomach. That makes sense for an acid-modifying drug, but then to assess its effect on esophageal motility. We were asked to do single and multiple dose kinetics standard sampling through 24 hours with a drug that has a half-life of one hour.

We were asked to study two to three different fixed doses of the drug. We were asked

to look at the kinetics and safety of the drug in neonatal mice and p53 knockout mice and then, in the infant studies to follow the children up through adolescence.

These are all things that at some point or another came out in the recommendations. Fortunately, these didn't stick--these didn't stick. We are finally getting our way to do this correctly. But why do I show you this horror story? It is not to make light of the agency, but when these recommendations came out I can tell you, from working with the sponsors, it was almost as if their head was put in a vice and they began to think how in the world could we do these studies; should we do these studies? Are they even in some cases ethically defensible to do--esophageal impedance in an otherwise health two-month old child? What parent would agree to have that done? So, there were a lot of issues.

[Slide]

Sometimes it is good to look at mistakes that might be made because is lets us improve what we might do. In this case, I have to admit it really is not the usual scenario. We know that from what we have heard today. I am picking at

off-the-wall examples to make a point.

The approach, if we look at this example, the approach now becomes not a solution but an impediment to pediatric drug development because of slippage in the regulations and their interpretation. How is that so?

If we look at the exclusivity provisions under the Best Pharmaceuticals Act which still brings a lot of marketed products to study in pediatrics, they enable labeling only if the disease process is substantially similar, the disease process. Now, every company that studies the drug, I can guarantee they are interested in labeling. There is a belief by some that dosing and safety information is not wholly sufficient for exclusivity or pediatric labeling but in every instance in pediatric a pivotal phase 3 study is necessary. That is not what the regulations say but there is enough slippage in the regulations to allow this interpretation to be propagated in the course of discourse between the sponsor and the agency.

Granting of exclusivity is increasingly viewed as a privilege and there is a control on it. About 25 percent of issued written requests for

pediatric studies have resulted in exclusivity. We are not breaking the bank with it. There is differential interpretation of the regulations by what I have termed the "Tower of Review Divisions." I can tell you that the review divisions that looked at montelukast took a very different approach than the review division that looked at sotalol and the review division that looked at the acid-modifying drug. So, there is not uniformity of interpretation across the board.

Problems and in some instances failures with regard to integration of both the Pediatric Division at FDA and Clinical Pharmacology with what the review divisions do. Much of the discussion this morning at the end-of-phase-2A, to me, goes toward solving some of this problem. Then, the entire pediatric initiative clearly largely remains an unfunded mandate. So, there are some problems that exist that turn into decision-making.

[Slide]

Let's go back to the decision tree for just a minute. You have seen it and I am going to modify it just slightly by getting rid of the first two things in the top box. Let me explain why I am trashing the top box.

[Slide]

If you look in pediatrics, from what I have been able to learn in the few years of dealing with it, is that in most instances the disease process is rarely substantially similar to adults. It is rarely similar with respect to onset, progression, expression of symptoms, and the disease environment-treatment interface. There are many, many differences. So, it becomes an interpretation issue to say is it similar or is it not, and I think we heard that with the last presentation. When you get down to the end of the day with numbers and you say is this a meaningful difference between these two populations, we ask the medical officers is it really different.

Now, what many people have shown is similar is the relationship between the concentration of the drug and the effect of the drug. It is often similar between adults and children. That is not to say that develop doesn't influence receptor expression certainly in the first few months of life but beyond that it is pretty much the same.

[Slide]

Ergo, here is what the decision tree might

look like in my mind. In the top box we have similar drug effect or mechanism of action. Is there similar concentration effect or is there similar effector response? This moves it away from disease and squarely puts it into issues regarding the clinical pharmacology of the drug. Once you satisfy a couple of those you march down, and march down in such a way as to determine tolerability and what is the right dose.

[Slide]

So, the "holy grail" of extrapolation, as I see it, is forget about the disease being substantially similar because in many cases it won't be. Focus on the drug response being similar. That is what clinical pharmacology does best. Again, in many cases this notion of a morbid-mortal outcome for studies because that is just not the way it is done. But base the assessment on drug efficacy and tolerability associated with similar--I didn't say equivalent but similar exposure. Then, mandate the use of a decision tree that is driven by the Exposure-Response Guidance, something that really lets us look to see if similarity exists. When that is done and it is woven together, like this

picture of an Indian blanket, it becomes not only a thing of great beauty but something of great function and potential significance.

[Slide]

But to do it we have to improve what we do in development, and it is real simple because if you think about it like Einstein did, which is to think out of the box and much of our discussion today has been about thinking out of the box, the problems and the challenges of pediatrics, many of which are insurmountable, we are always going to have small numbers, we are always going to be dealing with what you can do and what you can't do, what you shouldn't do, but if we apply the best that technology has to offer we can make effective solutions, and I think that is my last slide.

DR. VENITZ: Thank you, Greg. Any questions for Dr. Kearns? Larry?

DR. LESKO: Just a terminology question, Greg, what do you mean by tolerability in one of those boxes that you modified?

DR. KEARNS: That is my way, Larry, of saying that we never truly get safety data from any of the pediatric things that we do. For most of them that have less than 100 subjects, it is only

tolerance data.

DR. LESKO: Then, just to understand your point in the first box where you are suggesting to drive it by exposure response primarily, is that by demonstration with data that one would get during the drug development process?

DR. KEARNS: Yes. That was actually done in the pediatric labeling of famotidine by Merck where in a limited number of children and infants we were able to measure intragastric pH, calculate EC50, Emax, the pharmacodynamic parameters, compare those to the parameters in adults and we found that there was no difference. Then the approach that was used for the labeling of famotidine was one driven by exposure response and kinetics.

DR. LESKO: So, the assumption kind of is that we need to have response correlates. In other words, there is going to be a subset that do and a whole bunch of drugs that don't.

DR. KEARNS: But it is even possible I think to--one of the early pediatric studies, one of the early drugs that had some labeling was Tegretol, carbamazepine. Those studies on response were done using in vitro systems to show that the concentration-effect response of Tegretol on the

gating I think of sodium was similar to what it was in adults. But we have moved far afield of that now in terms of our thinking about pediatrics and I am saying if there are relevant approaches that come from animals or in vitro that deal with effect, that should be something to look at.

DR. FLOCKHART: Greg, I guess this is the pediatric internal medicine conversations. So, first of all, I totally agree with you that we to think a lot more carefully about the differences in disease progression and so on, but I would like to explore with you what some of those might be, just to flesh out some good examples.

Now, the first thing that strikes me is that the diseases aren't actually the same. You know, adults get high blood pressure and kids don't much. On the other extreme, you know, asthma would seem to be, to a very naive internist, not terribly different. The kinds of drugs we use in kids tend to be similar and that we be representative of a group of diseases where we have been somewhat successful in transferring adult methodologies--well, not methodologies but PK/PD relationships to kids.

This begs the question of the vast

untouched swath of disease where it is not similar. So, could you talk a little bit about what that might be. What would be diseases where there are very substantial differences that we might expect?

DR. KEARNS: Well, let me use asthma as an example. Yes, it is similar from the standpoint of what the symptoms are; that anti-inflammatory medicine is something good for all asthmatics. But if you look at the impact of development on remodeling of the airways, it is much different in a young infant than it is in an adult. If that has something to do with the long-term outcome of treatment in terms of morbidity and mortality, there could be very, very important things.

The other side of the coin is the acid-modifying drugs. Again, I go back to the example. For adults, probably 30 percent of adults in the room here today have some proton pump inhibitors in their kit. Certainly I d. They work; they work. They are given to infants not because infants have gastroesophageal reflux disease, not because there are many infants running around with Barrett's esophagus. They are given to infants who throw up and are unhappy when that occurs because of the acidic gastric content that

is thrust into their esophagus. So, if you can make that better, the baby still spits up but the kid is a lot happier and that is why the drugs are used.

Now, that may seem like a lame reason if you are a regulator, but it is the context of use. So, at the end of the day acid-modifying drugs, if you look at the proton pump and all the studies, or you look at H2 antagonists, they seem to work with the same concentration-effect relationship in babies that are a month old as they do in adults who are 40 years old. A lot of the disease stuff from a scientific perspective has not been well explored.

DR. VENITZ: Any other questions?

[No response]

Thank you, Greg. Our next presentation is by Dr. Rodriguez. He is going to talk about the regulatory experience with the very same decision tree that we just talked about.

Regulatory Experience in Using the Pediatric Decision Tree

DR. RODRIGUEZ: I am a pediatrician; I am not a pharmacologist so obviously what you are going to hear is from the perspective of a

pediatrician who is, however, as interested as we all are in the appropriate, number one, use of the drugs and the observation of effectiveness and the safety or tolerability depending where we end today or in the future.

[Slide]

This is one of the reasons why I am doing some of this stuff. We are starting here a few years ago with some of my grandchildren. The reason I do that is because my children used to complain all the time that I didn't pay much attention to them; I was too much at work or in the hospital, whatever, so now I spend more time with them and, therefore, I have them there as a reminder. But specifically they are the ones who are going to get the drugs that are studied appropriately and that is why I put them at the beginning and I put them at the end too.

[Slide]

It is interesting because the issue of pediatric labeling has been around for quite a number of years and, of course, Greg mentioned Jacobi's commentaries and, in fact, in 1979 there was a statement which I will read to you: statements on pediatric use of a drug for an

indication approved for adults must be based on substantial evidence derived from adequate and well-controlled studies unless a requirement is waived. So, that is a little thing on the side. That was in 1979.

From there we progressed to 1994 where we had probably the first almost legalization of the extrapolation. Essentially, we were allowing people to infer or estimate by projecting or extending known information in the field of pediatric drug therapy.

[Slide]

This '94 rule required the sponsors of marketed products to review existing data and submit appropriate labeling supplements. Do you know how many came in? Very few. Anyway, it applied to drugs and biologics and pediatric applications could be based or may be based on adequate and well-controlled trials in adults with other information supporting the pediatric use. Here we are talking about PK and safety data. However, there was no requirement to perform new studies in pediatrics and, in fact, some drugs have actually been labeled from information that is out in the literature essentially, and that could be

one way to look at it if the studies were well done.

[Slide]

The efficacy could be extrapolated in the '94 rule if the course of the disease and effects of the drugs, beneficial and adverse, are sufficiently similar in pediatric and adult population and, therefore, it would be permissible to extrapolate the adult efficacy data to the pediatric patient. So, sufficiently similar is a little bit more open than substantially similar. It is what the '79 rule was talking about.

[Slide]

Other supporting information included information which would be appropriate for the pediatric rule which supports use in that age group and minimum PK and safety data must be obtained. I am not wording this; I am actually getting it out of the regulation. However, if the PK parameters are not well correlated with activity in adults, a clinical study would more likely be requested.

[Slide]

So, an approach based only on PK is likely to be insufficient when blood levels are known or expected not to correspond with efficacy or, for

example, when there is concern that the concentration-response relationship varies with age, and we have heard about that today, and in such situations there is need for studies of clinical or pharmacologic effects. If the comparability of the disease and outcome of therapy are similar but appropriate blood levels are not clear, a combined measurement PK/PD approach may be possible.

[Slide]

So, today what I would like to do, among other things is, first of all, share something that we did within the agency where we actually got people together from various divisions and looked at drugs that were actually being studied or have been studied in response to written requests. I want to share that information with you because it might actually help us identify areas where there are problems and areas where we are likely to fail.

Where may extrapolation not be the right approach? For example, adult efficacy cannot be extrapolated or the response of drug may differ because of receptor differences or the disease manifestations may be different.

Difficulties may be posed also by the

child's inability to cooperate. You have heard about some of the pulmonary drugs today. Essentially, if you are trying to measure the effect of something used in a spacer, the four or five-year old kid may not be able to help you or may not be willing to cooperate in the carrying out of an FEV1 evaluation, although people have gotten strong enough to say if you take some of these young kids and you squeeze their chest real hard you will be able to find out some of the response, and it has been done, by the way, in the younger population but we are not pushing for that.

[Slide]

The extrapolation may not be the approach if the disease is different in etiology, pathophysiology and/or manifestations. There are some pretty good examples particularly in the area of psychopharm., such as neonatal seizures, infantile spasms and febrile seizures. Therefore, in those situations you would expect that there would be nothing to extrapolate from or that the therapy might be different. Antiepileptic drugs effective in adults may actually be ineffective proconvulsants in children, such as phenytoin and carbamazepine which may exacerbate certain

pediatric types; or vigabatrin, which is not approved in the U.S.A., and may exacerbate myoclonic seizures; or we may find drugs that are ineffective in adults but therapeutic in children, like ACTH and steroids in infantile spasms.

So, we have another way and that is important to keep in mind because if we sit around waiting for extrapolation we may actually not study drugs that could actually be useful in the pediatric population.

The pathophysiology may be comparable but the response to therapy may not be predictable in adults and children. This happens with many of the psychotropic agents. In fact, CDER had a program last week in the area of the use of extrapolation and the various divisions came that we invited. Essentially, some of the areas from pulmonary, etc. were actually discussed. And interesting one was drugs for allergic rhinitis where in the physiologic area the pathophysiology was understood and, therefore, the drug was approved for use in the pediatric population, whereas neuropharm. felt very uncomfortable in extending that type of process in some of their products.

[Slide]

The favorable scenarios where it may be okay to extrapolate are, for example, if the drug has been effective in adults and in children down to six years of age. You have heard about one exercise in which they went under that age group. In order to extend the labeling down to one month you must establish that the disease is similar; response to treatment is similar; plasma levels of drug dosing is in the therapeutic range; and the safety profile is acceptable--essentially what you have been talking about today.

There are some areas in which extrapolation has generally been very appropriate. That happens to be one of my areas of expertise, essentially antimicrobial and antiviral. I am an infectious diseases pediatric specialist. You heard about bronchodilators. In fact, in AIDS it is fascinating because there, even though the disease may actually differ in terms of the progress, the markers, for example, are looking at something as the viral effect of the drug and also looking at some of the markers like CD4 were actually used to approve drugs for use in the pediatric age. So, essentially, in some areas of the agency some of the stuff we are talking about

today has been used rather readily.

[Slide]

What I have in this slide is actually what this multi-disciplinary group actually said how about if we were to consider extrapolation in children to support the efficacy data. What would we actually be looking at? We looked at the nature of the evidence, such as empirical comparison; knowledge of mechanisms; known adult physiologic and clinical properties of the analogous drugs; known sensitivity of children to specific toxicities.

And, how do we get there? Let me give you a little bit of background. These were actually 35 drugs that had been turned into the institution in response to written requests. They are drugs that have been granted exclusivity, etc. The reason I am telling you this is because I want you to see that in order to get exclusivity you may not have to show that your study showed efficacy. However, you have to follow what the agency actually asks you and I will show you an example about that.

So, how do we get there? Well, non-clinical studies--I was very glad to hear that people might take a look at cell lines for example;

they might take a look at animal studies; they might take a look at patient samples. In fact, somebody was talking the other day about use of tissues from a brain that had undergone surgery for whatever reason, and looking to see how the drug acted in there. Looking at the pathophysiology, in other words, similar clinical and symptom markers in adults and children or the involved cell types; similar natural history in an affected population. Essentially, the continuity across age spans may be helpful, and similarity of response to therapy such as improvement in the same clinical signs and symptoms for example.

I have not been exhaustive there. There are quite a number of other factors that we have in there. But we felt that an evaluation of some degree of safety is essential. Granted, when we thought about safety in adult studies we have thought sometimes of 300-plus patients in a study essentially to pick up a signal that may actually be at a relatively high level, let alone the ones that are at a very low level. But if you take a look at the process of drug approval, you see the word safety used in phase 1, phase 2 and phase 3. Again, this has to be supported with

pharmacokinetic and exposure response.

[Slide]

I actually went to the regulation of '94 and said let me take a look and see how this really fits into the decision tree. Essentially, we can see that the first column would probably not fit into the decision tree and essentially there we have to include in pediatric use or limitations or pediatric indications, for example, the difference between pediatric and adult responses for the drug and other information related to the safe and effective pediatric use of the drug. We could be using the same example of ACTH and steroids in the issue of infantile spasms.

We move down the line and we look at pediatric use for the indications also approved for adults and the simple product that came to my mind was actually the use of drugs for inflammatory response in the eye or infection in the eye. We could conceivably say that in those situations we don't need to really get PK/PD. We are actually specifically looking at the response and could use the data from adults to specifically say that we would not need two well-controlled studies and we might be able to get away with one.

Of course, in the third row we have essentially the closest thing to the decision tree, which is indications based on raw data and that is where we are talking about use of the well-controlled information supporting pediatric use. In that situation, again, we still have to note that the course of the disease and effect of drug, both beneficial and adverse, are sufficiently similar in adult and pediatric populations to permit extrapolation. Again, we have to spell out the indications for that.

Essentially, I am not going to spend much time with this, I know that in April of this year Dr. Rosemary Roberts spent quite a bit of time going into the various drugs that fit into this tree and what I decided to do was to essentially show you--

[Slide]

I am sorry, before I go there, for all these drugs that we want to study we ask the following questions: What is the public health benefit for using the product in children? What is it? For what ages? What information is needed? What other products are available or approved for this indication? And, what type of studies are

being done or should be conducted?

[Slide]

Essentially, what I am going to show you over here is information which is as up to date as of September 3 and we essentially looked at the studies that were requested for written request in response first to FDAMA and then BPCA. You can see that 284 written requests were issued. Now, 93 written reports have come back to the agency as of September, by the way. Of those, 60 have already been labeled, which is quite a bit of progress. And, 85 have been granted exclusivity, which means that only 9 studies did not get exclusivity, and they didn't get exclusivity because they weren't providing or they haven't provided the information that they had agreed to provide in the report.

I think Dr. Lesko showed you something earlier, showing the percentage for efficacy and safety, PK and safety, and you can see it has changed very little over the period. You could argue, well, we haven't changed anything or we are getting the information that we need to go forward. So, there are two ways to interpret that.

[Slide]

Now I would like to share with you some

experiences and these experiences came from this group that was put together to look at drugs that have been granted exclusivity, have been labeled and have provided some type of information.

[Slide]

The first one that we have here is the psychotropics. I have selected the psychotropics because that is where we had the biggest problem in thinking about the way that the decision tree would help us.

Essentially, for this drug, over here, there was absence of prior data, according to the division, that would allow extrapolation. So, they actually went ahead. Our group went ahead and said, okay, what factors could be used for extrapolation? Essentially, we felt that there was similarity of symptoms in children at least over six years of age. We felt that the response to therapy would probably be similar and so would the natural history. Essentially, the division asked for multicenter, randomized, double-blind, placebo-controlled studies to evaluate efficacy and safety, and PK open-labeled escalation.

Let me tell you that there were well over 500 patients, almost 600 patients enrolled in

these. What did we come out with? Safety and effectiveness was not established in patients 6-17 years at doses recommended for use in adults. PK parameters, area under the curve and Cmax of drug was found to be equal to or higher in children and adolescents than in adults. Maybe in the future something like this may actually benefit from some of the stuff that we are talking about today but essentially that is what came. Let me tell you that this company did get exclusivity. Why? Because they did everything that was in the written request. So, essentially, that is the criteria for granting exclusivity.

[Slide]

Another example is the psychotropic fluvoxamine. Let me tell you first of all that exclusivity came to the agency on 1/3/00. Remember that these are in response to the FDAMA in 1997-98. So, within a couple of years we had this area on our hands. This was for obsessive-compulsive disorder. Essentially, again the group said similarity of symptoms and response to therapy would be areas where extrapolation could be done. There was a multicenter, open-label PK study and long-term open-label safety study.

The result was that, number one, we already had an efficacy study of this drug at the time this drug came to us. It was actually in the label but there were questions about why aren't we having some effect in the adolescents? Why do we seem to be having more effect in the girls or in the children 8-11 years of age with the doses that were recommended in the label?

To make a long story short, nonlinear pharmacokinetics was a part of the answer to this, and this was corrected and essentially girls 8-11 years of age may require a lower dose while the adolescent may require doses to be adjusted to actually be increased over what they were constantly getting.

[Slide]

Essentially, we are learning and we could learn more. This is gabapentin, an antiepileptic. Actually, that came to the agency on 2/2/00 and, again, it was labeled by October of that year. The concerns with respect to this drug were that safety and efficacy could not be extrapolated. Remember, this is in the psychopharm. group again where they have had some of the bigger problems for extrapolation.

But our group said that they could extrapolate on the basis of similarity of symptoms and response to therapy. Essentially, they actually did a double-blind, placebo-controlled, parallel group efficacy and safety study as add-on therapy; population PK; open-label extension study and single-dose PK. There were quite a few patients that were studied there, almost 1,000 patients.

[Slide]

The results were there was safety and effectiveness down to 3 years, however, we identified some neuropsychiatric disorders in 3-12 years old such as emotional lability with attention problems in school and hyperkinesis. The product clearance, normalized by body weight, increased in children less than 5 years of age. So, between 3-5 higher doses were required in that population.

[Slide]

The next two drugs were in the cardiovascular group. Again, there were some problems in the area of extrapolation. Essentially we have here hypertension. The thought was there was similarity in symptoms and that the natural history was similar. We have to remember that

hypertension in kids may actually be the result of structural abnormalities for example which may differ from the adult population.

There was an open-label PK study, double-blind dose-response study. The result was that the drug was labeled for one month to 16 years of age, and there was information on dose efficacy and pharmacokinetics and, more beautiful, there was information on preparation of a suspension. So, essentially, we had good information that actually made it into the label.

Let me just add here that we had at least two situations where there has been information on a suspension and five situations of the first 34 drugs that were approved where we had new formulations made for use in the pediatric population.

[Slide]

Here we have the last one that I want to share with you, which is fosinopril. Essentially, that drug came in on 1/27/03. The indication was hypertension. Essentially, areas that could actually be used for extrapolation were similarity in symptoms and the natural history. Essentially, there were open-label studies, multicenter,

single-dose PK studies were requested in one month to 16 years of age; multicenter, randomized, double-blind dose ranging and placebo-controlled studies in 6-16 years of age.

The results are as follows: New recommendation for dose in children weighing more than 50 kg; new information on PK parameters and appropriate dose strength is not available for children weighing less than 50 kg. The company did not come in with a formulation or with a preparation for suspension and even though data is available, that was not included in the label at this moment. Essentially, you can see that this is a two-way street.

[Slide]

So, what have we learned from the point of view of pharmacokinetics and pharmacodynamics? Some populations may need to start therapy at the lower end of dosing to avoid adverse events. That was for midazolam hydrochloride in patients with congenital heart disease and pulmonary hypertension.

Elimination half-life may be shorter in pediatric patients than in adults. That was in atovaquone/proguanil. Essentially what we saw is

that atovaquone clearance in children was 1-2 days--I am sorry, the half-life, not the clearance. The volume of distribution and half-life may differ in a fashion which necessitates doses higher in younger children than adults. That happened with etodolac.

[Slide]

Higher oral clearance by body weight in patients less than five years of age necessitated higher dose concerning gabapentin. You have already gone extensively over sotalol hydrochloride. Buspirone hydrochloride from kinetic parameters, area under the curve and maximum concentration of the drug may be equal to or higher in children and adolescents than in adults, and no demonstrated efficacy. As I mentioned earlier, in fluvoxamine there were nonlinear pharmacokinetics.

[Slide]

So, what are the gaps in information? There are many but I have selected three. Many populations such as infants and neonates, both term and pre-term, remain to be studied. There is still a lot to be learned in terms of clear exposure-response relationship across the various

special populations. Very importantly, it is very hard to meet these criteria in some of the drugs and essentially try to find appropriate pediatric formulations. But if somebody comes home with a correct formulation the agency is ready to look at it favorably.

[Slide]

This is the end of my comments and I am open to questions and if I don't know, I will communicate with you later.

DR. VENITZ: Any questions?

DR. FLOCKHART: Well, I would like to thank you too. I think this was really tremendously valuable to me in terms of my thinking about this from many respects.

I would like to ask you about two kinds of studies you presented. The first is the hypertension ones. I am an internist. Hypertension in children or adolescents, to me, is different in that it is rarely what I would call essential hypertension. As you indicated, it is much more neurofibromatosis induced or one of those things. So, are the studies that you are talking about ruling those out because they would be separately treated? And, you are essentially

dealing with essential hypertension in children which would be a very, very narrow group of patients.

DR. RODRIGUEZ: These studies, in response to written requests on which a protocol was developed, would specify clearly the diagnostic criteria by which the patients would be enrolled in the study. In other words, it was not all hypertension. It was stenosis for example.

DR. FLOCKHART: Right. The second question, you mentioned specific liabilities that children might have to side effects. What about actually testing side effects? I am interested particularly in the situation with HIV drugs--side effects that might occur more in adults, something like lipodystrophy, and less in children? Has that been the case also?

DR. RODRIGUEZ: To the best of my knowledge, no, but I am not sure. So, if you want I will give you my e-mail and we can communicate.

DR. FLOCKHART: Sure.

DR. KEARNS: Bill, that was a great talk, as usual. My question is based on the examples that you showed of the drugs recently studied, almost all of them had some type of efficacy study

associated with them. You showed the earlier regulations and went back to 29 CFR, dot, dot, dot. The third point that you made is that if pediatric use was based on adult data, then it could be the case were appropriate dose-finding safety studies could be done, which is very much part of the pediatric decision tree but, yet, your examples all deal with an efficacy study and in some cases with some of the psychoactive drugs it has been debated that those efficacy studies were probably under-powered to really assess an effect because the things measured in children are sometimes very difficult. So, if most or all of these are going to involve efficacy studies do we need to redo the decision tree that has the first box immediately going to an efficacy study?

DR. RODRIGUEZ: I thought I had said that but I will repeat it, one of the reasons I selected these drugs is because these were the drugs that we actually had some problems with, and these are two divisions, for example, that have had some problems--not problems, I should say maybe different mechanisms, I mean the psychopharm. drugs for example. So, essentially what I did was I selected the ones where the problems were because I

figured there were enough people here that might come up with some suggestions on how we can deal with that.

You raise a point. It might be the power. But when you hear about 500-plus kids, that is a pretty good sized study. In fact, one of the things I said was maybe those kids needed higher doses and that was my naive way to look at it. Anyway, I selected the problems on purpose. But if you look at the breakdown of the various requests, a lot of the drugs did not necessarily require efficacy. They had the PK/PD and, of course, they had safety.

DR. LESKO: To follow on the question that Greg raised, Bill, in the type of study, that is the study breakdown on the issue of written request, there are 284 or 660 studies, it looks like, and there is a percentage. In the written requests only 35 percent--getting back to what Greg asked--are efficacy studies, although for the ones you showed in the area of the antihypertensives and the psychotherapeutic agents it was 100 percent efficacy.

There are two questions. Of the 93 that you said came in, and you said 60 have been

labeled, does the percentage in terms of the type of study remain the same as it is for the written requests?

DR. RODRIGUEZ: I have that tabulation on the first 33 drugs that were labeled. That is over 50 percent of the drugs that have been labeled. We published this in JAMA.

DR. LESKO: Okay.

DR. RODRIGUEZ: There we have around 43 percent efficacy and safety; 34 percent PK/PD; and 12 percent were combination where the topics were actually safety.

DR. LESKO: So, it sounds like it is kind of similar in terms of what actually is done in studies as opposed to what is put in a written request.

DR. RODRIGUEZ: But if you take a look at that, we have almost 56 percent that were PK, safety; PK/PD and safety and 43 percent that were efficacy, safety.

DR. LESKO: Just continuing with that, can you think of several therapeutic classes--we know where efficacy studies predominate, for example, in the antihypertensive and psychotherapeutic agents, were, on the other hand, approvals based not on

efficacy studies but on other information, the PK, safety or the PK/PD--

DR. RODRIGUEZ: Well, you heard about the pulmonary allergy type reactions. That has been one where there has been a mix of drugs where some biomarker or some other finding has been used for that.

DR. FLOCKHART: HIV with a CD4 count.

DR. RODRIGUEZ: HIV with CD4, that is right. You see, the area where it is relatively easier is in the infectious diseases because if you draw a triangle and you put the human over here, you put the drug over here and you put the virus or the bacteria over there, you can do--I mean, we do a lot of things in vitro which adds validity. In fact, even there, there is a problem because, you see, when you approve drugs for viruses you approve drugs for viruses. When we approve drugs for bacteria we are sometimes approving them for otitis media or sinusitis or pneumonia even though, for example, in H. flu it would be H. flu or strep. pneumo., strep. pneumo., strep. pneumo. but we are applying it for the various clinical indications. But in the virology field it is easier because for some reason that rationale has actually prevailed.

I wouldn't be surprised if we progressed toward that direction. I am speaking off the top of my head right now.

DR. VENITZ: Any other questions? If not, thank you.

DR. RODRIGUEZ: You are welcome.

Committee Discussion

DR. VENITZ: Larry, I would ask you to put your last slide up so we can go through the three questions that you want us to give you some feedback on.

DR. LESKO: I actually don't have one. I don't have a slide on the questions but they are in the background package and maybe we can refer to that because there are only really two questions. One of the questions refers to the methods of analysis that Dr. Machado showed us in terms of determining similarity and exposure response between adults and pediatrics, and we did have some discussion of that already.

However, the second question really revolved around providing some feedback on the current way the pediatric decision tree is being used in the context of the numerous examples that were presented today. In other words, does this

seem like it is on the right track?

Furthermore, some suggestions were made that maybe there is room for other approaches than what we have in the pediatric decision tree based on what Dr. Kearns presented. Are there comments on potential alternative ways of thinking about, in particular, that first box? I think if we can sort of go in that area for discussion it would be helpful.

Maybe rephrasing the question, if we think of the current pediatric decision tree as the current situation, in essence a one-size-fits-all because that is the decision tree, are there any situations where a different approach might work, similar to what Greg had suggested, to approach it and drive it from an exposure-response mechanism of action point of view? For example, could that be an approach that would work well in areas of drugs that are well understood in terms of their mechanism of action, drugs which might be a third in class for example, a drug with a wide therapeutic index where pharmacodynamic endpoints are reasonably measured and are thought to correlate not as surrogate endpoints but with clinical endpoints? And, given certain criteria,

could an alternative approach be used to go down that decision tree? So, that is kind of an area that I would like to maybe hear about as well from the committee.

DR. KEARNS: Larry, I think one thing I would like to add to this, and Bill's talk alluded to it, is that the pharmacodynamic endpoints that are measured have to be appropriate so things can be done in children, and they must relate to the effect of the medicine. That is easier said than done. I mean, psychometric testing in young children is not an easy thing.

What happens sometimes is that in the course of pediatric drug development and trying to satisfy the questions we are faced with, almost being forced out of necessity or in some cases desire--and that is my impression, to develop endpoints in the context of the trial, none of which are validated and in some cases the endpoints have nothing to do with effect. Again, case in point, an acid-modifying drug doesn't influence esophageal motility. So, as long as we are basing what we do on the clinical pharmacology of the drug and doing the best we can, I think we get the best approach and at the end of the day the best answer.

DR. SHEINER: The example you used, the acid-modifying drug, that is a tough one. What you are saying is, look, it is getting rid of the acid and when the kid spits up it makes him happier and there is no equivalent adult disease per se. So, you are saying that here is an indication that doesn't exist in the adults, treated by the same mechanism as something that does.

If you find that the physiology is the same, the acid is turned off at the same concentrations, lasts as long, and everything like that, first of all I have a question, doesn't the indication have to be approved? Maybe your drug has some safety consideration that would make it approvable for something that was life-threatening but not something that as symptomatic, etc. I mean, I just don't see how you are going to be able to automatically find that because the physiology is the same after the drug, that because the indication is different you get approval in pediatrics. You wouldn't get it in adults. If it turned out that there was a new condition that was treatable--I mean off-label use is fine because the drug is approved but for approval you would have to show that it is efficacious in that condition.

DR. KEARNS: A good question. Again, my impression and I am not speaking here for the agency, but I referred to some of the slippage in interpretation. Children per se, young infants especially, do not characteristically have gastroesophageal reflux disease. Histologically many of them are normal or they may have a little bit of hyperemia but it is not the same thing in adults. Well, if we interpret that as saying, oh, well, that is a different indication, then as you interpret the regulations you could certainly go down and say, okay, we have to do efficacy studies of these drugs. So, you interpret the regulation. But if you went back to 29 CFR dot, dot, dot, and you read if pediatric use is based on adult data, and proton pump inhibitor use in pediatrics is based on adult data, and the data it is based on is the ability of the drug to modify the pH of the gastric content, not anything else.

So, there is a tremendous amount of interpretation that has to go on and that is why I said earlier it is imperative that the Office of Clinical Pharmacology and Biopharmaceutics be involved early and, hence the decision tree. Be involved early and try to work cooperatively and

collaboratively with the review divisions to make sure that the studies that we think we need in kids are done and that they are done right because some things in children you just can't do. Parents will not volunteer for repeat endoscopies in young infants and, arguably, they shouldn't be done because of the risks associated with anesthesia and stuff like that. So, we can't use the old adult ways to do the pediatric studies. But it is hard. There is room for slippage.

DR. SHEINER: But I think there are two issues there. You know, all my sympathies are with you. My guess is that you are saying is that modifying the acid production is going to help condition X whether it is adults or children, and what I have is approval of things that modify the acid production for condition Y. So why not? And there will be plenty of off-label usage of that and it may never-ever come to the FDA because they can sell it for that. We know lots of drugs where a given action turns out to be good for something else and people use it for that.

But if you want, you know, the "Westinghouse seal of approval," you have to show it for that indication. That is the rule. I am

not saying it is right. Therefore, this is not a pediatric problem; this is a general problem of discovering that a given action of a drug is useful for another indication and whether or not you can get the FDA to say, well okay, if you think so--it just doesn't do that, I don't think.

DR. KEARNS: Well, one of the worries has been the concern that if you put information in the label, if you put PK or PD information in the label absent information that proved efficacy in a condition, the label would then foster additional off-label use of the drug in children. You know, I think that is a little bit laughable because historically pediatricians have not been inhibited at all from using drugs off-label. They won't be compelled by that issue in the future, but what is helpful for many people is to know that if they gave a dose of X it would make exposure Y which was similar to that in adults. Then at the end of the day the medical practitioner has to make the decision whether he or she will utilize a medicine.

I don't have any trouble with labeling saying that this drug has not been evaluated in children and its efficacy is not known. I think that is okay because I am willing to use other

information to make the decision. But in an environment that is indication driven where the indications in adults and kids can be very different, it could set us back a little bit and the decision tree, if done right, can fix a lot of that.

DR. SHEINER: I won't get the last word in because I know you but--

[Laughter]

--one more time, the thing is that what you would have to say is that this has not been shown empirically to be safe and effective for this indication. That doesn't mean it isn't, it just hasn't been shown. The mismatch between what is approved for children and what is used in children--I think the attempt of the flow chart is to get close to that. But I think what you are saying is that in the end it is only going to get us part of the way there, and how should we deal with the rest of the way because it would be nice for the public to be reassured at some level that what the pediatricians are doing has been inspected to some degree. But I am not sure that we want to mix that with the issue here.

They have bitten off an easier part, the

same indication, and now can we establish that the concentration response is the same for the same indication, and then we can just approve with the PK, or something like that. That is an easier problem. Let's get that one all straight and then let's move on. As I say, I am totally sympathetic.

DR. KEARNS: And I appreciate that more than you know. The same indication and the same use is oftentimes different and that is the problem. If you look at the labeled indication for many of the acid-modifying drugs, it is to treat nocturnal heartburn associated with symptomatic GIRD in adults. That is nutty. You know, that is really nutty. But we use drugs in pediatrics for the same reasons. Whether it is hypertension, asthma the same target, the same therapeutic target is there so I appreciate your words and I will stop talking now.

DR. VENITZ: Larry, maybe just one comment, you are looking for scenarios where it is likely to use the currently modified decision tree, acute indications, symptomatic indications. You may be more likely to use pharmacology-driven approval/labeling rather than chronic indications.

DR. LESKO: It would seem like that would

have to be the case in the sense that it is the effect that you would measure early on in this decision tree. Thinking of the alternative or the pharmacological effect in an acute condition, I would expect that would be fairly close to the clinical endpoint in the sort of chain of events. As in Greg's example, you have a modifying of the acid secretion in the gastric pH and then there is an immediate benefit from that in the short term and the change in the environment of the stomach would be close to what you want to achieve at the clinical endpoint. It gets a little more complicated in terms of picking on the effect when you move into some of the therapeutic areas that Bill mentioned in the CNS area and the seizure area where you don't have the convenience of the same type of biomarker, if you will.

So, that was why one way I was thinking about this, you know, rather than one-size-fits-all, would be are there alternative decision trees that could be thought about in terms of what we have now and an alternative for those indications where use and indication are somewhat different but there is a close relationship between drug mechanism, marker and endpoint where you could

do something that could rely on less than efficacy studies basically. But that is the open question.

DR. VENITZ: But it might be those drugs as well that allow you to incorporate some of the preclinical information that he was talking about.

DR. LESKO: Of course. I don't know the extent to which that has been done. It makes sense and Bill had a slide on that where he had prior information. It was animal data. I don't know how much of that is relied on in the current situation. I don't have any first-hand experience with that so maybe Bill can answer.

DR. RODRIGUEZ: Without mentioning the drug, there is one drug that has been used off-label in the pediatric population and there have been concerns about some studies that were done in the rodent model. Essentially, the agency right now is actually conducting studies in primates, newborn, juvenile primates. We have already collected the animals, and everything, and the studies are about to start and, hopefully, we will answer the question once and for all. Not only have the animal studies been done but you wonder how applicable they are so you have to be careful about that. So, we are trying to get as

close as we can to the human primate with a non-human primate so we can then actually say, fine, let's forget about it; go forward and label this drug; it is okay.

So, we have to be careful about it but, on the other hand, Phil Sheridan was talking the other day about the tissues that were actually obtained from surgical interventions in patients with seizures and how those tissues were actually in vitro exposed to medications and the effect of the medication was actually being studied there. Of course, we cannot do brain biopsies on everybody so that is the problem there. But, essentially, there could be, again, primate models that could be used. It is expensive but actually in the long-run may be less expensive than the 800 million dollars that were mentioned over here.

DR. VENITZ: Any more comments to question number two?

[No response]

Then let's try to tackle the last question for today.

DR. KEARNS: To answer number three, first get a crystal ball.

[Laughter]

I don't think that we can ever know for sure that adjusting dose and exposure will give us what we want. I think that extrapolation is predicated upon assumptions that are reasonable from the scientific and clinical perspective; that are predicated upon approaches that are well proven and tested and show that they work, and when done by men and women who understand the scenario in which they are to be applied generally do produce good results. At the end of the day as perfection, I don't think we will ever achieve that but we have come a long way. I think the stuff Bill presented is evidence that we have come a long way with the pediatric initiative. I think we can improve it. It is a work in progress. Then we should be expected to deal with the deviations.

Tomorrow we are going to talk about pharmacogenetics and I am looking forward to that, and I can tell you that in doing phase 1 and phase 2 PK work, having pharmacogenetic data in children is very, very important to understand how much of that variability is really associated with age as opposed to a certain polymorphism and an enzyme. But I don't think we will ever reach perfection.

DR. VENITZ: Let me maybe add something

more specific to that. I think in general when we are adjusting doses based on exposure we are talking about exposures to the parent drug. So, I am always worried when I look at drugs that are highly metabolized. Phase one metabolites may be active or have safety issues related with them. So, as a general rule I would be more skeptical about dose adjustments for highly metabolized drugs that form potentially active metabolites, again, just as a way of stratifying risk. So, drugs that are readily eliminated via metabolism, I think adjusting the dose to achieve the same exposure with the intent to achieve the same response makes sense. But if you have a drug that has ten metabolites and three or four of them are known to be active and you don't really know how active relative to the parent, then adjusting the dose just based on parent exposure may not be reasonable.

Any final comments? It looks as if we are all metabolized for today. Everybody is ready to take a break. So, let me conclude our first day's meeting. Let me thank all the speakers and committee members for their valuable input. We will reconvene tomorrow morning, bright-eyed,

bushy-tailed, at 8:30, same place. See you tomorrow.

[Whereupon, at 5:10 p.m., the proceedings were recessed to resume Tuesday, November 18, 2003 at 8:30 a.m.]

- - -