ajh

ATDEPARTMENT OF HEALTH AND HUMAN SERVICES

FOOD AND DRUG ADMINISTRATION

CENTER FOR DRUG EVALUATION AND RESEARCH

CLINICAL PHARMACOLOGY SUBCOMMITTEE OF THE ADVISORY COMMITTEE FOR PHARMACEUTICAL SCIENCE

VOLUME II

Tuesday, November 15, 2005

8:30 a.m.

CDER Advisory Committee Conference Room
5630 Fishers Lane
Rockville, Maryland

MILLER REPORTING CO., INC.
735 8th STREET, S.E.
WASHINGTON, D.C. 20003-2802
(202) 546-6666

PARTICIPANTS

Mimi Phan, Pharm.D., Executive Secretary

ADVISORY COMMITTEE FOR PHARMACEUTICAL SCIENCE MEMBERS (Voting):

Carol Gloff, Ph.D.
Nozer Singpurwalla, Ph.D.

AD HOC SUBCOMMITTEE MEMBERS

Mary V. Relling, Pharm.D., (Chair)
Jeffrey S. Barrett, Ph.D., FCP
Edmund Capparelli, Pharm.D.
David D'Argenio, Ph.D.
Marie Davidian, Ph.D.
David Flockhart, M.D., Ph.D.
Brian F. Gage, M.D., M.Sc.
William Jusko, Ph.D.
Gregory L. Kearns, Pharm.D.
Howard L. McLeod, Pharm.D.
Wolfgang Sadee, Ph.D.

GUEST SPEAKER (NON-VOTING):

Douglas Mayers, M.D.

FDA STAFFS:

Steven Gutman, M.D.
Lawrence Lesko, Ph.D.
Janet Woodcock, 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 Introductions

DR. RELLING: Good morning again. Let's go ahead and get started. We have our transcription machinery here now.

I am Mary Relling. I am at St. Jude Children's Research Hospital in Memphis.

If we could start with you, David.

DR. D'ARGENIO: David D'Argenio, University of Southern California.

DR. CAPPARELLI: Edmund Capparelli, University of California/San Diego.

DR. SADEE: Wolfgang Sadee, Ohio State University.

DR. SINGPURWALLA: Nozer Singpurwalla, George Washington University.

DR. KEARNS: Greg Kearns, Children's Mercy Hospital.

DR. JUSKO: William Jusko, University at Buffalo.

DR. PHAN: Mimi Phan, Executive Secretary.

DR. BARRETT: Jeff Barrett, The Children's Hospital of Philadelphia.

DR. FLOCKHART: Dave Flockhart, Indiana University.

DR. GLOFF: Carol Gloff, Boston University, and an independent consultant.

DR. DAVIDIAN: Marie Davidian, North Carolina State University.

DR. LESKO: Larry Lesko, FDA, Clinical Pharmacology.

DR. WOODCOCK: Janet Woodcock, FDA.

DR. GUTMAN: Steve Gutman, FDA, from the Office of In Vitro Diagnostics.

DR. RELLING: Thank you, everybody.

Now, we are going to have Mimi Phan read the Conflict of Interest Statement for the group.

Conflict of Interest Statement

DR. PHAN: Thank you, Dr. Relling.

The statement of conflict of interest for the meeting of the Clinical Pharmacology Subcommittee meeting of the Advisory Committee for Pharmaceutical Science. Today is November 15, 2005.

Topic 3: An update on the critical path biomarker-surrogate endpoint project, the use of biomarker information in labels to facilitate individualizing pharmacotherapy and the analytical and clinical validation criteria for approving a clinical assay.

The Food and Drug Administration has prepared general matters waivers for the following special

Government employees who are participating in today's meeting of the Clinical Pharmacology Subcommittee of the Advisory Committee for Pharmaceutical Science to discuss and provide comments on an update of the critical path biomarker-surrogate endpoint project; the use of biomarker information in labels to facilitate individualizing pharmacotherapy; and the analytical and clinical and validation criteria for approving a clinical assay.

This meeting is being held at the Center for Drug Evaluation and Research. Waivers for the following doctors:

Dr. Nozer Singpurwalla, Jeffrey Barrett, Edmund Capparelli, David D'Argenio, Marie Davidian, David Flockhart, William Jusko, Gregory Kearns, Howard McLeod, Mary Relling, Wolfgang Sadee, Brian Gage, and Carol Gloff.

Unlike issues before a committee in which a particular product is discussed, issues of broader applicability, such as the topic of today's meeting involve many industrial sponsors and academic institutions.

The committee members have been screened for their financial interests as they may apply to the general topic at hand. Because general topics impact so many institutions, it is not practical to recite all potential conflicts of interest as they may apply to each member.

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

We would also like to disclose that Dr. Douglas Mayers is participating in this meeting as a guest speaker, giving a presentation on the Use of Biomarkers in Clinical Development and Labeling: An Industry Perspective. He is employed by Boehringer Ingelheim Pharmaceuticals, Inc.

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

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

DR. RELLING: I would like to add for the record that two additional members have joined us.

We have Dr. Howard McLeod from Washington University and Dr. Brian Gage, Washington University also.

I am sorry, I need you guys to state that specifically for the record.

DR. McLEOD: Howard McLeod, Washington University, St. Louis.

DR. GAGE Brian Gage, Washington University, St. Louis.

DR. RELLING: I think we are ready to proceed then. It's a pleasure to have Dr. Woodcock, who is going to give us an update on the Critical Path Biomarker-Surrogate Endpoint Project.

Update on the Critical Path Biomarker-Surrogate Endpoint Project

DR. WOODCOCK: Thank you and good morning. This is a follow-up to a discussion I believe that we had with this committee in March on what FDA was going to do to try and develop a more detailed framework, a more specific framework on how to qualify new biomarkers for various uses and surrogate endpoints

I do regret to say we haven't come along as far as we had hoped we would on this project, but we are making considerable progress, and I would be looking forward to the committee's discussion on this.

[Slide.]

So, there are really two related projects here I want to talk about, because they play off one another. First, is the general framework for biomarker qualification, and then some of the things we have been

doing with pharmacogenomic biomarker development, because we are using that as a worked example.

It is very useful to have some real assays or other biomarkers in front of you that you are working up and starting to qualify, and that really gives you an idea of what the issues are in qualification, so I am going to go through that in some detail.

[Slide.]

Now, our original plan, as we discussed with the committee, was to conduct an internal survey of the use of surrogate markers and other markers in the new drug review process, and we started that, and we encountered immediately some obstacles.

It turns out I think the terminology confusion that reigns around this issue would have rendered our results uninterpretable if we had just gone out with the survey asking for surrogates.

People who were using long accepted surrogates didn't think of them as surrogates--that is how clinicians think--and therefore were extremely surprised when we said, well, maybe an x-ray of the sinuses is a surrogate. That was a very surprising finding to some people.

So, we decided what we needed to do is take a survey of all the measures that were used to assess drug effectiveness and then from that, we can go back and try to

construct a hierarchy or a framework or whatever for how we think about these things now and what we are using.

We also used composites, composite endpoints in a number of disorders, some of which are biomarkers, some of which have more clinical utility, and so forth, and so we just need to collect all this stuff first and then we can go back with the staff and discuss it more.

[Slide.]

So, at the working level, we just don't have enough clarity to directly collect surveys.

There is disagreement or maybe confusion, I am going to talk about this more, over the nature of the measurement instruments or the technology, and people confused the marker itself with the assay, x-ray, whatever, that is used to assess the state of the marker. So, that was confusing people quite a bit.

Since we didn't have a comprehensive inventory of what efficacy endpoints are actually used in trials across all different indication areas, this seemed like a very useful thing to do. Probably once we collect this, we will share it with the committee.

[Slide.]

So, we have structured a survey and we have done our pilot testing of the survey with our staff to make sure

they can answer these questions, that they are understandable and they can answer them.

We will be conducting this over the next several months, and we expect we will be surprised by what we learn from this, but we think it will be iterative, because we will get answers that we think will be uninterpretable, will have to go back and get clarification, and gradually construct an inventory, which I think will be useful for a number of purposes.

[Slide.]

But we will be able to see within this inventory which endpoints actually would be considered biomarkers, and that should help us in developing some kind of hierarchical classification.

We have established, and we talked last time, and this is widely promulgated in the literature, a clinical endpoint and a biomarker are somewhat different things. Clinical endpoint measures how a patient feels, functions, or survives. That is the definition people use for it.

I personally, and this is one of the issues here, I personally believe these clinical endpoints are also biomarkers. They are simply biomarkers that we accept their validity of, but you could say they are biomarkers or clinical endpoints, or you could say clinical endpoints or a subset of the general universe of biomarkers.

It really doesn't matter, it's just a matter of definition, because, in fact, as we discussed last time, a clinical endpoint usually requires measurement technology regardless, it requires some sort of measurement to assess and quantify whether it is pain or suffering, another kind of suffering or morbidity, or whether it is even death.

I mean right now we may require, whether it's an epidemiologic definition, case ascertainment definition of survival, or whether it is actually whether somebody is dead or not, you know, have they reached a certain state.

If you have ever actually run certain trials, you know what I am walking about. Anything that you measure in a trial requires some sort of application of some sort of definitions or technologies to quantify it.

This measurement technology requires validation in and of itself, and as I am going to get into, that is what the device center for assays calls "clinical validation," is the assay measuring what you ask it to measure, what it is supposed to measure.

For clinical endpoints, though, we require this type of validation, is it measuring what you intend to measure, but we don't require demonstration of utility.

We assign that face validity. It is useful to the patient to know whether or not the pain is gone or whatever. That is something that has face validity versus

a serum marker or something, if your cholesterol goes down, and you are patient, that does not immediately make you feel better. That is something that you are projecting will associate with a clinical endpoint.

You can see that these ideas are somewhat convoluted, and this is why it is very difficult to take this all out and to do a survey and to find out what people are actually doing when they do trials with all these measures.

[Slide.]

Now, as we have started to work through, though, the qualification concept, it turns out it really does hold tremendous promise for achieving progress in biomarker utilization.

I think what I told you before when we met in March, that one of the huge problems we have had with biomarkers is a surrogate endpoint issue, and people are thinking it's either useless or it's a surrogate endpoint.

Then, people have this huge fight about whether or not it had achieved surrogacy or not. In fact, it is almost irrelevant.

Qualification concept says is it useful for what decisions you are intending to make with this biomarker, and that turns out, as I may talk about in pharmacogenomics, that is an extremely important concept

and is actually a sort of healing concept that we can get over the issue of surrogacy and everything, and move on to how we are going to use these, because it turns out the acceptable performance, what is acceptable in the performance of a biomarker varies greatly with what you intend to use it for and what the consequences of that might be.

Use of a pharmacogenomic marker, for example, to pick out people, say, who are going to respond to a cancer treatment, that is not a surrogate marker, that is simply patient selection, however, that may have tremendous consequences for a patient.

It may be of paramount importance that you sort people correctly into those who are going to get this chemotherapy or this targeted therapy, and those who are not, you know, for their survival, for their ultimate clinical care.

Therefore, the performance required of that marker, and it might be sensitivity, it might be specificity, you have to work through the individual case, it is very high. Certain expectations, you have that marker, even though it is not going to be used as a surrogate marker.

In other cases, it's anatomic markers that are used as surrogates, the bar may not be that high, because

there is a tremendous amount of face validity. If you lose half your myocardium and you can demonstrate that, you know, most people are willing to buy that as a bad thing clinically happened to you.

So, we really need to start thinking in a more sophisticated manner about these markers and how we use them in trials.

What we are going to need to do in addition, and this will have to happen after we go through this first step, and this is why I think Steve Gutman is here and others, we are going to have to merge the concepts of in vitro diagnostics, which have a very different conceptual framework, predictive value, and so forth.

We have to merge that, as well as people who do radiology and imaging, those concepts, or psychometric measurements which have a whole different type of set of concepts of the instrument, with the ideas of using pharmaceutical development.

These two things are not going to come together that easily, but I think the qualification concept will provide a bridge to match up, say, in vitro diagnostic concepts with drug development concepts in a way that can work for both sides.

I think from pharmacogenomics, which is moving so rapidly now, and I heard you had a good meeting yesterday

on this one topic, that is going to be our test case I think of how to do this, because in my mind, we need assays out there. We cannot simply have these to be research assays in various people's laboratories. It is not going to work in clinical practice. we need really simple tests that people can understand.

[Slide.]

So, let's move on to pharmacogenomics since I think that will be the test case. I think pharmacogenomics, and it sounds like it was demonstrated yesterday, and I know people in this room believe it, that there are a lot of people don't understand or believe this, will help move therapy from trial and error, say, in dosing, to a scientifically-based prediction. That is really where we must go as rapidly as possible.

It also will help us refine definitions of disease, and that is where, for example, the cancer example I was using, you know, we need to get down to the molecular definition of cancer, not the organ system definition of cancer if we are going to use targeted therapies appropriately.

We are also going to be able to avoid certain adverse drug events, not necessarily from dosage, but actually be able to predict people who are at risk of getting them, and not have them be exposed to the agent,

and then we are going to be able to select patients for therapy based on their predictions of their response.

So, this field has tremendous promise for drug development, but, of course, it's going to be a struggle now we are in that transition phase. One of the things that we need to do is make sure we understand how these assays are going to be used in a technical manner in development.

[Slide.]

Now, I think we have made significant progress in recent years, a lot of it due to Larry Lesko and his colleagues. We have had all kinds of public workshops, as well as coming before this committee. We have had guidances. We have a functioning process to submit the voluntary submissions to the FDA, and we have had about 20 of these submissions of pharmacogenomic data. We are reviewing those.

The Device Center has approved a number of pharmacogenomic diagnostics that are making this real. A clinician can actually get hold of a kit in their laboratory and you can actually test people for pharmacogenetic, for drug metabolism.

We are making efforts on co-developing a drug and a diagnostic together, how would you actually do that. That is where the issue of biomarker qualification arises.

[Slide.]

Now, we recently had a workshop on genomic biomarker qualifications, a little bit confusing. They are both scientific, they are clinical, they are nomenclature issues that I just went over, and then they are procedural issues about how to go about this.

But the basic question which is true also, going to be true for imaging and all the other biomarkers we would like to qualify, is how do we get to tests that are usable for regulatory decisions in drug development, that is key, also, interpretable and valuable in the clinic, so that they will be taken up in the clinic once their value is shown.

[Slide.]

It will be very important to do this work because if we believe, as I do, that this is going to improve medical care, we have to provide persuasive data on the value of doing this. Otherwise, the clinical community is not going to change practices.

We need evidence that can be used for cost effectiveness analysis, that will help payers in making decisions around reimbursement, because if using more biomarkers is just seen as adding more tests and adding more costs, and isn't seen as providing value, then, again, health care is not going to take this up rapidly.

The validation that is done we think during drug development can help establish protocols for the use of these tests in clinical medicine, because clinical medicine itself doesn't really have many procedures for figuring out how to use things properly.

[Slide.]

Now, what do we mean, though, by validation? That is really the subject of my whole talk. We prefer, as was discussed in March, not to use the freestanding term "validation," because it means many things to many people, and usually, it sounds like some huge burden that we can't accomplish, but we don't know what to do. We have to do something very difficult, but we don't know what.

In contrast, the concept for a test kit, you know, diagnostic, of analytical validation is fairly well understood, and, in fact, if you talk to the imaging folks, the same is true for various types of imaging.

It has all the same things, inter-observer variability, or inter-test variability, all the kinds of things you would do is fairly well understood.

We are working right now, for example, in FDG PET, trying to look at how it might be qualified for tumors as a response measure for tumors, cancerous tumors. The same issues have come up, very interesting.

Platforms haven't been standardized, and the validation work hasn't been done, so you could take an image from over here in this center, and then you can compare it to this center, and you don't know if you have apples or oranges.

So, we are also, in addition to pharmacogenomics, we are working on developing the standards with the professional societies, developing standardization for FDG PET, so that it can be used as a biomarker.

So, this is kind of the sine qua non, I think for biomarker is whatever you are using to measure it, the measurement technology has to be stabilized.

Now, in the medical device context, they use "clinical validation" to reflect the idea, as I understand it, that you know what you are measuring, that the measurement, the test really measures what you think you are measuring.

Now, in pharmacogenomics, this could be a big problem, because you don't know what you are measuring often. A lot of the microarray tests, you are just developing correlations, you are not trying to measure something that already exists, but what we are going to use, as I said, for pharmacogenomics, is thinking about qualification for use, because the exercise is going to be very different depending on what use you are aiming for.

Steve and I have talked about this, and we think we can merge these concepts in a seamless way.

[Slide.]

So, let's consider what this might mean, something is valid.

[Slide.]

There are three interrelated concepts here, and this is part of the confusion that people get into. The biomarker, we all agree is something that is measured, right, it is not the measurement technology itself. We can have a discussion about this, because I would be interested in your views.

Most of us believe, though, you know, it is sort of a scientific belief that there is a real physical state or reality that is being measured, and if you take something that has been well established, serum electrolytes or pulmonary function tests, whatever, you could measure those things using different tests, and hopefully, you would be measuring the same entity, so that, I think is the biomarker, that is my theory at least.

A test then is the measurement technology that is applied to that to try and quantitate it. For genomics, it can be a very straightforward test, but usually, it is going to be a pretty excruciatingly convoluted type of test

that has to be applied with many steps, and so forth, to get hopefully, a reproducible result.

A pharmacogenomic test is a genomic test that has meaning, clinical utility, vis-a-vis some aspect of drug therapy.

So, that is kind of the different validities we are working on here.

[Slide.]

Now, what we are beginning to think through here is what characteristics of the marker itself contribute to validity. Well, for some of them, a tremendous amount of mechanistic knowledge exists, so we generally accept that for drug metabolizing enzyme polymorphisms, they have been studied a long time, we understand mechanistically their role, so there is a tremendous amount of validity that simply applies to the fact that there is a lot of scientific evidence.

Molecular drug targets, most of you have probably been reading the literature, for example, on EGFR, and all the controversies about for targeted drug therapies, what that target represents, what the different measurements might be.

They have less--there is a mechanistic link, though. We believe that for targeted therapy, there is

something about that receptor, either its expression or something that probably correlates with drug response.

Tissue injury gene expression sequence, that is one of the things that is now being studied, is you simply injure tissues and then you see what genes are turned off and on, and so forth, at different stages after injuring the issues.

That is pretty empirical. Then, you have purely clinically, empirically derived correlation patterns.

So, those are the ranges of types of pharmacogenomic biomarker data that you might get from something that is very well understood, the biomarker, to something that is just emerging, and you have no idea what mechanistically it means.

[Slide.]

Now, where you have mechanistic knowledge, this contributes support for the validity. So, your confidence is really a lot greater when there is a physiologic, pathophysiologic or pharmacologic link, a plausible one to what you are trying to measure.

Empirically derived associations, on the other hand, have only one line of evidence for the link, the correlation that you have done. You might do it multiple times, but it requires more robust data for that type of

evidence. You may have to repeat that correlation more times to believe that the marker really has some meaning.

But the goal overall, we need to understand the marker, that physical marker, in a context of disease process, in other words, embedded in a matrix of scientific knowledge and adding to our understanding of clinical medicine.

Of course, many of the new biomarkers that are coming out now are not doing that yet, because we don't have the context, we don't understand their correlations, their mechanistic correlations whatsoever.

I guess yesterday, the data that you heard is really starting to embed the understanding of the metabolism of warfarin into the matrix of the clinical outcomes for that drug.

[Slide.]

So, when you are talking about "degree of validity" of the biomarker, in principle, you are talking about the physical marker, not the specific test that you use, but obviously, the specifics of the assay are make or break on this, as we all know.

In the pharmacogenomics guidance that we have, when we are talking about known or probable valid biomarker concept, we are talking about the marker itself and all the

information that pertains to marker. We are not talking about a specific test necessarily.

Independence from a specific test actually, that elevates the scientific robustness of the biomarker up even higher where you get the same results if you apply different tests.

[Slide.]

Now, as I said earlier, when you think now about qualification, a pharmacogenomic biomarker can be used for many purposes, and you can use them for animal toxicology, early and late drug development where you don't intend to use the marker out in healthcare, you are simply using the marker to aid in drug development decisions, or you may use it in drug development for clinical decision-making and you fully intend for that marker then to be used in conjunction with the drug once it's out on the market.

So, there are a variety of different scenarios of use. We did do a survey of biomarker use in drug development with our staff about a year and a half ago, and we looked at imaging and a couple of other things. There is massive use of various biomarkers during drug development. It is just very few of them are used in that clinical decision-making piece.

[Slide.]

So, as we already said, the type of qualification or validation differs. All tests, though, that you do must actually--I say "recommended" here--but really ought to achieve reasonable analytic validation. You shouldn't be making drug development decisions or clinical decisions on a test whose results you can't interpret, because it is not reliable.

Now, biomarkers that are used, though, to eliminate candidates, for example, drug candidates, or choose amongst various alternatives, which might be safety or other biomarkers, need less certainty, however, because you are really assuming, the company or the developers assuming that risk that they are using unreliable decision-maker.

Biomarkers used to select or reject patients need a higher certainty, but the level of predictive value required depends on the use, as I already said.

For example, in cancer, where you are doing a pick what choice of treatment a patient has, that is a pretty high bar, but, say, where there are a lot of alternative therapies available, say, NSAIDs, what if you had--and I am a rheumatologist, so, you know, we are highly empirical, and we firmly believe that different patients respond differently to different NSAIDs, now, if you had some miraculous tests that could actually predict better a

patient response to NSAID, that would be very good, however, it wouldn't be life or death, because what we do now is we give them one NSAID, and if it doesn't work, we give them another NSAID. That is what we mean by trial and error medicine, that's what we do.

So, that has a lower threshold obviously, of certainty, needed for it, and surrogate endpoints really need the highest level of assurance, because there, you are actually making a decision to put the drug onto the market based on data from that marker, so you had better believe it.

[Slide.]

Now, analytical validation for genomic tests, as I said, this varies. I am very familiar with validation of psychometric tests. I am becoming familiar reluctantly with how you validate imaging technology. It is very interesting, but different.

Here, this is like an in vitro diagnostic. You have to go through a lot of steps to determine how accurate and reproducible your measurement is, how precise is it, what range does it function in, what sample conditions work, how do you run the test, blah-blah, and you all know all this. Maybe this is partly for the benefit of the audience.

What is very interesting in genomics, to me, and this is somewhat true in the radiology area, it seems like, is that many of the people actually developing these are nontraditional diagnostic developers, so they don't have the benefit of all this experience of all these pitfalls that are required for analytical development of a diagnostic test.

[Slide.]

Now, a lot of issues have arisen in genomics about use of stored samples and how valid that is and everything. One of the things you have to do is make sure that storage conditions, for example, don't affect your result. That could like lead you way down the garden path, but you could perform analytical validation on stored samples if stored samples are okay, and that helps simplify this part of the work quite a bit.

It is really desirable to configure tests and do this analytical validation prior to employing the tests in real-time clinical trials. We can say this a lot, but this often in practice is not what happens.

Often, of course, during development, people change their tests, reconfigure it to make it, you know, you could run it more rapidly, run larger samples at once, and everything. Whenever you do that, you are going to

have to bridge between the old and new tests. As I said, these are standard things.

Another thing that really hasn't been focused on in the genomic area is specification of what result is positive, negative, and so forth, what are your thresholds and cutoff points.

The Device Center has recommended for most diagnostic tests that they use receiver operator characteristic curves or other analytical methodology simply to determine where your cutoff is, but I would point out again, and I know I already said this, but where you put your cutoff depends again on what you are going to use the test for and how you want to balance specificity and sensitivity, and so forth, against one another.

There is going to need in the genomics area for attention and focus on these issues regardless, but how stringent you want to be depends on what you are going to use the test for.

[Slide.]

Now, as we move down in pharmacogenomics, the tests the Device Center has approved are freestanding, in other words, they weren't approved just with a drug. They were approved to look at 2D6, different metabolism alleles.

A test labeled to be used with a drug can be developed in quite a different way than a test that is

going to freestand and say you can test any patient and you will get this range, and you can find these ranges of genes.

It really depends on the amount of pre-existing scientific knowledge on the clinical utility of the result, so the Device Center is able to approve a lot of freestanding drug metabolism tests I think because we have so much mechanistic knowledge about that.

For a lot of the new genetic information I expect many may be approved in conjunction with a drug initially, because as I am going to talk about in a minute, the drug trials will be used to validate or elucidate the clinical utility of that test, and that is a special case of co-development of investigational test and investigational drug.

Now, I am going to go over this pretty fast, because I know this committee is not interested in this, but one of the real promises here in pharmacogenomics is actually to get better safety biomarkers, and at the FDA, that starts with the animal toxicology.

Actually, that is a tremendous test system, because what we do, what is done in drug development is, first, expose a lot of animals in routinized protocols and then you go ahead and expose humans, so you have the bridge, animal to human bridge for those data.

Then, you may go on if a drug is successful, which doesn't, unfortunately, happen all that often, but does happen fortunately, and then you may expose tens of millions of people, so you have a very rare case where you could actually start evaluating a predictive value of genetic biomarkers in animals and look all the way to the human outcomes over time.

Now, animal testing those traditionally used to select starting dose in people, identify the potential target organs where the toxicity might emerge, and also to identify special toxicities that are poorly tested for in human trials for ethical reasons - reproductive toxicology and carcinogenicity.

Identifying new markers that could provide more precision and predictability in animal tests doesn't require a high bar on those markers, however, it could give the field a black eye if, in fact, you started relying on markers that were not very useful.

Identifying markers, which is the goal of some, that ultimately might substitute for animal testing would be a much higher bar, much more difficult, but we need to start, and I am pleased to say this work is getting started in getting a lot of the biomarkers validated in animals.

We can access their performance or predictive value across a wide range of settings and drug types if

companies are willing to share their markers with each other in drug development, so they can use those assays. We have been encouraging this to happen.

[Slide.]

Now, I already talked about metabolism is a special case because such a large body of existing data exists. The assays might be approved as "freestanding," but we, I think the people at the Drug Center, are extremely interested in then putting that information into labels of drugs, so they start being used more rationally as far as dosing.

Development of drugs that are subject to polymorphic metabolism is a specific area of interest.

Now, in general, the pharmaceutical industry's response to finding a candidate that is subject to polymorphic metabolism is to eliminate that candidate, but in some cases, it might be the only game in town, and therefore, that would provide a very good opportunity to actually, where there are no other treatment alternatives, to go ahead and develop that drug with the dose directed by metabolism.

As you all know very well, this has not really happened to date. Right, Larry? Right. It's happening? That's good. It has not finished to date, yes.

[Slide.]

Human safety biomarkers, what we would hope, and there is early evidence that this actually will occur, can occur. We could take those animal markers, develop the human analog marker and provide a more sensitive screen for early toxicities in people.

What we do now is wait until we use our old-fashioned biomarkers, liver enzymes, whatever, electrocardiographic QT prolongation, whatever we use that has been around a while, and we wait until those develop to declare defeat.

But we could with better biomarkers try to monitor this early and do trials that would look at withholding therapy as toxicity emerged, but that would raise the issue of using the test postmarket. If we actually develop the drug this way, then, the safety would be predicated on the use of this test. That is a higher burden.

We would like to work with people as part of the critical path, and we are doing this to develop new genomic safety biomarkers that will help better predict organ toxicity in investigational trials.

Of course, this would need to be published, become generalized knowledge, so it could be picked up and used, and the assays would need to become available, so they could be used.

[Slide.]

So, if they needed to become available, they would need to be qualified either as freestanding tests or for use with the drug or several drugs, and a commercial test configuration would need to be developed, and this, of course, is a huge barrier right now to availability.

[Slide.]

Now, this is what everyone is interested in, human efficacy biomarkers. Of course, we can use these in a variety of different ways, many of which may not require that they be used later in the clinic with the drug, some which might.

[Slide.]

What we are doing here is well under the critical path, is trying to develop consortia to get existing biomarkers further qualified. There are many candidate pharmacogenomic markers. The performance data for these reside in one firm or within an academic setting. This is a class problem we have been dealing with. The data may or may not be public.

Wider acceptance in actual use is going to require further performance evaluation, further qualification in multiple hands with a variety of therapeutics, so we start believing these are real, and biomarker consortia provide an ideal setting in which to

perform such work, and I can't elaborate on this right now, but these are being set up.

[Slide.]

We need novel processes because the current models for any of these biomarkers I have just been talking about are either nonexistent or they have been unsuccessful in developing new biomarkers.

I think you are going to talk about HIV and maybe some biomarkers for infection. That is one of the few successes over the last perhaps couple decades where we actually have some new biomarkers.

Many non-pharmacogenomic markers have been available for decades, but their utility either in drug development or in the clinic is still unclear, and this cannot be the fate of genomic biomarkers.

That is why under critical path, FDA is trying to intervene in this and get this qualification worked on, because we have to build a robust qualification model to improve both the safety and effectiveness of drug therapy in this country.

[Slide.]

Now, in many cases, as I said, the pharmacogenomic tests and the drug will both be investigational, but we should look at this as an opportunity, not a barrier, because in this case, you can

rely upon the clinical phase of the drug development program to also provide the evidence of clinical utility that is needed to approve the diagnostic test, so can kind of piggyback onto the drug trials.

In this case, the claim for the test would be for use with the drug, not a freestanding claim, and the drug would be cross-labeled for use with that diagnostic, but other parts of the drug development program and the device diagnostic development program would proceed as usual.

Because this is pretty tricky and novel, we are working obviously very closely, the Drug Center, the Device Center, and my office are working very closely on this, and we do have to get our guidance out soon, our draft guidance on co-development, which will go into many of these issues.

[Slide.]

The questions that have arisen in many of the workshops we have had, and so forth, and get to the larger issue, many of them, of biomarker qualifications, how would you design trials to accomplish the objectives of proving the usefulness of both the drug and the diagnostic in a series of development trials.

To what extent are you able to identify the biomarkers and then qualify a genomic biomarker in the very same study? That does seem to be asking quite a bit, given

what I earlier said about the mechanistic robustness, and so forth.

There is a lot of issues related to generalizability of results when you have used a genomic test's select population or have moved people around, or whatever, how generalizable is that result to a broader population. Those are policy issues that have to be solved, a similar degree to which a study in a rich population actually pertains to a broader group or not.

That leads to questions about approval of a drug in a newly identified subgroup of a larger population.

Now, I am personally not all that sympathetic to this issue, because frankly, it seems like over the years, you know, a long time ago, we just had rheumatism and then all of a sudden we discovered rheumatoid arthritis, osteoarthritis. Now we have many different subgroups, the same with cancer.

So the history of biomedicine is actually more carefully defining disease, and if these new subgroups actually more specifically, mechanistically, define disease, we should not resist this.

In today's healthcare world, though, there is always the problem. You approve it for some carefully defined mechanistic group, and then people start using it for everything.

That is one of the problems we have, but that should not be permitted to prevent the advance of science to a more scientifically directed drug therapy than in the past, but we will have to deal with this.

[Slide.]

So, in summary, the inventory we are going to do of efficacy outcome measures will assist in building our biomarker qualification framework. We have been thinking about this a lot, and we think we will get a lot of data in that will help us arrange this in a way it will be understandable to people.

Pharmacogenomics is really moving along, and the markers are being utilized rapidly. We think also the experience with these markers as worked examples will inform our overall biomarker effort, so we are going to be out in front with the co-development of an investigational diagnostic and drug, but other technologies will then be brought into that.

So, I thank you very much.

[Applause.]

DR. RELLING: I think we actually have a few minutes for discussion right now, so questions? I see Dr. Flockhart.

DR. FLOCKHART: Janet, thanks. Lots of food for thought. Just one thing that came up yesterday and I think

is a valuable thing, but which didn't come up as one of the qualifying criteria for a test, and this is I guess is as much for Steve as anybody else, and that is the incremental or the iterative value of a test in a clinical environment.

I dillydally between whether you should use the word incremental or iterative. I prefer the term iterative I think because it implies some quantitative thing, but the sample I used yesterday is in breast cancer where if you are looking at how effectively a drug works, you have, well, the stage, grade of the tumor, you have the number of nodes, you have the age of the person, you have their general state of health, you have all these things, and you have the estrogen receptor, and you have HER2/neu, and then on top of that, some test has to iteratively or incrementally improve.

I think we need to be thinking about ways to quantitate that improvement, not that those would be generally applicable in every setting, but there is a big difference between a test that provides a small incremental value, however good statistics, and one that provides a big thing, so that it provides a predictive ability that we just didn't have before versus one that just allows us an ability to get a better handle on how an antihypertensive works on blood pressure.

That is the first thing and then I think the second comment I would make, and then I will shut up, is I think that we talked yesterday about what subjects the committee might get into later on a little bit.

There has been great value in the discussions about drugs that are off patent, if you like, which we, in the academic setting, and obviously people in the environment in this committee, have been very interested in, but for those in particular, you run up against a really difficult issue, and we have run up against this several times already, with validating multiple populations, going to multiple places to do it, so the specific issue of the biomarker consortia comes up very directly.

So, I wonder if you could--you said you weren't going to expand on that--but I wonder if you could expand on that just a little bit, because it is very important to lots of things this committee might do.

DR. WOODCOCK: The second question first, the biomarker consortia. It has come to many people's attention, I mean it is intuitively obvious that biomarker qualification, whatever the process might be for any given marker, is not something that really can be accomplished usually by a single company, an academic center, one NIH

grant, et cetera, et cetera, in other words, the usual mechanisms. That is why it hasn't happened very much.

So, FDA has been exploring, as have others, setting up consortia between public and private partners to get some of this work done, and the goal there is that many people share the pain, many people can contribute various resources.

Qualification usually will have to happen in several settings to provide the adequate generalizability, and for a biomarker used with a drug, you might want to use a number of different therapeutics to evaluate the response of the marker or its predictive capacity in a variety of situations.

All those things mean that we really need to have a consortium so people can bring everything to the table. Yes, we are trying to get these set up, this is happening. Consortia that are public/private partnerships in general need to have, as a central core, a nonprofit organization or some other organization like that where the partners can come together on neutral ground.

So, put differently and more simply, there is a lot of like it's very lawyer-intensive for a while, and there is a lot of paperwork and discussions and plans, and everything would have to be drawn up, but we are working on that.

We would welcome other people approaching the FDA, you know, other foundations or whatever, if there is interest in setting up other consortia around, but the consortia trying to be set up on the animal toxicogenomic biomarkers, on clinical biomarkers, and so forth, pharmacogenetic biomarkers, and I think these will happen. It is just frustratingly slow, and that is why we can't talk about them or announce them.

DR. RELLING: Dr. Barrett.

DR. BARRETT: Dr. Woodcock, thank you for the presentation and that clarification.

One of the other things, as you pointed out, with the big consortia, is, in fact, the difficulty in mobilizing everybody to consensus, so I implicitly agree with that.

I wonder if there is any smaller efforts in kind of targeted areas where you have a little bit more mobilization and energy in a particular area.

One of the things that comes to mind, at a recent AAPS/FDA/Pharma biomarker conference, there was a discussion about the QT guidance and specifically on the area of whether or not there could be some movement in the preclinical area of establishing that safety-toxicology link to clinical outcomes, and there is a lot of dancing around this issue, but no real effort to move it forward

because it does require the collaboration of several of the stakeholders in conjunction with FDA and possibly Pharma, to, in effect, do something specifically in that area. It would be a nice example if it were possible.

But I wondered if you had other specific areas where those kinds of efforts could be moved forward.

DR. WOODCOCK: The question that was on QT prolongation, cardiac repolarization abnormalities, and whether a consortium could be formed around some of the preclinical safety assays to work in the animal setting to develop more predictive assays first.

Well, it has been announced that we are working with Duke University with their nonprofit to set up an ECG warehouse, that is a human setting, to share ECGs and do analysis. That kind of setting might be good also to do the animal work in, or perhaps animal toxicology consortium I was talking about, which is larger, though.

The problem with most of these is getting somebody to drive them, figuring out what projects need to be done and talking everybody into coming together and doing it, get Larry to do it maybe.

You know, we don't need different players for different areas, but I agree with you, and obviously, that is one, cardiac repolarization is a big problem, we don't have really good predictive markers. We are doing large

clinical trials of unknown utility, and that is a big problem.

DR. RELLING: One last question. Dr. Sadee.

DR. SADEE: I have some question about the degree as to what biomarkers are used. It is clear that biomarkers are going to move forward together with drug applications, and so then we set a language that might say, well, there is a biomarker available, you might want to consider it, or it is recommended, or it becomes mandatory.

The language or the barriers between these various steps is not entirely clear to me, and I think there are just really tremendous implications, because this is going to move forward very quickly, and at what point does it become standard care or we say this is recommended, and then you read in the literature that this indeed improves therapy, and then all of a sudden becomes standard care.

So, how do we deal with these issues?

DR. WOODCOCK: Well, that is where we are trying to deal with them through developing some standards. We started this with the pharmacogenomics guidance where we talked about biomarkers that are not known to be valid, and we said we are not even going to use those for regulatory decisions, but you can voluntarily submit them, so we can discuss them.

Now, the field is moving rapidly along and it is going to start begging the question soon when are they going to be--they have to be part of the regulatory submission, okay, when are they valid enough that really you can rely upon certain ones of them, and then your question is you are reflecting the usual clinical route, which is like a number of university professors published series on this, and then the clinical people pick it up and start using it sort of in an ad-hoc basis, and then it starts being used.

I think what I am saying is I don't think that is a very good way of developing things. It doesn't develop an adequate evidence base. Look at all the tumor markers that have been around for the last 20 years. We still don't know how to use them, and we don't accept them in drug development, the FDA doesn't accept them as outcome measures. So, that is a very bad outcome. Even if they are used in clinical care, they are not used in developing the drugs to treat those conditions.

So, I think what we are trying to do is find a better way, but it is not easy.

DR. RELLING: Thank you very much, Dr. Woodcock.

In the interest of staying on time, we will introduce Dr. Lesko, who is going to talk on Use of

Biomarker Information in Drug Product Labels to Individualize Pharmacotherapy.

Use of Biomarker Information in Drug Product Labels to Individualize Pharmacotherapy

DR. LESKO: Thank you.

[Slide.]

We thought it would be useful this morning for the committee to hear an update from Dr. Woodcock on the critical path initiative, and what you heard I think was a very broad update touching upon many things, but certainly the topic of biomarkers.

The next three presentations are really designed to hone in on biomarkers as a specific critical path initiative to set the stage for the subsequent discussion of the committee later on this morning.

I think the message from the critical path initiative is that if biomarkers are used successfully within the constrains of the continuum between preclinical drug development and clinical practice, they have much value.

They can facilitate drug development, they certainly can help regulatory assessment of the things that we see, and they have the potential to lead to diagnostics that can improve patient care.

Of course, along that continuum are many issues that come to mind, and what I will be doing in this presentation is really to give the committee a perspective on how we might think about using biomarker information to inform clinical decisions.

[Slide.]

Now, thinking about this issue of labeling and biomarkers and how they can be used, I will go back to yesterday and recap what we discussed. I will use the two or three biomarkers that we discussed yesterday as examples.

We talked about for the anticoagulant 2C9 and VKORC1. We heard from Dr. Powell about the viral load, and we heard about blood glucose from Dr. Wang.

The purpose of these biomarkers is basically to match patients to dosing, maximize success in clinical trials, and the innovative part of critical path is how can biomarkers be better utilized and what types of innovations can help get us there.

[Slide.]

So, the critical path initiative seeks solutions to the productivity problem. Within that document, there are many sections that talk about opportunities, but one of the heavy emphasis in this document has to do with biomarkers, and these are some of the quotes from the

document that talk about new tools or new science that include, amongst other things, biomarkers, clinical trial endpoints, and as we heard about yesterday, model-based drug development.

Furthermore, and I think Dr. Woodcock emphasized it, is that technology is moving forward whether it be genomics, imaging, proteomics, and these are going to provide even more biomarkers, so that the impetus is to get our hands around these and learn to use them effectively, not only in drug development, but also in clinical practice.

[Slide.]

Now, this is a standard textbook definition of biomarkers, but I have added another dimension to think about with biomarkers. For example, we have common biomarkers, which would be the more traditional ones that represent a single feature.

We heard yesterday about INR, which confirms the activity of a drug. We also heard about warfarin levels that predict activity of the drug in terms of clinical outcome.

We also heard yesterday about more accurate biomarkers, if I can use that based upon predicting variance in dosing, that use multiple features. We heard from Dr. Caldwell when he did a multivariate analysis that

used a whole bunch of patient demographic and environmental factors to prognose on disease outcome.

As we talk about genomics, we even get to the molecular level with more precise biomarkers in terms of haplotypes and gene expression, as we heard from Dr. Gage on the VKORC1 talk.

[Slide.]

So, there is really levels of biomarkers that are emerging, and they all have a potential for being pharmacodiagnostic agents, but I think it is safe to say that biomarker discovery programs are growing at a rapid rate. Virtually, every company has them, and these are again related to the technology development, as well.

But the issue that I would like to raise for potential discussion today is these biomarkers not only help drug development, but create a potential for individualizing treatment, whether it be on a subgroup basis or whether it be on individual patient basis.

We can put it in the framework of bringing a scientific basis to the art of medicine, sort of what Dr. Woodcock talked about in terms of her nonsteroidal example.

So, one of the questions to think about for our discussion period is how can we obtain biomarker information effectively and efficiently during drug

development, and when is it important and critical to translate this into labels for direct inpatient treatment.

[Slide.]

Now, biomarker development programs really start from square one in most pharmaceutical companies with a heavy emphasis on understanding disease pathophysiology and drug mechanism of action.

What I have listed on this slide is really a bucket of information that comes out of drug development with regard to biomarkers. For example, moving from preclinical to early clinical frequently, plasma drug concentrations are the target in terms of designing proof of concept trials.

Later on, moving into late Phase II/III trials, biomarkers focus on risk factors or individual measurements that direct the therapy, or entry criteria for identifying responder patients.

There is also dose finding data in most drug development programs that are used typically to select Phase III doses.

So, my point is that there is a lot of information generated during the drug development process, and the question is how much of that information is directed towards qualification of biomarkers and through some of the critical path activities, can we enhance the

qualification of biomarkers through, for example, disease, drug, and statistical models that we saw yesterday.

As I say on the bottom, I think this is actually done relatively infrequently in drug development programs, and as a result of this, there is limited opportunity to use biomarkers as diagnostics in clinical practice.

[Slide.]

So, if we reflect upon late phase trials in particular, the randomized controlled trials, these are obviously traditionally been intended to provide the best evidence of rejecting a null hypothesis of no treatment effect. The goal of these trials principally is to demonstrate efficacy and, to a degree, safety.

Generally, in analyzing these trials, one assumes a homogeneous population. They are not designed, generally speaking, to qualify relevant efficacy and safety biomarkers prospectively. That information occurs generally earlier in drug development, and once decisions are made based on those biomarkers, further utilization of them in terms of qualifying them or validating them is generally not done.

Those studies, if they were to be done, need to address a different set of questions. They need to focus more on the heterogeneity of patients, not on the homogeneity of patients, and it leads to another question

to think about in our discussion period - how can better biomarkers or diagnostics be incorporated into drug development, that is to say, how can we develop prospective hypothesis testing beyond the early trials that are used for things like go and no-go decisions.

On the bottom, I have indicated I am not talking about surrogate endpoints here, I am talking about biomarkers that have potential application in patient care.

[Slide.]

So, what kind of questions am I talking about that would need to be addressed more specifically along this critical path?

Well, one of the questions is generating more fully dose-response relationships for benefit and risk. I think it is important that we begin to move towards understanding the inherent variability in these relationships, and thinking about it as a response surface, how changes in dose-response occur with patient co-factors and different dosing regimens.

Further, biomarkers most suitable to adjust doses in clinical practice could be better identified. We had a long history going through the '80s and '90s of therapeutic drug monitoring, which pretty much came to a stop in the early '90s, and we have not had many new drugs for which blood level ranges have been identified, and as a result,

we have had little progress in the diagnostic area in terms of a simple biomarker, namely, plasma drug levels.

Yesterday, what we tried to show you was that quantitative methods is a way to qualify biomarker associations with clinical outcomes by bringing together all of this disparate data that occurs throughout the drug development process.

[Slide.]

So, our model-based drug development program very much depends on biomarkers, availability and analyses. It is one of the critical path initiatives, as we showed yesterday.

But it makes extensive use of biomarkers, and the whole goal of doing that is not only to improve decision-making, as you saw yesterday, but also to begin to think about these biomarkers as potential diagnostics and what it would need to move in that direction.

Again, I am not talking about diagnostics in this context, of limiting the drug to a specific population necessarily, but diagnostics in the context of refining how we can design the dose for individual patients and adjust the dose appropriately.

[Slide.]

We have relied on a conceptual framework for biomarkers in model-based development. It is, in fact, a

framework that many companies are reorganizing around, namely, the learn/confirm paradigm that was introduced by Dr. Sheiner some years ago, and we have added to that the principles of evidence-based medicine to begin to identify the standards for thinking of biomarkers as potential tests to guide dosing.

[Slide.]

Dr. Woodcock touched on another issue, and that is drug safety and what can be done about this, and we have had all sorts of calls about what needs to be done about drug safety, but I think the biomarkers represent something that we can do without a lot of pain and bring that to bear upon drug safety.

[Slide.]

When we talk about drug safety, we talk about adverse drug reactions, and the source of those adverse drug reactions is no secret.

It relates to the inherent predisposition as we saw with 2C9 and VKORC1 and the environmental factors that define a patient scenario, and as Goodman and Gilman frequently says in their various volumes, new drugs are inherently more risky because of the relatively small amounts of data, and one could think of biomarkers about their effects.

[Slide.]

So, many adverse drug reactions are avoidable. These are the typical high incidence, less severe effects. Most of them occur within the range of approved doses, and depending on the source, depending on the study, the survey, anywhere from 70 to 90 percent of these adverse drug reactions are related to exposure and, thus, biomarkers.

So, what is needed, I think, is for us to begin to talk about better pre-marketing approaches to bring biomarkers into drug development, and eventually translate some of those into information in the labels.

[Slide.]

So, we sort of adopted a philosophy that risk management using biomarkers must be built into the process, and not simply monitored in looking at it in post-marketing surveillance.

As you can see the critical path illustrated here, is this continuum of defining benefit-risk and the role that biomarkers play in that continuum along with model-based drug development.

[Slide.]

Now, moving to the label, we talked about it yesterday, and the regulatory mandate to provide informative labels, but mainly, we communicate with providers and with prescribers, and it is an integral part

of the drug development process, targeted to the intended label beginning in the preclinical area.

So, the goal of labeling is to assure that prescribers have access to useful data that suggest ways to optimize efficacy, and not eliminate, but rather, manage risk.

We think one of the key pieces of information in labels ought to be dose-response or PK/PD relationships that are particularly important to individualizing therapy.

[Slide.]

We recently did an assessment of pharmacometrics in regulatory decisions. That was reported on in the pink sheet in the reference that I have indicated there, and the analysis was done in a limited number of therapeutic areas, primarily in the area of cardiorenal oncology and neuropharm, and a survey was done of 244 NDAs.

Of that number, 42 of those NDAs had a pharmacometric analysis applied to the database, and after that analysis, 26 of those analyses were pivotal or supportive of the NDA approval, and 32 provided evidence for label language.

Now, you might look at those numbers and say why wasn't there more of an impact of pharmacometrics, of model-based drug development, and the reason was that the number was not higher because sponsor applications lacked

the necessary data on dose-biomarker relationships, but when it is there, it has tremendous potential for both regulatory assessment and also for designing label language.

[Slide.]

So, my perspective on the current situation is that much of the value inherent in the development of a new drug, whether it's $800 million or less, which is the estimate of developing drugs, is lost in uninformative labels. That is to say, labels are effective to a degree, but I think we can do much more in leveraging information in terms of what is being done in drug development.

The thought that I had on this was that a physician without information, thinking of drug safety and patient care, cannot take responsibility, a physician who is given information cannot help but take responsibility, and that stands on the point of informative labels.

[Slide.]

Now, when we talk about informative labels, I emphasize the knowledge of dose-response or PK/PD data, and this is not foreign to drug development. This data is obtained typically for most drugs in Phase I and Phase II. Dose-response relationships are typically used to pick the doses for late-phase trials.

Furthermore, data obtained in Phase II/III trials usually contain, amongst other things, extensive information on biomarkers and sometimes blood levels of drugs, and the question is where does that data go after decisions are made within the drug development process.

In addition to that, we have these targeted studies. We call them "special population PK studies," where PK drug levels are intensive with changes in levels associated with decisions to label a product with dosing adjustments, so the data is in there.

[Slide.]

Surprisingly, we don't see very much development of exposure-response modeling throughout drug development. Most of the information that I just described is obtained in the early trials, and Phase III typically is not a point where more dose-response trials are conducted as part of the efficacy safety trials.

[Slide.]

So, if you think about the importance of the relationship between effect and dose, it turns out that very few labels actually contain exposure-response information, and it would seem to me in clinical practice, knowledge of these relationships are important when it comes time to design a dosing adjustment based on the phenotype.

A dose is started, response is observed, what do I do next to increase the dose or decrease the dose? There is two examples of dose-response information in labels that come from the cardiorenal area.

The one on the left is irbisartan. It shows a dose-response for effects on diastolic blood pressure. In that label, there is also a graph on systolic blood pressure, but it gives a sense of the dose and when you begin to plateau on the dose in a patient that may be unresponsive.

In the case of the metoprolol, there is a text rather than a graph, but it has key pieces of information that I think would be beneficial to a prescriber - where is the most sensitive part of a dose-response curve, what is the plateau of that dose-response curve either in terms of dose or blood levels, again, information that is potentially useful in clinical scenarios, but we don't have many examples of that in the label.

[Slide.]

We recently had an example that involved an antiviral drug, tipranavir, for the treatment of HIV. Generally, when we are talking about treatment of HIV, we are concerned about drug plasma levels remaining above the IC_{50}, or in some cases, the IC_{90}, to achieve viral suppression and avoid resistance development.

There is a metric that can be derived called the inhibitory quotient, which is defined as the C_{min} divided by the IC_{50}, and this has been proposed by some as a metric to predict efficacy, for example, percent responders at 24 weeks of therapy.

In looking at this application, the data contained in it, we thought there might be a potential basis for individualizing dosing in patients receiving the combination of tipranavir/ritonavir.

[Slide.]

This was discussed at an Advisory Committee, and these were the relationships that were presented by Dr. Jenny Zheng from our office. On the lefthand side, you see a curve that relates on the x axis, the inhibitory quotient, as I just defined it, C_{min}/IC_{50}, to the percent of responders at Week 24.

You can see it takes the shape of a dose-response curve and probably would have the same value in terms of guiding dosing based upon where one would be on that curve relative to a decision to increase the dose, decrease the dose, or perhaps turn to an alternative therapy.

That data was derived from a meta-analysis of several clinical trials that were submitted in the application, and on the righthand side was a relationship for risk that was correlating or showing the association

between the trough level and the percent of patients with Grade 3/4 ALT toxicity.

So, I am using this example to illustrate how biomarker information from drug development processes can be analyzed in terms of a model-based approach to generate relationships between exposure and benefit and risk, so that taken together, this information can be used to make recommendations in the label.

[Slide.]

There were issues, however, with inhibitory quotient. It is controversial. Some of those issues that were unresolved at the time we discussed this in May or June at the Advisory Committee was the exact degree of protein binding adjustment that was necessary to estimate the IC_{50}, there was a lack of consensus on that.

The estimate of IC_{50} itself was very much patient dependent, and the generalizability of that information was in question. There was some sensitivity and specificity issues with C_{min} to predict toxicity.

There was relative variability in the C_{min}/IC_{50} and having a number in hand, it wasn't clear whether that variability was related to the C_{min} component or the denominator, however, there was an agreement that future studies of drugs in this class would be conducted to determine the usefulness of inhibitory quotient of C_{min}.

So, this is an example of an emerging biomarker that has future potential in drug development programs with antiviral drugs and ultimately, the possibility of translating that into a package insert to help the provider describe the drug.

We will have a further discussion of this in the next presentation by Dr. Mayers. You can see the complexity of this, but nevertheless, the potential value of it, as well.

[Slide.]

The final label for this drug, it was approved as combination therapy. The studies that were submitted pointed towards the need for individualization based on the large degree of variability in PK. There are many drug interaction studies that influence the exposure up and down, and so on, and the label did say something about individualization, but on the genotype/phenotype for viral resistance.

So, we recommended at the Advisory Committee that individualized dosing be considered as a postmarketing experience and pilot studies will be conducted to look at this more closely, but you can imagine the potential, depending on the results of this study, to utilize this type of biomarker information and personalization.

[Slide.]

Before I give you another example, I will give you a flavor for what I am talking about. This was a case of zoledronic acid in cancer patients where our goal of using biomarkers and translating that information to the label was to optimize benefit-risk.

This is a drug that is indicated for pain and fracture reduction in patients with metastatic bone disease, and the major clinical concern is the risk, what dose would optimize clinical benefit, but not cause unnecessary renal deterioration, which was the major toxicity of the drug.

We had a lot of data on this. We applied quantitative pharmacology to understand the tradeoffs and to select the dosing in conjunction with the company's analysis.

[Slide.]

The first step in this analysis to illustrate the integration of biomarkers into decision-making was to look at the factors that influence the kinetics of the drug. Through a series of analyses, looking at various patient cofactors, it was concluded that renal function was the only major covariate of interest in terms of the pharmacokinetics, so obviously, biomarker as a renal function would be important to translate into a label.

[Slide.]

The other step that we did was to look at efficacy. Bone turnover response biomarkers were used to do this and what you see on this graph is really a dose-response curve. As the excretion of these biomarkers go down, that's a good thing from an efficacy standpoint and it is shown as a function of a log of dose. So, it gives us a sense of the dose-response for benefit.

[Slide.]

We then combined this with a dose-response form of relationship for toxicity where the x axis here is area under curve, another biomarker, and it shows how doses can be selected both to optimize the efficacy based on turnover of bone response biomarkers, and also to limit the dose based on toxicity relationships, in this case, of renal deterioration.

So, this is an example again of how biomarkers can help the regulatory assessment and begin to think about and lead to decisions about dosing recommendations in the label.

[Slide.]

In fact, this example did translate into label in terms of the Warning Section where it gave extensive information about how to use this drug prospectively and safely in terms of renal function, and it gives all sorts of monitoring information about the rate of rise of renal

function, the initial measurement of renal function, how that influences toxicity, and a lot of that information came from the analysis which I showed you.

So, this is an example of how biomarker data of one sort or another can be translated into labels for the purposes of optimizing dosing.

[Slide.]

So, in setting the stage, the summary is that critical path is, by definition, an opportunity for innovation, and in the critical path is a recommendation that there should be a systematic effort to include biomarker data in the labeling.

You might think of this as translational science - when it is available from drug development, when validated assays can be available, and Dr. Gutman will talk a little bit about that, when we have meaningful clinical outcomes, and there is a lot of detail in what that meaningful word means, when they are potentially useful in guiding dose adjustments as we saw yesterday with some of our biomarkers and when they can be used as an adjunct, and not a substitute for the traditional clinical monitoring of patients.

So, this is a framework to think about as we go into the rest of the morning.

[Slide.]

A couple of acknowledgments of my colleagues. I want to bring your attention to the article on the bottom. That does a more thorough analysis than I presented in today's examples of how biomarkers can improve clinical drug development, but this represents, for us at least, an important critical path initiative.

Just to continue the context and continuity, the next talk you will hear is getting into the specifics of tipranavir. It will illustrate for you the data in drug development, translational challenges in moving information to the label, and it will give you an appreciation of some of the questions that we will talk about during the discussion, but I will pause at this point.

DR. RELLING: Time for a question or two.

I have a quick one, Larry. Janet had mentioned the idea of developing biomarkers in concert with drug development. What is your feeling on that? It seems to me, I can understand that it might be necessary to develop a biomarker and help validate it in the process of developing a drug or refining usage of a drug, but it seems that the biomarker method development should be something that can stand on its own, because many things are likely to be applicable to other situations down the road.

DR. LESKO: Right. Yes, I think in Dr. Woodcock's presentation, she made some distinctions between

the different types of biomarkers that could be thought of as tests or diagnostic tests.

I think at the highest level, which also corresponds to the highest level of evidence necessary, would be the tests that one might develop in a co-development context with the drug product, such that the drug can only be given to patients that test with that test. I mean Herceptin would be an example of that.

I think that is where much of the emphasis in her presentation landed, identifying those tests that you can use to maybe identify your responder group, classify them or stratify them for drug use, and if they don't have that classifier, then, they would not get the drug.

Conversely, people at risk, I think what I emphasized a little bit more are those biomarkers that represent tests that would have other purposes. The patient is getting the drug, it is not restricting the drug in any way, but that drug has inherent benefit and it has inherent risk, and how can you manage that benefit-risk ratio by the use of diagnostic and test information.

At its most basic level, we used to do this with drug blood levels. At its more sophisticated level, it might get into things like derived parameters like I showed with the inhibitory quotient.

So, I think there is a range of biomarkers and a range of testing that can have a whole bunch of different roles in patient care and drug development, and, of course, one of the things we will be talking about as time goes by is what is the evidence necessary for each of those cases, and it is going to be case by case, and I think as Dr. Woodcock mentioned, contact-specific for the disease that we are talking about.

But I think it is worth thinking about the structure around that, and I think that is what we are focusing on this morning.

DR. RELLING: One quick question?

DR. SINGPURWALLA: Would you like me to help you state my name? Nozer would be helpful.

I don't know exactly what to make of the talk that Dr. Woodcock gave and that you gave, but the impression I get is the talk is focused around advocating the use of biomarkers in the activities that you undertake. Is that correct?

DR. LESKO: I think in part that's true.

DR. SINGPURWALLA: Then, I would like to say the following, that it's pretty natural to me that one should use biomarkers or whatever you wish to call it to make predictions about whatever it is that you want to do.

The important question is how well they correlate with the end event that you are looking for, and so if the correlation is strong, the biomarker has value; if the correlation is weak, certainly, the biomarker has no value, but if you try to incorporate it in your labeling, you may be incurring some kind of a risk.

The other comment I would like to make is that the use of biomarkers is also prevalent in two other disciplines where you may want to look for sources of literature and ideas.

One is in economics. The economists don't call them biomarkers, they call them leading indicators, that is, they have something which does then which predicts what else is going to happen, and they are being managed to study correlations between two stochastic processes that are connected with each other.

The other arena where you may find useful information is in engineering. They call it degradation modeling, because what they want to do is they want to look say, for, example, at the wing of an aircraft which has very, very small cracks, and those cracks are predictors of wing failure. Every wing has a little crack, but, you know, when the cracks become very big, the wings fail.

There are some very sophisticated techniques by which they do those analyses and correlations, and what I

would like to suggest is that you may look into that source of literature for possible ideas and for possible future development.

DR. LESKO: I think that is an excellent suggestion. I thought you were going to say the military applications of modeling and simulation which also have their own intents, but one point you did mention, and this is something that has not been discussed thoroughly, is where on that predictive scale does one lie for a particular test.

That is to say, how do I quantitate predictiveness, and you will hear a little bit about that from Steve later on this morning, but what is the appropriate metric for predictiveness of a biomarker, and where is it for its intended use, and that will be an interesting discussion and one we haven't really had a discussion of. That might be another topic for framing in the future get together.

DR. RELLING: Thank you. I think we should move on. Now, we are going to hear from Dr. Douglas Mayers from Boehringer Ingelheim Pharmaceuticals on the Use of Biomarkers in Clinical Development and Labeling: An Industry Perspective.

Use of Biomarkers in Clinical Development

and Labeling: An Industry Perspective

DR. MAYERS: My name is Doug Mayers. I am the Therapeutic Area Head of Virology at Boehringer. I coordinate antiviral drugs from in licensing through postmarketing phase medically for HIV and hepatitis C.

I want to give a little story to illustrate one of the problems we are going to face today. I was talking to Dr. Lesko before the meeting, and we were talking about our travel schedules, and I related that my wife thinks that time averaged, I live over Iceland, but I am never actually in Iceland, I am either in Frankfurt, Germany, or I am in the United States, and I think one of the problems you get into with some of these models that we will talk about.

[Slide.]

I want to thank Dr. Lesko for the opportunity to talk with you today about biomarkers from a clinical development and labeling with an industry perspective.

[Slide.]

I am going to do a general overview, then, talk about the use of biomarkers in our tipranavir clinical development program where we had a very rich dataset of data, talk about use of biomarkers in labeling, and then have a discussion of philosophy on TDM to optimize individual patient therapy and the requirements that we

believe will be needed to use TDM with an antiretroviral drug.

[Slide.]

As Dr. Lesko showed you earlier, there has been a definition of a biomarker that has now been carried through a number of documents, a characteristic that is objectively measured and evaluated as an indicator of normal processes, pathogenic processes, or pharmacological responses to a therapeutic intervention.

This isn't a novel concept or a new concept. Biomarkers have been used extensively in the past, but I think it is being much more aggressively pursued now that we have a lot better understanding of molecular level of many of the disease processes that we are developing drugs to treat.

Biomarkers are often used in label to describe drug mechanism of action or describe target populations of patients, and some common examples would be using the IC_{50} or EC_{50} against a drug target, hemoglobin A1c for diabetes, and HER2 receptor or CEA-125 for oncology indications.

[Slide.]

When you move to a surrogate endpoint, this is a biomarker intended to substitute for a clinical endpoint, it is often used in label to document efficacy and requires a prior validation process, which we will talk about in a

second, and some commonly used examples, which I pulled out of the FDA documentation were blood pressure for antihypertensives, cholesterol for statins, for cardiovascular indications, prothrombin time or INR for coumadin, and HIV RNA or CD4 cells for antiretroviral drugs.

[Slide.]

In HIV, we have two surrogate markers that have been validated, and I think they provide a useful model for the correct process to validate endpoints. This occurred about a decade ago.

It was a collaborative effort involving multiple academic groups, industry, the NIH, and FDA through a series of meetings, and had a number of requirements. They required widespread availability of quality assured, well characterized, quantitative assays.

At that point, we had two assays that were available, we knew their variability on an assay basis, we knew their variability on a patient basis. We had well characterized their properties.

The assays were biologically plausible and the results were related to the natural history of HIV disease in several natural history cohorts, so the CD4's and viral loads were predictive of progression to AIDS and death.

Changes in these surrogate endpoints were then related to clinical outcomes, AIDS progression or death, in several large clinical trials of antiviral drugs, so we then showed the intervention which altered these parameters was associated with a clinical outcome.

No single sponsor or group could have done this. It took the data from all these groups put together to get a convincing database that CD4's and viral load measurements were appropriate to use for drug development and for patient management.

[Slide.]

Some of the uses of biomarkers in clinical development. I think the really exciting area that we are seeing a lot of advances right now is in the preclinical arena where models are being used to develop predictions of drug toxicities in man and biomarkers that can be carried across from the animals to human, for example, for hepatotoxicity, are being developed.

Animal models that predict efficacy in man, developing biomarkers that would be carried across the animal model into human clinical trials is also moving forward briskly.

In the clinic, we use biomarkers extensively. We use them for dose selection in our initial human studies, for target validation and proof of principle, to select the

patients for our trials, and I will show you how we use this very specifically in the tipranavir development program to put patients in the trials where we thought they would get the maximum benefit.

Phase II/III trials, endpoints, primary and secondary endpoints, for clinical monitoring, treatment guidance, and ultimately, hopefully, for individual patient dose adjustment in some instances.

[Slide.]

The critical issue, as was mentioned earlier, is what is the link between PK/PD and how tightly is that link correlated for what the therapeutic utility would be for drug level monitoring.

I think the ideal circumstance is a direct relationship where the plasma concentration is directly related to pharmacologic effect, such as a drug on a specific receptor.

More commonly, and especially in HIV, we see indirect relationships where plasma drug concentrations are related directly to a peripheral compartment, for example, with the nucleoside AZT or D4T, it is actually the intracellular triphosphate that matters, not the blood level in the patient at a particular point in time.

[Slide.]

As you start to look at sources of variability, the variability in the dose we give the patient is usually quite small, and the FDA makes sure that it remains quite small.

As you move into measuring that drug level, you have assay variability that is under 15 percent if you use internal standards and quality control.

When you move into the clinical trial and you start looking at individual patients with a given dose, you will see variability gets up to around 50 percent, and then as you move from the PK to the PD, the variability can increase up to 100 percent or more and outcome data is highly variable, requiring large trials and potentially search for biomarkers.

So, this just sort of shows the cascade of uncertainty as you move from a dose and an assay into the individual patient management.

[Slide.]

Moving to tipranavir's clinical development program, HIV RNA and CD4 cell counts have very clearly accelerated HIV drug development and have allowed us to get drugs to the patient quickly and in an efficient manner.

Tipranavir's Phase II/III program was a data-rich source for biomarker and drug level data. In this program,

all patients had baseline genotypic resistance for the HIV virus, viral load measurements and CD4 cell counts.

All patients had several C_{min} determinations determined during the course of both the Phase II and the Phase III studies with serial viral load and CD4 data.

In the largest resistance database to date, approximately 860 patients had baseline phenotypic drug resistance determined for their isolates, which allowed us to then relate the genotypic resistance in the virus's enzyme sequence to the phenotypic resistance.

Tipranavir C_{min}, baseline genotype and phenotype, inhibitory quotient were all related to viral load response in our Phase II and Phase III studies, and the results were used to design our clinical trials.

[Slide.]

Just quickly showing you tipranavir. Tipranavir is a novel nonpeptidic protease inhibitor, which was actually developed as a new treatment option for highly treatment-experienced patients or patients with virus resistant to multiple PIs.

So, it was actually the first protease inhibitor developed specifically to target drug resistant virus. It has potent in vitro activity against HIV-1 and 2, and the majority of multiple PI-resistant HIV viruses seen in the clinic.

[Slide.]

Looking at PK, the dark blue curve is tipranavir when it was administered without ritonavir, and we could not get high enough drug levels for long enough duration to achieve the targets we needed in the clinic, so tipranavir was then combined with ritonavir, which gave a significant boost, which is the light blue curve.

Now you can see that we get 9-fold greater exposure, and a 48-fold increase in C_{min}, so that we were able to get well above the concentrations needed to inhibit virus in the clinic with twice-a-day dosing with ritonavir.

[Slide.]

Looking at our early Phase I study to show this drug had had activity, in 14 days with an HIV drug, you can do very efficient dose ranging. The light blue curve shows about a 0.7 of a log response when tipranavir was given as 1,200 milligrams without ritonavir twice a day.

The purple curves show that you get about 1.7 log response when tipranavir is given with ritonavir, and we move forward in clinical development with the drugs given together thereafter.

[Slide.]

We had two large Phase II studies which were inconclusive as to which dose we needed to take into our Phase III program, and so we did a three-dose study of

tipranavir at 500 and 100 ritonavir, 500 and 200, and 750/200 twice a day, in 216 patients who were highly treatment experienced with three classes of antiretroviral drug treatment and two prior PI regimens at least.

This was an innovative study in that we looked at the functional monotherapy viral load response in these patients at 14 days for efficacy measure, and we looked at toxicity measurements at 8 weeks because most of the toxicities with tipranavir occur early on in treatment, and so we were able to very quickly dose range and find that 500/200 mg was our optimum dose for Phase III.

The 500/100 dose was eliminated because it was not as active a drug as we needed for the highly resistant viruses that we were going to be treating in our Phase III program, and it had a little bit higher PK variability.

The 500/200 and 750/200 doses had similar efficacy and PK profiles, but the 750/200 mg dose of tipranavir/ritonavir was eliminated because it had higher rates of ALT/AST elevations and higher treatment discontinuations, and it didn't appear to be as well tolerated as the 500/200 dose.

We subsequently are evaluating the 500/100 and 500/200 doses in treatment-naive patients where we don't need to attain as high drug levels potentially.

[Slide.]

During that Phase II program, we also looked very carefully at drug resistance, and we specifically looked in the HIV protease enzyme at four positions, Position 33, 82, 84, and 90, which had either been selected in the test tube when we grew out drug-resistant virus, or were seen in our patients who were failing on a tipranavir-based regimen and had decreased susceptibility to tipranavir in the clinic, so we found these four mutations to be predictive of resistance to tipranavir.

In our Phase II program, we confirmed that multiple mutations at these sites were associated with decreased responses to tipranavir, not surprisingly, but the surprise was they were also associated with broad, high-level resistance to all of the other available protease inhibitors, saquinavir, indinavir, lopinavir, and amprenavir.

As an example, if the patient's virus had three mutations at those positions with lopinavir, there was 100-fold resistance to lopinavir in the test tube.

So, we used these mutations to select patients who were unlikely to get a durable response to any single PI-based regimen who were offered a dual-boosted PI regimen containing tipranavir.

So, if the patients had less than three of the key mutations, we put them into our single-boosted PI

pivotal trials, and if they had more than three of these key mutations, we put them into a dual-boosted PI Phase II drug interaction study, and if that study had been effective, we would have moved to a third pivotal trial. Unfortunately, drug interactions prevented us from doing the third trial.

[Slide.]

This just shows our Phase III program. We had two very similar studies, RESIST-1 in North America and Australia, RESIST-2 in Europe and Latin America, looking at tipranavir versus the best alternative boosted protease inhibitor, and for patients who screened into these studies, and had higher levels of resistance, we allowed them to go into the dual-boosted Companion Study, where they got tipranavir/ritonavir combined with either lopinavir, saquinavir, or amprenavir to try and get more activity for this group of patients.

[Slide.]

Looking at efficacy.

[Slide.]

In our pivotal trials, it was very clear that tipranavir expected as we had hoped with a 1-log viral load reduction at 24 weeks was our primary endpoint, 41 percent of patients had this with the tipranavir arm, but 19 percent had with the comparator arm, and as you can see, we

had significant p values for all the viral load in CD4 measurements.

With AIDS progression, we had a trend toward a benefit, but it was not powered or sized to have an AIDS progression benefit in that study.

[Slide.]

We then did the most extensive evaluation of phenotypic and genotypic drug resistance that has been conducted to date.

Basically, we took 291 baseline genotypes from our Phase II program and looked at all 99 amino acid positions in the protease, and related all 99 amino acids in a univariate and in a multivariate manner to tipranavir phenotype, viral load reduction at two weeks and at 24 weeks to get a score that was predictive of resistance and lack of response.

We then took this score and evaluated it in our Phase III program. We had 569 baseline genotypes with phenotypic data to confirm the relationship, so we had essentially 860 patients with genotypic and phenotypic data related to phenotype, drug response early and late in treatment.

We came up with a score that predicted reduced tipranavir susceptibility or reduced responses, that had 16 amino acid positions and 22 mutations, and simply you just

add up the number of positions that have a tipranavir-associated mutation, and that gives you the score.

We don't expect the clinician to do this, but this is available to the diagnostic companies who can then provide it to the clinicians for guidance.

[Slide.]

This is the model that we think most accurately predicts responses to tipranavir in the clinic. Basically, if you get tipranavir/ritonavir, it gives about a 1.25 log response at 24 weeks.

If you add T20, which is the most potent additional drug we had available at that time, you get about an additional log of response.

For each available drug in the optimized background, which means the resistance test said it was either sensitive or partially resistant, you got about a quarter of a log response, and when you added those together, that told you roughly how much viral load response you get to the drug.

Then, the tipranavir score per mutation decreased that response by 0.17 logs, so the higher your tipranavir score, the lower your ultimate response.

The real challenge with tipranavir, it is not that it isn't an active drug, it is a very active drug, it is finding the second and third drug to form a regimen that

will completely suppress the virus, so the real challenge, and we are working with other companies right now, is to find additional drugs that are active in this population to combine with tipranavir.

[Slide.]

Looking at two-week viral load reductions by drug concentrations, you can see that if you have a drug concentration above 6.5 micromolar, that you get a 1-log response, and there really isn't, when you get above that, a clear trend for increasing response once you get over that level.

That means that 95 percent of patients had a drug level that would give them a 1-log response at two weeks.

[Slide.]

Unfortunately, when you start to look at 24-week viral load responses by tipranavir concentrations, the relationship and correlation is very, very weak, and I am not sure how I would use that to manage a patient.

[Slide.]

Moving to inhibitory quotients, this is the C_{min} of the drug over the protein-adjusted IC_{50}, this gives you a sense of how much of a barrier to resistance you have of your drug, and how much over the minimum required amount you have, and we adapted this for the currently available measurements, because what the clinician gets from Virco

[ph] or Virologic is a fold change, so the equation we put out was C_{min} over 3.75, which is the protein binding correction factor between 100 percent human serum and 10 percent fetal calf, which is what is used in the in vitro assay, 0.058, which is the median IC_{50} of a wild-type virus in the Virco assay, and the fold change for the IC_{50}, and the clinicians can get the C_{min} and the fold change from laboratories outside the clinic.

[Slide.]

Looking at 14-day responses in our Phase II monotherapy study, and we think this is probably the most accurate reflection of the drug, because is where tipranavir was given as the only new drug when we switched the patients, you can see that when you had an IQ above 30, that there was a 1-log response, and if you were below 30, you clearly had a suboptimal response to the drug.

Of interest, an IQ of 30 correlates very nicely with a C_{min} of 6.5 micromolar when you do the conversion.

[Slide.]

Looking at 24-week data, here you have in olive the response data when tipranavir was given without T20, and in dark blue when tipranavir was given with T20, what you can see is there is a clear dose-response the higher the IQ, the more effective the drug was.

In olive green, it's almost a continuous relationship because in many of these patients, tipranavir was the only active drug they were receiving in their regimen, but again, you can see there is very large inter-quartile variability.

When you go to the green, which is blue on my projector, that is when you add T20, so this says what do you need when you have a potent second drug to combine with tipranavir.

Here, you can see it looks like you need an IQ of approximately 60 to get a durable response with a second active drug, with T20 being the example of that drug, but again there is a very, very wide variability in individual patient responses.

[Slide.]

Looking at safety data.

[Slide.]

It was clear that there were two areas where tipranavir has safety that needs to be monitored in the clinic. The first is elevation of AST/ALT where there was a 9 percent increase in ALT with tipranavir versus 2.3 percent increase with the comparator, and cholesterol and triglycerides where again there was an excess of cholesterol and triglyceride elevation in the tipranavir arms of the study.

[Slide.]

There is a weak trend. As was shown previously, between 20 and 120 micromolar concentrations, there is a weak trend for increasing ALT/AST elevations although there really isn't a significant jump until you get above 120 micromolar where you actually see that 45 percent of the patients now had an elevation, a Grade 3-4 elevation, but again this is only 1.5 percent of the patients in all the clinical trials had that high level of tipranavir concentration.

[Slide.]

When you looked at individual patient data, again, you can see there is a very, very extensive overlap of essentially the same median value between those who had an elevation and those who did not, and the majority of AST/ALT elevations actually occurred in the range in which you actually target the drug, because that is where most of the patients were.

[Slide.]

So, our conclusions were that tipranavir trough levels above 6.5 micromolar were associated with a greater than 1-log response at 2 weeks, showing the activity of the drug.

Tipranavir trough levels greater than 120 micromolar were associated with increased risk of AST/ALT

elevations. That meant that 93 percent of patients had tipranavir trough levels that would produce a 1-log reduction, and did not produce excess risk of hepatotoxicity, significant excess risk.

There were weak trends associating tipranavir trough levels with hepatic events and treatment responses, but the large inter-patient variability could limit the utility of these measurements in practice because they were not predictive of either safety or efficacy in the individual patient level.

The exception is the potential use of TDM to optimize dual-boosted regimens where we give tipranavir/ritonavir with a second protease inhibitor, and you have to make adjustments to try and get an optimal dual-boosted regimen for the clinicians who are trying to use these regimens, because we cannot as a company make a recommendation that says use this amount of the drug with tipranavir. So, these are being explored in some pilot studies especially in Europe.

[Slide.]

When you look in the package insert, we ended up having a descriptive label of the PK/PD relationship in which we basically took the median IQ, inhibitory quotient, was 75, and showed that if you had a lower inhibitory quotient, you had a lower response rate, if you had a

higher inhibitory quotient, you had a higher response rate, and showed how T20 improved those responses in the patients.

It was specifically noted that these IQ groups were derived from a select population, were not meant to represent clinical breakpoints.

[Slide.]

So, where are we with biomarkers and HIV drug development? I think that HIV drug development has truly benefited from the validated surrogacy of plasma HIV RNA and CD4 cell counts both for drug development and for patient management.

Phenotypic and genotypic drug resistance testing have been validated in multiple prospective studies and integrated into clinical drug development and patient management in the DHHS guidelines. They are well characterized on the tipranavir label.

IQ calculations are used to determine our target drug levels in our Phase I. We try and get an IQ that will be above the first resistant variance to our drug, and then we clinically confirm these IQ measurements in our Phase II/III trials.

All the companies I think are moving toward genetic assessments to look at PK variability and the risk of drug-specific toxicity. The best example has been with

abacavir with toxicity and there is emerging data with nevirapine, and we have a large genetic study looking at nevirapine early events, as well.

[Slide.]

Use of biomarkers in labels. I think they are already well integrated into product labeling at mechanism of action, patient selection, dose selection/adjustment, and clinical monitoring for both safety and efficacy.

Validated surrogate endpoints are very clearly integrated into our current labeling especially with our HIV and hepatitis C drugs where they are actually used as the clinical endpoints in our trials.

[Slide.]

Moving to TDM to optimize clinical management. Therapeutic drug monitoring to adjust doses to optimize patient management is far more common in Europe than the United States, and I think that we have been thinking about this a lot as to why we see this difference, and probably it's because in Europe, many of the academic pharmacologists actually are integrated into the clinical units and can provide rapid turnaround levels in a day to their clinical colleagues, and have an ongoing dialogue with their clinical colleagues, which is a relationship that many of us do not have in the United States.

You have to have a correlation between the drug concentration and clinical effect and it has to be, the tighter it is, the more useful TDM will be.

Adjustment of drug levels must result in a clinical benefit, and I think this is a major issue because as I will point out later, adjusting drug levels can result in clinical harm.

Use of TDM requires a widely available, rapid turnaround, quality assured, clinically validated drug level assay with a clear algorithm for dose adjustment, and we have these. In antiepileptic drugs, these are very clearly monitored with drug levels, aminoglycosides, we all use them in the hospital regularly, and antiarrhythmic drugs.

[Slide.]

Moving to antiretroviral drugs and IQ measurements, it is much more mixed picture. There have been two prospective studies looking at the use of drug level monitoring and TDM in HIV infected treatment-experienced patients, and both of them came up with negative results. There was no clinical benefit to TDM in those populations.

There has been a study that showed benefit. It was in treatment-naive patients with unboosted protease

inhibitors, and that is the one study that I am aware of that has shown benefit.

There are lots of challenges at this point. There is significant variability of drug levels between patients. There is an absence of standardized drug level measurements. We don't have the laboratory infrastructure to provide these measurements at this time, especially in the required time frame which is going to be 48 to 72 hours to make it meaningful.

There is an absence of consensus on target drug concentrations or inhibitory quotients. We are beginning to get a consensus for treatment-naive patients, but we really haven't developed the same consensus for treatment-experienced patients on what level target you need to have.

Often TDM is only performed on one drug in a three or four drug regimen, so you can't optimize the whole regimen, which I think is the goal of therapy, and then we get to the most important issue, which there currently is a 2- to 4-week lag in obtaining drug levels and drug resistance results in the clinic.

The problem is that HIV doesn't stand still. The HIV virus was designed to escape the immune system and it uses that same mechanism to escape drugs, so the virus makes a swarm of viruses. It makes continuous mistakes trying to get around the pressures it sees.

It has a doubling time of one day, so basically, if you have a drug that has a single point mutation that produces high-level resistance, and an example of that would be nevirapine and lamivudine, in 14 days, you can have the whole virus population turn over and become drug resistant.

With the protease inhibitors, by 28 days you can see partial resistance that has emerged, so the problem is the virus that you got your IQ measurement for at the beginning is not the virus you are trying to treat if you haven't had an effective regimen.

So, by the time you get the data back right now, it is simply too late, and I think that is why both the prospective studies that have been conducted did not have a positive outcome.

[Slide.]

So, what do we think will be the requirements for labeling therapeutic drug monitoring? I personally believe biologic plausibility is not a high enough standard. TDM requires you have a certain level of infrastructure, widely available assay with a rapid turnaround, quality assurance, quantitative drug level measurements.

The drug must have a large enough safe and effective range to allow reasonable dose adjustments. In tipranavir, we only have significant data on the 500/100

and 500/200 dose. What do you do when you give a drug dose that is outside the range that has been tested in your pivotal trial program because of TDM measurements?

There has to be an algorithm for drug level adjustment that must be prospectively validated, and most importantly, must demonstrate a clinical benefit. Since dose adjustment can introduce new toxicities when you increase the dose, you can see increased hepatotoxicity potentially.

It can result in loss of therapeutic effect because you won't go from a dose that was effective to a dose which now can allow breakthrough of the virus, and perhaps most importantly, it can decrease adherence, and there is actually data from an ACTG study in which one arm continued the same dose, and the other arm allowed dose variations by an algorithm, and the patients actually had lower adherence because if my doctor can change my dose based on data, why can't I change my dose based on how I feel, and you have to have 90 percent compliance with these drugs for your lifetime or you break through, and these aren't like cardiac drugs.

When you go back and you take them right the second time, they no longer work because of drug resistance. So, we think there needs to be a clinical benefit before we move this into the clinic. This level of

proof does not exist for the majority of drug including antiretroviral drugs in clinical use today.

Thank you.

[Applause.]

Committee Questions to the Speaker

DR. RELLING: Questions? Go ahead, Dr. Gage.

DR. GAGE: You mentioned that in the higher doses, there was hepatic toxicity and that you will be studying potentially genetic predictors of that.

Suppose that the FDA had a voluntary program where industry had archived DNA for patients on protease inhibitors that had had hepatic toxicity, would that type of resource be useful for you when you conduct these studies?

DR. MAYERS: I think that we are all moving toward getting banks. Unfortunately, in the tipranavir program, we didn't. We will get genetic data from some of our post-approval studies to get that type of a bank available.

I think most of the companies now, in the antiretroviral area at least, are collecting cells from Phase II onward, specifically, to have those banks available to look at those types of questions, so, yes, they would be useful.

DR. GAGE: So, then the corollary is do you think that your company would be willing to supply DNA on a voluntary basis from consenting patients to do these nested case-controlled studies that require adequate sample size to predict genetic based hepatotoxicity?

DR. MAYERS: I think we would be willing to participate in the process. I am only an international head.

DR. RELLING: Dr. Jusko.

DR. JUSKO: That was a very nice presentation and results from a complex situation. I wanted to ask you about how much you knew ahead of time when you designed these studies. It sounds like you already had a good indication that the IQ, the C_{min} over the IC_{50} ratio was going to be relevant, and you designed your studies accordingly because you only collected C_{min} samples.

With a lot of anti-infectives, people are concerned that sometimes C_{max} or the AUC versus the C_{min} might be relevant, and you didn't have that opportunity to assess that question because you didn't collect the relevant data.

On the other hand, we heard a presentation yesterday that the FDA people utilized a mechanistic model that sort of confirmed that what you did at a certain point in time seemed to be consistent with basic principles of

turnover of the viruses and the rest, and that helped your case along the way.

So, returning back to my original question, did you know this ahead of time, or did you need the FDA's help to confirm that your empirical choice was the right one?

DR. MAYERS: We always like to collaborate with the FDA, but I think in this particular instance, there is a very good body of data that as long as you stay above the IC_{90} of the virus, you completely suppress and doesn't replicate, and as you drop below that, it does start to replicate, so at least there is a sort of general consensus that for non-nucleoside agents and protease inhibitors, you want to get a target C_{min} of some fold above the level you see in the clinic, and then the clinical data tells you whether you guessed right or whether you need to go higher or lower based on your viral load response data.

But it gets a little more tricky with the nucleosides where I think the general consensus is you need an AUC exposure as opposed to a C_{min} exposure. We did get pop PK data, though, so there are patients where you do have population PK in our studies to try and get some of the other parameters you were talking about in these studies, but in general, we believe that C_{min} and inhibitory quotient based on C_{min} from what we have seen across a number of these drugs would be the right measurement although

whether that is as firmly established as people would like could be debated.

DR. RELLING: Dr. Flockhart.

DR. FLOCKHART: This was a really nice presentation and an interesting series of studies.

This is a question based in ignorance, but it gets at the incremental values in drug monitoring in general, but simply in this kind of situation. What is wrong with just following the viral load?

I understand if you just measure the viral load, you don't know if it's resistant or not, because you haven't got a genotype of the virus, but could you just kill that one?

DR. MAYERS: Actually, when we have had these discussions as to how you would best try and adjust doses if one were going to try and do it, I actually think that viral load and ALT are probably the best parameters we could use for dose adjustment rather than using TDM, because if I give the drug and in 7 days the virus hasn't gone down, I don't have enough drug. If it has gone down, it goes undetectable, I have enough drug.

So, I think one could argue that if you wish to try and tailor therapy, that basing it on viral load and CD4 would actually be a more efficient and more likely to

be successful way of doing it than using drug levels per se.

DR. FLOCKHART: And viral load comes back quickly, what is used, and you don't have to create a new infrastructure like clinical pharmacologists in hospitals to do it.

DR. MAYERS: Yes, I would agree with you.

DR. RELLING: Can I just expand? I mean the reason to do TDM may be in addition to viral load would be so you adjust the right drug. Don't all these patients receive more than one drug?

DR. MAYERS: If you knew how to adjust the other drugs in the regimen, potentially, it would be of use. I mean the real problem right now is the time frame. I mean we have got to get this time frame down to 24, 48 hours if you are going to have any chance of success.

So, the doc out there can't get these tests back in a time that is relevant to the virus, so that I think that if we knew what targets to hit, and we did several prospective studies that showed that hitting these targets provided a clinical benefit, and the French have tried twice, unfortunately, not successfully, then, I think one could argue that there was a utility to adding this to clinical practice.

In the one study in which they used drug resistance measurements, and they used drug concentration measurements, drug concentration adjustment did not provide a benefit, adjusting your drugs based on drug resistance did, and so we think at this point, for what is available to the doc, we have the right tools for them to optimize their regimen.

DR. RELLING: Dr. Sadee.

DR. SADEE: I also was wondering about the use of multiple other drugs, and if you use ritonavir, I did not see any mention that this is a boost because it inhibits 3A4, MDL1, and so on, and so on.

So, shouldn't we then also consider what happens to statins, and shouldn't that be all specified in the label, because all of a sudden, you are going to non-metabolizers for a number of different enzymes that would raise flags for all kinds of different drugs?

DR. MAYERS: Yes, we actually have done extensive drug-drug interaction trials, and worked with the agency, so that we have labeled all the data where we have the drug interactions, the tipranavir/ritonavir combination, because it gets a little tricky, because the ritonavir completely inhibits hepatic and intestinal 3A4, but in combination, the tipranavir induces PGP in the gut, so that you can get some funny interactions.

But we actually have worked closely with the agency to label all the known interactions. We have looked at statins, and then we also have gone through the theoretical bases and informed the clinician where we believe the drug is likely to go up or go down and where we cannot tell where we recommend clinical guidance.

So, I think we have a very extensive drug-drug interaction label where actually, there is a post-approval commitment now going to do a cocktail study to fine-map the net effect of tipranavir/ritonavir on each of the individual hepatic CYPs, which would then allow us giving more informed label for the practicing doc.

DR. RELLING: I think the last question, Dr. Barrett.

DR. BARRETT: I really appreciated your presentation and working through the evolution of the biomarker work, but I couldn't disagree more with this last slide.

I think we are well beyond biological plausibility, and I don't see how you can walk through that evolution in terms of using this biomarker to justify dose and as the basis for an approval, and then step away from it on the TDM side, particularly the argument of compliance and adherence, this is very circular.

I mean one of the values I think of TDM is that it would give you a tool to assess patterns of noncompliance, and the issue of the doc can change my drug, so can I, I don't see that as being a clear identifiable pattern in this population.

I think the other point on the analytical side, you know, there are several other instances in which point of care tests become available when TDM has been shown to be of value, and I would think if this moves forward and we do get the compelling data, I am agreeing that I don't necessarily feel we have the data we would like to have in order to implement it uniformly, but I think, you know, as you point out, it is complicated, but it is not unsolvable.

DR. MAYERS: I would agree although as I said, the two prospective studies that looked did not work, and with resistance, we have multiple prospective studies that did show a clinical benefit.

So, I think because we have a dynamic situation with changing virus, that we need to develop enough knowledge to implement it in some prospective studies, show we can produce a clinical benefit that is meaningful, and then I would agree then it would be time to move into the clinic.

DR. RELLING: Because Dr. Mayers may not be here, we are going to go ahead and take two more questions. First, Dr. Capparelli, and then Dr. Lesko.

DR. CAPPARELLI: I appreciate the presentation, I really enjoyed, as I think Jeff mentioned, going forward with really trying to understand the components, and especially the multivariate approach and looking at these as complicated patients, multiple drugs.

But you are not going to get an approvable TDM study. Even in the naive population, you may be able to get it, but you are going to take thousands of patients, and there have been presentations that have looked at trying to power these things.

The other value in a lot of the points that you brought up besides what was mentioned before really are logistic issues. A 48-hour turnaround time, you need a quick one, but often the sensitivities may not be back in that time. You are not going to be at steady state because you have got induction going on.

So, I think, you know, the follow-up, as you mentioned, in terms of different approaches may be some of these questions of utilizing different techniques of monitoring in some of these multiple drug studies.

It would have been really nice knowing this population that you are targeting for this drug to have had

more information on the 750 dose, because, you know, even if you have got a higher resistance, even if you aren't doing TDM monitoring with drug levels, there is going to be that temptation to go up on dose, and without having some information and having some comfort level at least on the safety side, it puts us a step further back in where we want to take this therapy.

DR. MAYERS: We actually probably have about 120 patients at or above 750, so I know in a rough way what the issues are, but you are right. I mean I think one of the challenges is going to be if we are going to propose to use different drug levels, then, we are going to have to establish a large enough Phase II/Phase III database to support the safe use of those drug levels in some manner.

But we are going to be doing a pilot with both Europe and the FDA, hopefully, the same pilot with Europe and the FDA, looking at TDM measurements with tipranavir, so we will be probably coming back with more data.

DR. RELLING: Dr. Lesko, last question.

DR. LESKO: Doug, thanks. The question I had is that a lot of the data that one might imagine being used in TDM was actually used in the drug development program to make decisions, whether it was to explain variability and outcome, select a dose, interpret a drug interaction.

In fact, as I recall in this label, there were something like 21 or 22, or something like that, drug interactions, most of which had no specific recommendation other than arrows going up and down to reflect the area under the curve.

So, I guess what I am trying to get to is the perspective you have on relative value of information, and what you have shown is that we don't have a perfect case for monitoring either blood levels, C_{min}, or inhibitory quotient.

On the other hand, you have a clinical situation that is characterized by significant inter-patient variability in response to a select dose, and I just wondered if it isn't a case of more information being helpful at least to point a direction towards where to move with a dose in the face of a clinical outcome.

For example, when you have all these drug interactions in a patient setting, how do I know from the label when the arrow goes up or down, what am I supposed to do, adjust the dose, change the dose, try another drug, and without some more indicative information about the direction of my decision, how does a physician deal with that?

DR. MAYERS: I think for better or for worse, most of the truly significant interactions, the ones that

are contraindicated, the ones that produce serious toxicity are 3A4 interactions that are driven by the ritonavir component, so thankfully, the providing community has learned how to give ritonavir-boosted protease inhibitors in combinations with multiple other drugs by trial and error, if nothing else, so we benefit from that large accumulated knowledge of how to use this.

For specific drugs, especially for protease inhibitors, if you want to use a second protease inhibitor, I agree with you, there is no way to predict, and if you are a doctor who wants to try and use that type of regimen, and you have access to TDM, that might be a place where you would want to use it.

DR. RELLING: I thank the speakers and the participants. We are going to shorten our break, so we will start back up at 11 o'clock, please.

[Break.]

DR. RELLING: We are going to hear from Dr. Gutman from the FDA, the Office of In Vitro Diagnostics, the Center for Devices and Radiological Health.

He is going to talk on the CDRH Perspective on Analytical and Clinical Considerations that go into an FDA Approval of a "diagnostic test." A Presentation of Case Studies.

CDRH Perspective on Analytical

and Clinical Considerations

DR. GUTMAN: Good morning. As Dr. Woodcock suggested, that while you are playing with lab tests in the course of drug discovery, FDA has some interest, but not a passionate interest in the well-being of that test, but, in fact, if a test is going to accompany a drug to the marketplace, and is going to be used for decision-making, then, FDA's interest becomes more intense, and if that is a diagnostic that is going to be sold in interstate commerce, across multiple labs, that test actually becomes subject to scrutiny by you get two for the price of two by a second center, that being the Center for Devices.

I represent the Center for Devices. I appreciate the opportunity to be here. I will provide you with a paucity of facts and figures in this talk. It will be broad and structural, and, as always, I appreciate the fact that Larry put me last since I like to have the last word.

[Slide.]

FDA has been regulating medical devices in general and in vitro diagnostic devices in particular since the Medical Device Amendments were passed in 1976, and those amendments put into place a variety of general controls on medical devices including the requirement for registration and listing of the devices for Good

Manufacturing Practices and for medical device reporting of adverse events.

As a consequence of those general controls, the first time in history in this country, probably for the first time in history on the globe, there was an actual menu of tests on the market, a list of tests on the market.

There were mechanisms for ensuring manufacturers made products consistently over time, and there was a system for identifying real world adverse events, so that FDA could collaboratively work with companies and hold them responsible for fixing whatever had gone wrong.

[Slide.]

Those amendments also introduced for the first time, the requirement for premarket review of new diagnostics, new versions of old tests were processed through a process called the 510(k) process of the law, and fundamentally, new devices were processed as premarket approval applications.

[Slide.]

While administratively, there are actually marked differences between those two processes, the heart and soul of lab tests remains the same no matter how you look at them, and as Dr. Woodcock suggested this morning, you are not driving with gas, you are not cooking with gas until you have analytical performance.

So, as a starting point for all tests, one needs to be certain that you know what you are measuring and that you are measuring it reasonably well, that you get the same measurement if you decide to measure the same sample more than once, or as you are measuring over time, space, and analytical systems, that there be analytical, there should be specificity that you know interferes with the test, and that in some cases, you may be interested in limits of detection or measurements of performance.

For certain tests, in fact, the story stops here, in fact, for many of the analytes, perhaps for most of the analytes that my workgroup sees, this story stops here, because analytical performance is embedded in enough clinical history and standards of practice that that does the trick, so if we had a new test for hemoglobin, in fact, we would want to analytically characterize the hemoglobin. We would not start asking sponsors to demonstrate that hemoglobin was associated with anemia. That would be, of course, preposterous.

[Slide.]

But if a test is not so well pedigreed, if an analytic link to clinical behavior is, in fact, less certain, then, whether we have a 510(k) or a PMA, our workgroup is likely to start asking nosey questions about clinical performance.

The gold standard for doing that is to establish a gold standard, some kind of a yardstick for truth, and to determine how the analytical performance of that test compares, how the signal compares to truth and to characterize performance in terms of clinical or diagnostic sensitivity or specificity or, if you want to humor Larry, then, we would probably develop a likelihood ratio, and when we can't find a gold standard, we will use a silver standard or a bronze standard or a lead standard, we will use what we can get, and we won't pretend that is sensitivity or specificity, and we won't pretend that you can get a likelihood ratio out of it.

What we will demonstrate is that we have some measurement of agreement with a silver, aluminum, or lead standard.

[Slide.]

Where my workgroup generally does not tread, where angels or devils do, in fact, fear to tread, is in what I think Dr. Woodcock was referring to as the qualification of the test and what I would call the clinical utility of the test, that we are very intent on having analytical performance well characterized, and we are very intent on having clinical performance well characterized or at least well understood.

We are certainly intent on understanding, our term of art would be instructions for use, would be able to create instructions for use, so that the user of the test actually understood how the signal could be plugged into his or her practice, or if it's an over-the-counter test, into his or her self-management, so it is very important that information be used, and it is very important that there be a possible value and benefit, and that, in fact, that value or benefit outweighs the risk.

But, in fact, we don't do outcome studies, we don't look at drug effectiveness, we don't look at any kind of treatment effectiveness, we don't look at the impact on morbidity and mortality of the new tests. What we do is define the test and then try and link it in some plausible way to a good outcome.

As Dr. Woodcock suggested, perhaps when you start making a selection of highly toxic drugs, that may have been enough for the Center for Devices, and in more plebeian terms, it may not be enough to sell the whole package to the agency.

[Slide.]

It would be my view as a biased clinical pathologist, that in the year 2005, there is no excuse for a poor evaluation of diagnostic methods. There is a very rich literature to draw from.

There are dozens of FDA guidances on virtually every aspect of method evaluation, and CLSI and the international standards organizations have collectively crafted dozens and dozens of actual guidances and standards, so there is lots of information on which to use on every aspect of design, every aspect of biology, and every aspect of statistical analysis of new tests.

[Slide.]

In fact, there are two very explicit roadmaps that have been published on the appropriate way to demonstrate performance of a new test.

The one I am most familiar with is the STARD work. You can go into Google or into Pub Med, you can type in STARD, and you will see that that is such a popular evaluative technique that it has been published in almost a dozen journals.

A newer initiative of action I only learned about yesterday is one being promulgated now by the NCI for credentialing of new cancer biomarkers, and that is an initiative called REMARK, and the lead author is Lisa McShane. It can be found on the NCI web page under her name, and is an extension of STARD. It is actually using the same principles as STARD, but the intent is to be a little bit more cancer marker specific.

[Slide.]

From the perspective of my workgroup, the bad news is as you move into the area of genomics, and if you become really daring, into the area of proteomics, we are dealing with very complex science.

You are dealing with a lack of material and method standards. You are dealing with very well recognized sources of data bias, verification bias, spectrum bias, and in terms of sample accruement, you are dealing with some peculiarities in informed consent requirements that separate FDA from NIH.

[Slide.]

The good news is that we have a very refined regulatory toolbox for dealing with new diagnostics. We have a pre-IDE program that is our code word for protocol review. It is one of the few things that FDA still does for free.

A company can submit their protocol, we will review their protocol, we will try and do it within 60 days and provide comments and even meet with the company if they are interested.

I characterize the pre-IDE as akin to a pop quiz except you give out the questions ahead of time and you let the sponsor argue with you if he or she thinks you have asked the wrong questions, and you try and negotiate those, and if the sponsor actually pays attention to the

questions, it means when the test comes in, you get 100, and it goes through quickly. Now, that is not always the case, but that is the idea.

We have expedited reviews, which mean that really cutting edge stuff can cut in line, go to the front of the line, and it is supposed to mean we still deal with it dispassionately.

That is completely disingenuous, because when we see fantastic new technology, how can you treat it dispassionately, but we do treat it fairly, so an expedited review, if it is a high quality submission, will result in a very rapid Yes, and an expedited review.

If it's a lousy submission, it will result in a very rapid No, and if it's somewhere in between, there won't be a rapid anything, we will just struggle for months and years, and try and get it right.

We have de novo classifications, which allows us to overcome the peculiarity in the law that would say that if you had a bandaid that had not been invented at the time the law was passed in '76, and somebody suddenly developed adhesive tape or a bandaid in 2005, by default, it becomes a Class III, high-risk device, and we have a way for dealing with that nuance in the law, and we have real time reviews, we follow the Nike tradition in a variety of our work products, which is to just do it.

[Slide.]

And the good news is that we have a congressional mandate to be least burdensome, to keep our questions focused on the regulatory, not academic thresholds.

We have user fees, so we have been able to hire talented young scientists, a whole cadre of people with expertise in genomics and proteomics, and we have the capacity now to educate old people like me, so that maybe we are not quite as out of it as we used to be, and we have a seasoned program where we increasingly understand the opportunity cost of delaying getting new diagnostics into the marketplace.

[Slide.]

We have experience with two metabolic enzymes, CYP-450 and UGT-1A1. They followed a standard pattern. In both cases, they were playing off of a de novo classification, the fact that we used a body of knowledge to mitigate risk and to suggest that we could process these administratively in a more streamlined way than if we had not had the de novo classification.

In both cases, there was a model of drug labeling that we could turn to, in one case, Strattera, and in the other case, irinotecan.

In both cases, the core review issues were focused on analytical data using diagnostic truth and using

extensive precision studies. In both cases, we were actually somewhat impressed by the quality of the ability of these companies to create certainty in analytical accuracy.

Then, in both cases, we actually defaulted in terms of our clinical assessment to the use of literature. There is a threshold problem in the use of the literature in that particularly for CYP-450, you do have to go through the various variations and decide which ones you believe are credentialed and which ones are great ideas, but aren't quite ready for prime time.

Actually, we sorted through the UGT-1A1 with some of the same issues, but the clinical literature did hit the spot in the case of CYP-450.

It's an old enough enzyme, I studied it when I was in medical school, and there are literally thousands, perhaps tens of thousands of hits, so it is not exactly an arcane activity, and then we labeled it for what it was with incredible transparency, so that, in fact, part of the ability to bring these two products to market was not their strengths, but their weaknesses, in fact, that the whole story may not be told by one enzyme, and if a physician doesn't understand that, maybe the problem isn't the test, maybe the physician needs a bit more education.

[Slide.]

FDA's mission is to promote public health and to protect public health is obviously a tension in the duality.

[Slide.]

We ground all of our work in IDE in good science. That means we try to focus our review and try to ask the relevant questions. It would be my view that if we do a good job at that, that even if you firebombed my workgroup, you couldn't get rid of the questions.

Thank you.

[Applause.]

DR. RELLING: Thank you, Dr. Gutman.

I wonder if there are any questions for Dr. Gutman at this point. Dr. McLeod.

Committee Questions to the Speaker

DR. McLEOD: That was great, Steve, thank you.

One of the difficult things is when these tests are developed, they are developed as very well credentialed devices, but their clinical utility is variable.

Whose job is it? I mean does there need to be more interaction between CDER and your group, or how do we end up with products that come out that are not only analytically high quality, but actually, are worth using, or is that just not an FDA mandate?

DR. GUTMAN: Well, it is just a difficult job, because of the heterogeneity of practices and because of the heterogeneity of choices that people can use to bring tests to market.

There actually is a name for an uncredentialed test. I guess Larry's group would call it not a valid biomarker, I forget the exact term. We would call it "investigational use device," so a test that was analytically well credentialed is perfectly legitimate to use.

You shouldn't pretend it's a clinical tool, so it should be used either blindly or if it is actually going to be used in patient management, it should be used under IRB with informed consent. The patient really deserves to be enfranchised enough to understand when he is getting a test that may have incomplete meaning.

There is an alternative path I didn't mention, but is a legitimate alternative path to market, which is home brew, which does not provide for an investigational phase. Home brews spring to life, it was like Athena, I think she just spring from Zeus's brain.

Well, home brews do the same thing, they just spring to life based on a report or sometimes they spring to life with more rest even. One doesn't wish to oversimplify, but CDER and CDRH are I think connecting

better. We probably still have a ways to go, and there are choices that clinicians get to make and researchers get to make.

DR. RELLING: I have one question. When you talked about bringing the approval of the P-450 and the UGT-1A1 test, and you stated that your common approach that starts with de novo classification was helpful in facilitating that, can you elaborate a little bit more?

DR. GUTMAN: I can. Because CYP-450, which was 2D6, which was the first one that we brought to market, had no predicate, it would have been by default a Class III product. We probably could have processed this actually in a similar amount of time.

It might have warranted a panel meeting although you can sometimes waive panel meetings, but what we decided was that based on what was known, based on the fact that drugs had taken the first step and already labeled the drug with this test, that we felt comfortable sorting through the literature.

We did sort through the literature in a fairly methodologic way because we didn't just allow any variations, any alleles to pop up. We actually tried to base ones that seemed to be more plausible and have a stronger database.

We also understood the weakness because the recommendations on Strattera are clearly there, but they are not highly specific, and, of course, there are no recommendations in our labeling, they are quite general.

We understood that physicians would have hopefully enough insight and to understanding that they needed to use the test prudently, but that we weren't going to start for each of the 80 or 100 or 120 drugs that might be impacted by this diagnostic, we weren't planning to create a label that would provide specific guidelines to physicians on the use in each case.

We crafted--the de novo, it is like going to church or synagogue, it is a very formalized process with all kinds of different steps, and what you do is you find the product nonsubstantial equivalent, which would normally be bad news, but then the company petitions and then you turn around and you find it approvable, and you create a special control. That special control is a matter of public record. It is actually on our web page.

We made it very broad, so we called it metabolic enzymes, and then UGT-1A1, its similarity to CYP-450 was that it was a metabolic enzyme. So, we have allowed ourselves some wiggle room.

In all honesty, if you came along with the Steve Gutman enzyme or the Steve Gutman allele, and we didn't

know how the hell to connect it to anything biologically meaningful, we probably would say this is great, market it as an investigational use device.

DR. RELLING: So, it is not really tests-based or type of analytically test based, it is more functional?

DR. GUTMAN: Intended use, yeah, function, and we do that quite frequently.

If you look at the way we handle--this is with more specificity--but the way we handle markers for myocardial infarction, the demonstration of myocardial ischemia, in fact, we moved from creatinine phosphokinase to the MB subunit, to troponin, to ischemic albumin, all off the same intended use, very, very different analytes, more similar perhaps in methodology, but quite different in everything, in their half-lives, in what interferes with them or helps them, but the same intended use.

The rationale for that is that the issues of safety and effectiveness are common.

DR. RELLING: Thank you very much.

Open Public Hearing

DR. RELLING: At this point, we have our open public hearing, and I would like to request if there is anyone in the audience who would like to speak at this time.

[No response.]

Committee Discussion of Questions

DR. RELLING: If not, we can move forward with the committee discussion of the questions that have been brought up this morning.

I can try to recap a few of these issues and I will ask for you all to help me. I think what is being asked of us is to provide comments to the FDA on whether we think biomarkers can be better integrated into not only drug development, but also into issues of drug labeling and drug usage.

I know that there is also some specific questions that are in our agenda, which we will go over also. I think we learned that the integration of biomarkers is integrated into the critical path initiative, so that they have some importance for the agency over the near term and long term.

It was pointed out that we have to do a better job of defining what we mean by biomarkers and deciding how to both, quote "qualify" them and validate them, and I think we have just heard some recent clarification on there are sort of two issues.

One is deciding to what extent the biomarkers have clinical or functional utility versus being absolutely confident about the analytical performance, so that they

can be utilized practically in multiple geographic areas with timeliness throughout the country.

It is also then pointed out that pharmacogenomics, which of course we spent a lot of time talking about yesterday, might really be the ultimate test case or one of the early test cases for how well biomarkers are going to be integrated into drug development and drug labeling.

I think that is one of again the main questions that we are being asked to address is how do we get these biomarkers and/or genomic tests, so that they can be usable in drug development and in drug labeling.

Dr. Lesko pointed out several examples, and I think we heard from some of the material presented by Dr. Mayers, as well, that drug labeling could benefit from incorporation of more information that relates drug concentration, measures of drug sensitivity, and relationships between drug dose or drug concentration and toxicity into the labels.

We also heard some good points from Dr. Mayers about what to do if TDM suggests the use of drug doses that are outside the range of previously studied doses for that drug.

Then, we heard in a clear presentation from Dr. Gutman about in development of biomarkers, particularly

developing biomarkers to be used on a clinical basis, there is really no excuse for confusion on how to evaluate the analytical performance of such new biomarkers or tests, that there is multiple roadmaps that are available to providers of such tests to determine how to start to qualify or validate them, and that there is already experience within the agency for approving such tests for pharmacogenomics.

Are there other major points that should be summarized at this point?

How do you want to proceed, do you want to go over the questions that we have listed?

DR. LESKO: Yes, I think the questions are a good starting point, and many of the comments that were made during the course of the morning will lead into those questions, and we might both discuss the questions and some of the comments that were made by the different speakers.

DR. RELLING: Why don't we start by reading the Questions to the Subcommittee.

The preamble is that clinical biomarkers are used during drug development for identification of individuals at risk, e.g., the QT interval, prediction of treatment outcomes, e.g., viral load, selection of appropriate doses for individual patients, e.g., TPMT genotype, and

monitoring therapeutic effects of treatments, e.g., plasma drug concentrations.

With regard to the latter, the first question is: When is it desirable or necessary to include plasma drug concentration information in package inserts, and where in the label would this information be most useful to providers and patients?

Comments?

DR. LESKO: Maybe I can just give a little context to the question based on some of the things that were discussed this morning. One of the points that was raised, I think it was by Dr. Mayers, was basically why have we not progressed beyond anti-epileptic drugs, anti-infective drugs, and bronchodilators in terms of therapeutic drug monitoring.

As many on this committee realize, these were, and are, commonly used tests for therapeutic drug monitoring of patients to individualized therapy, but when you think about it, over the course of time, there has not been very many new tests that have come online.

I can't think of very many. There are kits out there to measure viral load, there are some kits out there to measure blood levels of antiviral drugs that are commercially available.

I think the point is, though, that it isn't because we have stopped dealing with drugs that are toxic or drugs that are ineffective. I think we have as high an adverse event rate as we have had 30 years ago.

I think we have drugs that have been reported to have efficacy rates that range down to 10 percent and sometimes not much more than 30 percent.

So, it is hard to say there is not a problem in optimizing therapy and therapeutics. One of the questions is, well, when would blood levels be necessary, so we could have a discussion about the framework as to the attributes or criteria of drugs for which therapeutic drug monitoring might be useful, and thus, information in the package insert would be interesting to have.

I think one of the reasons why we haven't seen the progression of therapeutic drug monitoring is that the information, particularly on new drugs, on relationships between dose and response, and between exposure and response, is not public.

This information is certainly obtained during the course of drug development. We have dose-response relationships quite frequently for drugs that are submitted to us. We actually do modeling of exposure-response relationships in terms of plasma levels almost routinely in all of the NDAs in order to determine from pharmacokinetic

studies when dosing adjustments are necessary in subpopulations.

We try to set therapeutic boundaries that if blood levels or area under the curve go above a certain range, you need a dose adjustment, and when they go below, you don't.

So, I believe the relationships between exposure and response in NDAs is potentially useful as information for a package insert in terms of the TDM. The question is when would that be relevant, that is, what is the criteria, would it be a disease like AIDS where we have a life-threatening disease as opposed to what Dr. Woodcock said nonsteroidals were, I just take a dose and titrate to effect.

Would it be where you have large inter-individual variability, and you don't have a sense in the case of a therapeutic failure or toxicity, whether I should increase the dose or decrease the dose, or if I have multiple regimens?

So, I think it would be helpful to begin thinking about a framework which would strive us towards thinking about when this kind of information would be useful.

The second part, which I tried to emphasize and maybe get a committee response to in my presentation, is that when we have dose-response relationships in a new drug

application, I pointed out that after surveying labels, we have very little examples of dose-response information in labels.

So, one question would be why not put drug information in labels that include dose-response relationships, graphs, or things of that sort, is there any down side to doing that, could that be, as somebody pointed out yesterday, more information that could be harmful, or could it be useful information to steer a physician, determining whether to increase a dose, decrease a dose, or change a drug, knowing what that relationship is for that particular drug and how it might relate to safety.

So, I am just trying to kind of frame that first question to maybe get some comments from the committee with regard to a direction we might think in terms of developing some framework to these questions.

DR. RELLING: Dr. Barrett.

DR. BARRETT: Larry, one of the things I thought that came out of the morning's discussion was this maybe gauntlet as far as what would be criteria that you would like to have in place before you would move things forward for not just the TDM, but even the genomic testing or putting that as part of the label.

One of things I think the committee, it is incumbent upon us is to be able to help frame that, because

just because we can do something doesn't mean we should, and that is specific for TDM, as well.

However, as we have discussed in the past, I liken this to like emissions testing. I mean I may not like to bring my car in every year to have this done, but there is some merit to having it done, but as cars improve over time, I don't necessarily feel I should have to do that, so maybe that's a personal thing, but on the topic of the blood levels with some of these indications, I think there is clear guidance in this notion of biological plausibility.

While I think it is an important first step, we have more than that. We have data in some of these cases where there is a clear connection on exposure and outcome, so drug levels being an important surrogate in the area of generics, of course, where you have a precedent for doing this and putting some of that information in the label, but the extent to which it becomes informative to make decisions on dosing, that is a clear indication and where we need to actually reflect it in the label.

Of course, there is a necessary complement to having the tools to perform the test, and, in fact, acceptable criteria by which you can take that information and provide practical guidance.

But I think one of the other issues would be is that it is in the right spot in the label and the dose adjustment piece, and not so much embedded into along the way of drug development we have discovered these things.

You asked at the very beginning here where should this be. For certain, it has to be defined on a per-drug disease basis, but I would just put forward that to the extent at which this is going to guide clinical outcomes and manage patient care, it should be in the right spot in a label in the dose adjustment place.

DR. RELLING: Dr. Singpurwalla.

DR. SINGPURWALLA: I am going to try and respond to Larry's statements. I think you know what my position is. I think information overload is bad, and I think you, in your question, used the word "patient," and when you say patient, I think of myself. I don't know what blood plasma means, and if you put it on the label, it wouldn't mean anything to me.

It is just like these diet labels on foods. We just don't read them. We just eat what we want to eat. The same thing is going to happen here. You put too much information, you are going to take the risk of lawyers suing doctors because they read these labels and say you did not do this when it was clearly stated.

So, I think you are entering into a territory which is rather risky and where do you draw the line? Today, it's blood plasma, tomorrow, it's, you know, genes. The day after tomorrow, it is something else. So, I think there has got to be some kind of a constraint on how far you want to go with these labels.

Already, the labels are hard to read. They are in fine print. If that is what you have in mind, then, I am opposed to it, but if you have in mind information for the providers and the physicians, then, that there may be a technologically efficient way to communicate all this to them, and I think you should.

So, that's my position.

DR. LESKO: I think just to clarify, I think the latter part of what you said is, in fact, what I was trying to convey in my comments. I am not suggesting, for example, that we put this information on the label of a prescription.

Nevertheless, there are certain disease areas where patients may want to access labels more than other areas. I think HIV would be one of those areas, oncology, probably another, where information about their care would be of greater interest.

But I think when I say providers and patients, I am thinking of the provider using information to the

benefit of the patient as opposed to a patient getting information that they are going to use to self-direct their own therapy.

The other side of that coin, though, occurs when you have a biomarker that is compelling as a predictor of risk, and it is not used, and somebody could claim that it was not in a label and be sued again.

So, I mean it works two ways with regard to biomarkers and lawsuits. I am not thinking in that context, I am thinking of science and clinical medicine here today.

DR. RELLING: Dr. Kearns.

DR. KEARNS: Thank you. I actually do read those labels now. I used to not when I go into the cafeteria, but now I read them to make sure I don't take something that has too many carbs or I give up the really good things that way.

You know, Larry, I think you kind of said a moment ago what, in my mind, focuses this, and that is, situations where we do have dose-exposure response information. To me, that is critical for inclusion somewhere in the labeling.

On a personal level, we do a clinical consult service, and I remain always frustrated when I am asked to see a kid with a transplant and asked to comment on

mycophenolate concentrations, or one of those drugs, because I am forever looking at where is the target. You know, if you give me a target to shoot after, being somewhat reasonable in kinetics, I could usually find the target, but there is no consistent information.

Since these data are collected during drug development increasingly here in the modern era, we have to figure out a way to make that information, appropriately qualified and validated, transparent to people who have the ability to use it, and use it wisely. I think that is critical.

Using another example of the importance of this, and I am going to go back to my arena, which is the pediatric arena, as we work on drugs for children, we are always fighting this battle of whether we can extrapolate a disease state between a child and an adult, and depending upon the data, the week, the time of the day, and the time of the year, people will argue that you can extrapolate, and then argue that you can't extrapolate.

Usually, when those arguments are lost, it is in the concept that is rooted in everybody's mind, and that is, children are different, ergo, the disease must be different, so we have to default to learning about drugs in a very traditional way.

I would posit, on the other hand, that if we had good dose-exposure response data, we would see that there are more similarities than differences. We could really improve, not just therapeutic use of drugs, but the study of drugs maybe in subpopulations as opposed to reinventing the wheel again.

A last comment and a bit of a story. Not long ago I was contacted by a friend who used to be a major league baseball player, and suffered a severe head injury, and he was being seen at a very prestigious medical center somewhere in the Midwest in the State of Missouri, that wasn't Kansas City.

He talked to me about symptoms he was having and had been put on the old medicine phenytoin, and I suggested to him that he go to his neurologist, who was a very esteemed individual, and suggest that the neurologist measure a blood level of the drug.

The neurologist originally told him that it really wasn't necessary, blood levels were meaningless, you know, but if you want one, I will get one. Well, of course, it was high, the toxicity was there. It's something that those of us who have been around for 30, 40 years know about very clearly.

But I asked myself the question, was this person's reaction to his patient something driven out of

assurance or something driven out of ignorance, and if it was out of ignorance, what kind of information should be available, if we can't convince a physician, to at least convince the patient and help the patient.

So, yes, I think the information needs to go in there, needs to go in the right place, it needs to be the right information, and done using tests that are the right tests.

DR. RELLING: Dr. Sadee.

DR. SADEE: I think it's a very dangerous road to go down if we say, well, let's adjust to the minimum qualifying level of competence of, whatever, patient and physician and scientist might be there.

I think a lot of information is good. I would like to emphatically state that currently drug therapy has still lots wrong with it. There are efficacies that range from 10 to 50 percent even for our major drugs. There are adverse side effects or leading cause of death in the United States and of morbidity.

So, there is an urgency for us to actually use the data that is available to minimize the effects and maximize the efficacy.

Actually, I was called by a lawyer to--we are not concerned about this, but on the other hand, it illuminates a little bit that point of view--and he asked me, well, why

on earth aren't we using these informations that are out there, why are we not doing this, isn't that practically malpractice because it's already in the literature, and it was a lay person reading the literature saying these are the connections between genotype, plasma level, you name it, and so on, why is it not used.

So, I think we have an obligation to do this quickly and efficiently, and more information is good.

DR. RELLING: I wanted to just make the point, Larry. You asked what kind of criteria should we use to help decide when it is desirable to include such information in package inserts.

I guess it is obvious, but I would say it would be for drugs that have narrow therapeutic indices, for diseases that are life-threatening and for which drugs can be life-saving, and drugs for which other, easily available, monitorable criteria, laboratory criteria, or otherwise are not readily available.

I think that is sort of the principle of why we do TDM, and if you build it, they will come. So, if the information relating drug concentration to toxicity or effect is more available, eventually, in our capitalist society, somebody tends to fill that void.

Dr. Capparelli.

DR. CAPPARELLI: I would just like to add a couple other situations which relate to that. One is where you are using combination therapy and there may be a lot of drug interactions.

One of the things that comes forward in these studies is you do have a more homogeneous population in your Phase III studies, you are trying to get information, but when that is applied outside, you really need to extrapolate that to the situation and as a caregiver to the individual patient.

The other situation that you will run into is where you--and this happens in oncology and HIV--where there isn't this sort of second chance, second opportunity to look at things, and you really only have one good chance to get things right.

If you are off initially, and you could have prevented or could have at least optimized therapy, you actually can create a much greater benefit.

DR. RELLING: Dr. Gage.

DR. GAGE: To summarize what I hear both of you saying is what matters is not just that we have the information, but that it is likely or at least may be clinically relevant.

Is that a fair summary, because that would sort of motivate what goes in?

DR. RELLING: Dr. Capparelli.

DR. CAPPARELLI: I think it does relate to the clinical relevance, and it has to do with understanding the dose-response, and there are a lot of interrelated terms here. If you have got a huge therapeutic index, you know, having that information, even if you have drug interactions, is not going to be all that helpful.

If you have got a wide phenotype from the disease state, while in a certain subpopulation you may be okay, if you anticipate there may be a need of pushing the exposure issues and having dose-response or exposure-response on toxicity and efficacy, may be of help, and having some integrated measures of efficacy would be of help.

But I also think the degree to which you have some certainty in the quality of the data tells you where you need to put it. I like the idea that we talked about graphically during dosing, I really like the graphs that had the IC_{50} information and maybe not being as specific, but allowing the providers to actually utilize that information on individual cases rather than being very directed in terms of how to address each situation.

DR. RELLING: Dr. McLeod.

DR. McLEOD: I think there are situations where a good monitoring should be done, that is not currently done, but I also think that the reason why good monitoring

stopped being of high value was just that, it stopped being of high value.

There were too few examples where blood levels truly correlated with the effect, and therefore, that, in cooperation with other intangibles, such as the decrease in clinical pharmacology training and those types of things led to the demise of the discipline, or at least the reduction in the use of it.

Dr. Mayers showed a nice example where the pharmacodynamics were rapidly changing, and even if you optimized the pharmacokinetics, or at least if you tried to catch up and optimizing the pharmacokinetics, realizing your point, Ed, if you start at the right place, you are probably okay, but if you don't start and you try to catch up, using blood levels, you are probably not that much wiser, because the dynamics are now out of control.

So, I think that there are situations where the pharmacokinetics will be of value. There are clear concentration-response relationships, and that information should be in the public domain anyway, but I think that that will not lead to therapeutic drug monitoring. Just knowing that there is a relationship there doesn't mean it should be done.

The last thing is that there are a lot of market forces against this. I mean who wants to develop a drug

where it might have to be monitored? I mean the marketing folks certainly aren't looking for that, so at some point in time, there need to be some dramatic examples that drag this forward, so that the marketers get it in their head that this is not evil, and then it will start to happen.

But I think at the least, the information needs to be available, if not in the insert, on the website or such.

DR. RELLING: Dr. Barrett.

DR. BARRETT: I just wanted to point out that in many places in the existing structure of labels, there are clear guidance with respect to dosing modifications, specifically in the area of renal and hepatic impairment, and the underlying evidence by which that data makes the label is, in fact, typically, a pharmacokinetic study in which blood levels are measured.

So, one of the issues with the TDM approach is the fact that patients can't otherwise classify themselves unless, in fact, a level is measured. So, there is no potentially any other characteristics which would allow them to identify themselves as being a candidate for dosing modification or their prescribing physician.

Then, there is the potential for time and variance. You know, when we look at somebody as being renally compromised or obese or pregnant, or whatever other

state we are going to use to guide dosing modification, there is that snapshot in time by which they can associate themselves with their characteristics and guidance can be provided, but we don't think of ourselves as having a blood level in excess of some threshold unless the measurement is, in fact, taken.

So, to some extent, this is really driven by the mechanism by which the clinical investigation proceeds, so if you knew a priori that drug levels, drug monitoring was relevant, then, that discrete experiment could be summarized to the standpoint of putting it in this, but the issue of assays and analytical competence around this is really a detail.

I know that that is not a trivial, but the fact remains that if, in fact, this is shown to be clinically relevant, there will be a market for having this done in an expedient fashion, and not to the standpoint of not being available to the patient. At least that is my opinion.

DR. RELLING: I think if it's okay, we will move on to the second question, which I think is highly related to the third.

The second question is: What evidence should be available to support the use of plasma drug concentration information in package inserts?

The follow-up is: What is the best approach to obtain that evidence during the course of clinical drug development preapproval, or as part of a recommended postmarketing study?

Open for comment.

DR. LESKO: If I can make a few comments to sort of steer the discussion in terms of what we would like to hear some opinions on. I think the question at its heart has to do with quality of data.

When Steve was talking about diagnostic tests that are approved by CDRH, he focused quite a bit on analytical and clinical validation. In fact, analytical and clinical validation is something we would look at in a new drug application as a way of accepting or not accepting a pharmacokinetic study or a dose-response study or a PK/PD study, so there is a given that the analytical validation and clinical validation are appropriate.

That leaves the clinical utility of the information, and during the course of the morning, there were some questions related to predictability or how to assess clinical utility, and predictability, one of my questions is does the committee feel that is the right question.

In other words, we put information in the label, whether it's liver function tests, QT prolongation, or even

plasma levels that are not absolute terms. They don't say a patient is going to be toxic or not. They don't say whether a patient is going to be efficacious or not.

What the numbers give you is a probability of an outcome, and I think the context for this question in terms of what level of clinical utility is appropriate, could that be answered in the context of, well, am I able to characterize a patient as being more likely to have a given adverse event or more likely to have a benefit if they are in this particular plasma level range.

I think the other context for comment is related to what Dr. Mayers showed, and that was a lot of scatter between a trough level and a population of patients. That could well be from the study design itself.

It may be that trough level is not the right metric to look at for exposure, but I think when it comes to plasma levels and therapeutic drug monitoring, what is important is where is my patient, not where is the population of patients, and thus, does a patient with a lower level or a higher level have a greater or lesser chance to be successful in terms of benefit or risk.

So, that is kind of the context here, it is population versus individual, it is information on probability versus predictive values.

DR. RELLING: Dr. McLeod.

DR. McLEOD: With your specific question, at least I think, or are you merging 2 and 3, Mary, are you just on Question 2?

DR. RELLING: I was merging them.

DR. McLEOD: I think that most of the studies that are designed currently are not designed to answer the question that you are wanting, that you are answering this question, or answering the questions we might want for defining a concentration for therapeutic drug monitoring.

I am not sure that is it is reasonable to expect that, because that is not the goal, or society has not set that as a goal or an expectation for drugs. Rightly or wrongly, that is not the current expectation.

So, I think that in the postmarketing, having sufficient data, the best data possible available, and working for postmarketing analysis in that context is the way to go. I think we already mentioned previously, some of the obstacles to that, and that there is very little incentive for a company to do that.

The stakeholders who do have incentive, such as CMS or Blue Cross/Blue Shield or other payers, are not organized enough to do that sort of analysis, and the FDA, it is really not the responsibility of the FDA to develop that.

So, while I personally would like to have that information and think that having targets for blood levels would be ideal, if I were a patient, I don't think that the expectation should be to have to develop those studies, and in the absence of that, I think we would be overinterpreting any data in terms of trying to define levels, and that scatter you saw was, like you said, a lot of it because of study design.

DR. RELLING: Larry, I guess I had interpreted, there are sort of two different things. One is are you asking whether drug concentration data should be more incorporated into the label even if they are not used for therapeutic drug monitoring, or is your primary goal to have it in there for real therapeutic drug monitoring.

I guess my opinion is there is a utility in putting in some information that relates drug concentrations to effect even if there is very little practical hope that it will ever be done on a clinical, CLIA-approved, FDA-approved TDM type basis.

DR. LESKO: You may see the line of thought in this series of questions, and it was basically maybe getting toward what Howard just said, that this is not for everything, there is going to be subsets of the general population of drugs where there will be attributes of the drug or the disease, where this may, in fact, be something

you would want to get during drug development prospectively. This can be discussed and managed during the IND process.

Studies could be designed when it is deemed, for that set of drugs that meet these characteristics, that kind of study could be designed, or perhaps the information could be gleaned from what is being done already in that particular drug development program.

I think the second part of it is the label and what goes in there. I think we need to separate out the issue of having information--let's say we had an effective development program that clearly had reasonable relationships between a dose or a PK and benefit and risk.

That is separate from therapeutic drug monitoring, and I think keeping those separate is useful in terms of thinking about these questions. I showed the irbisartan label where you had kind of a nice shape of a dose-response.

Now, I will be the first to admit that is a drug with a lot therapeutic index or broad therapeutic index, and not a narrow, but nevertheless, the information was there.

What if that were a drug with a narrow therapeutic index? It would seem useful, then, to have

that kind of relationship especially if you combined it with an adverse event biomarker of some sort.

So, I think there is information which would be things like dose-response, and then there is information like TDM that would be more action oriented specifically. I think the latter is more problematic than the former.

DR. RELLING: Dr. McLeod.

DR. McLEOD: The data, the concentration effect data, how often is that missing from the package insert? I know there is not data to the level of doing TDM, but the two examples you gave seemed like fairly standard bits of information that are in a lot of the labels that I have looked at, and I am not an avid label reader, because I try to limit the amount of excitement in my life, but I think there is a lot of blood level data in there, in a label now.

DR. LESKO: I think what you see in labels--and this may depend on a time frame we talk about--you know, if we are talking about the last five years, the last year, whatever, I think there is a lot of data that is not in labels for a variety of reasons, descriptive data that characterizes exposure-response.

I showed, for example, that little survey about pharmacometrics, and there were 200 and some NDAs, but only

a subset of those had data that we were able to analyze to help make a regulatory assessment or impact labeling.

It wasn't that there was data in the other 100 and some that wasn't useful. It was that there was no data that we could do.

DR. RELLING: Dr. Sadee.

DR. SADEE: So, the purpose of the information in the label on drug levels and relationship to response could be several fold. One could be to provide sort of an average, the other one could be to provide the range that one might expect. Then, a third one might be to provide a guide for therapeutic drug level monitoring.

I think the way this has emerged over the past quite a number of years is that initially, we thought, well, drug level monitoring is good.

Then, we began to understand the range and had more information on it, and then one could actually design optimized treatments a priori that would minimize the number of adverse events or maximize the efficacy.

So, by doing that, finding a dose somewhere in between, we minimized the need for drug level monitoring, because now we were educated on the extremes and the median, and so on, what is in the population.

So, I think different types of information should be there depending on whether one has made a decision that truly drug level monitoring is still useful.

DR. RELLING: How about if we move on?

The fourth question is: To what extent is it necessary to have actually studied the efficacy and safety at doses recommended in the package insert, based upon existing relationships between plasma drug concentrations and clinical outcome?

DR. LESKO: This gets back to some comments that were made this morning related to recommendations of a dose that was not within the range of doses studied in the clinical trials.

The question that really is in front of the committee is to what extent can we move away from the dose given that there is a wide variability between dose and concentration, and asked the question differently, what is the concentration related to the range of concentrations that were studied in the clinical trial, and could that, in turn, lead to the recommendation of a dose that was not, in fact, studied.

So, if a concentration requires a lower or higher dose than was actually studied, but that concentration is within the range of concentrations observed in, say, pivotal clinical trials, would that be appropriate.

Now, I realize without other information, it will be hard to answer that, but let's say we had a dose-response relationship, let's say we had a mechanistic understanding of dose-plasma level relationships, under those circumstances, is there any time, if ever, that that would be reasonable to think about.

DR. RELLING: Dr. Gage.

DR. GAGE: Because clinical trials are typically of younger, healthier patients than is a target audience, I think there are times when we have to extrapolate particularly in the elderly and use and recommend ranges that are different than those that were tested in trials.

DR. RELLING: Dr. Barrett.

DR. BARRETT: What we have seen over the last couple of days is the rollout of a number of tool sets that give some comfort in terms of being able to predict and extrapolate information particularly when there is good relationships established between the therapeutic window defined by indices of safety and efficacy and concentration.

So, I think there is some comfort now in being able to move to that kind of an approach. Having said that, I think there is still somewhat of a risk when you get into more sensitive populations and you are worried

about the extremes, how does the extreme perform with a given dose regimen.

There is little concern if you are talking about interpolating or suggesting a dose that hasn't been studied, that is lower than perhaps what you studied, but the extrapolation I think becomes more of a potential safety concern, but again I would kind of guide that by your assessment of the sensitivity of the population and the variability of the dose-exposure relationships, and assuming that those are not show stoppers, I wouldn't have any problem in, in fact, proceeding.

DR. RELLING: Dr. McLeod.

DR. McLEOD: We just trying to think through if there are any examples where we have gotten to the point through the evolution of medical practice where we stop thinking about dose.

In the area of oncology, one of the platinum agents, carboplatin, is currently dosed based on a measure of renal function that depending which institution you are on, it depends how you measure your renal function, but we really don't think about dose for that drug anymore, we think about the area under the curve, of 4.5 or 6.0 depending how you are treating, what you are treating.

That was a situation where we started off with milligram per meter squared with dosing and through trial

and error, got to the point where predicted blood level was the goal.

So, it certainly is possible and feasible, but it took a long time and it took a high level of perceived confidence before that was indeed there. So, I think your concept is good. I think that the timing is probably wrong in that the drug development phase is not sufficient to get that level of confidence.

Also, there is the psychological barriers that also have to take time to overcome, so I think it's a noble goal, I am just not sure that it can be achieved within the drug development phase.

DR. RELLING: The last question is: What analytical validation data are appropriate for recommending therapeutic drug monitoring information in the package insert?

Is this referring to the package insert for the drug or for the device, Larry?

DR. LESKO: This was referring to the drug product, and one of the contexts for this question relates to approved drugs in which we--well, first of all, think of label revisions, we revise labels based on published literature, for example, a drug interaction between, as I mentioned yesterday, proton pump inhibitors and warfarin,

there is no prospective study, but it comes from the clinical literature and, on its surface, looks valid.

So, the context for this is that frequently, sometimes in antiviral areas, there is compelling evidence from the literature that a blood level would be beneficial in certain circumstances in monitoring the patient.

In that case, if we were to update the label, what kind of analytical validation would be necessary beyond that which is in the literature, and with Steve here, because we always work together on these CDER/CDRH issues, but would there be a need, I mean what would be a CDRH perspective on including blood level information assuming the evidence was out there in a drug label, would that come under the purview, Steve, of CDRH in terms of approved devices?

Let's say I wanted to come in with, I wanted read a label, an antiviral, with the blood levels of the drug, because they correlate nicely with viral load or preventing toxicity, and the method is HPLC or RA or something like that, and we relabel that, would that be a circumstance where you would be interested in approving that drug from a device standpoint?

DR. GUTMAN: You would certainly prefer that it be approved if CDER were going to sanction it as a test, to have a couple of choices. One would be that it be FDA

approved. I suppose there is not any legal reason you couldn't default to a home brew assay.

You have done that before, and I think in the case where you have a product where CDRH hasn't analytically credentialed it, it would be a wise move for CDER in some way to analytically credential it, and the more information you could provide about the requirements of the assay, you know, if you believe in transparency and honesty, then, you need to do something to determine which test should be test, what quality test should be used, what quality exists.

So, I think, as awful as we are, it might be easier to go through us, but if you didn't go through us, I think you then carry a football that you owe the healthcare providers and the patients to communicate as much as you can about your expectations and qualifications of the test.

DR. LESKO: We answered our own questions here while we chatted, but certainly we are all welcome to hear more comments.

DR. RELLING: Dr. Singpurwalla.

DR. SINGPURWALLA: Question No. 5, you used analytical validation data. Do you mean data or do you mean information, because data, to me, is just numbers; information to me, is knowledge. Numbers, of course, contribute to knowledge.

So, I presume you include the word information, because you have a lot of models besides data.

But to answer this question specifically, if this question pertains to biomarkers, then, I would think that cross-correlation data between the biomarker and the effect that you are really interested in should be provided.

The negative is what if the cross-correlation is very small, like, say, 0.5, which to me is small, then, it's ineffective putting it up because the individual might think that it's as good as tossing a coin and taking an action.

Am I clear?

DR. RELLING: Dr. Jusko.

DR. JUSKO: I think a lot of the discussion that we have just heard needs to be further addressing the question for the future about measuring biomarkers including this kind of information in the label.

In measuring drugs, I have a lot of confidence that when one uses methods like LC mass spec, and such, that a great deal of specificity and accuracy is available in those assays, and then simpler assays can be compared to these.

But when I see a lot of measurements made for biomarkers, measuring proteins, measuring things that may be altered in the body, and there is uncertainties about

what these assays may mean, when some of these techniques are said to be validated, I am not so sure that in a variety of kind of patient situations, again, modification of proteins, that the analytical specificity is there and that the trust may be there and what these measurements may mean.

A lot of what we talked about this morning seems to me deja vu back in the '70s, in the era of evolution of therapeutic drug monitoring, a lot of the rationale for therapeutic drug monitoring was established and the reasons were well stated this morning.

It seems like many of these same issues come up with the biomarkers. When we have good biomarkers, we don't need drug assays in part, and measurements like INH, glycosylated hemoglobin, and such, provide good indication of what may be happening in drug therapy, but the time course of the dynamics and the complexities of the dynamics makes their interpretation a bit more awkward unless they are predicated or underpinned by mechanistic models that interrelate what is happening with the drug, physiology of the patient, the disease progression.

I think as we saw implied with some of the simulations and situations yesterday, the combination of approaching therapeutic questions on the basis of face value empiricism plus modeling that incorporates basic

mechanism is probably the preferred way to go whenever we can.

DR. RELLING: Dr. Barrett.

DR. BARRETT: One comment I had pertaining to analytical method quality, I guess, is that during the course of drug development, it is well appreciated that these analytical methods evolve and they presumably get better.

They also potentially get more intensive in terms of their requirements on analytical quality, as well as the human element, to actually conduct the assay, so they are not necessarily always portable to a broad-scale clinical setting.

So, I think you need to accommodate really both settings. In those instances in which you don't need a commercial kit or an assay to accommodate a huge volume, that there has to be some leeway by which, you know, you have a few patients that may require an assay, and then the other extreme in which you would benefit from having several perhaps commercial kits with different reagents or whatever, so there has to be some issue in terms of the robustness of those assays across the commercial kits and then some portability relative to the methods that were actually used to define the assay.

The only thing I think I would mention at this stage is that there needs to be some flexibility in terms of the clinical environment that these assays find themselves in, as well as perhaps the robustness metrics that are used to, in fact, compare assays across manufacturers potentially.

I think the only other thing I wanted to mention on the topic was the extent to which they are used for dosing guidance really has to be well defined in terms of where you expect to see that level occur.

So, when you have an analytical criteria that defines the use of the assay, that you pull the sample at the appropriate time, so that there can't be any misuse of the actual technique, sample, et cetera, relative to the operating characteristics of that assay.

DR. RELLING: I see no further comments from the group. What is next?

Summary of Recommendations

DR. LESKO: I am not going to try to summarize all the recommendations, just do a quick recap and say that I feel the day and a half that we have had here has been extremely valuable for us at FDA in terms of the questions we brought before the committee.

Reflecting on our discussions, we arrived at some very specific questions yesterday morning. That is very

helpful to us in our next steps with regard to genomics, warfarin, and relabeling, and that was the intent of bringing that to the committee.

The topic we discussed in the afternoon yesterday and this morning, both are works-in-progress in terms of model-based drug development and developing informative labels, and I think the information that you provided us in the case of these two topics not only have advanced our understanding of the current situation scientifically and clinically, but also helps us frame what our next steps are going to be in these areas.

I anticipate, like a lot of topics we bring before this committee, being more general that drug-specific, that the discussion we had yesterday afternoon and today will form the groundwork or basis for some subsequent meeting topics that we will be planning in 2006, so we will look forward to continuing to discuss these as they resolve to specific recommendations for the FDA.

So, with that, I want to thank the committee for their advice and participation in this meeting. I want to thank our colleagues and guest speakers for their presentations especially the team that worked together with me to get ready for this Advisory Committee.

Finally, I want to thank Mimi and Jane, and the rest of the support staff of the Advisory Committee. It

was a lot of work behind the scenes that you don't want to know about, but you can tell from the shuffling of chairs yesterday, it wasn't all that simple getting this program to run smoothly. So, I want to express my appreciation to them for that, and look forward to another meeting in 2006.

Thank you.

DR. RELLING: We are adjourned.

[Whereupon, at 12:25 p.m., the proceedings were adjourned.]

- - -