agents. New drugs in development--they should be encouraged if they show activity against resistant virus so, on the preclinical level, just to reinforce what you said, characterization of resistance at baseline and, importantly, activity against clinical isolates that already have resistance would be important to demonstrate.

DR. FLETCHER: Just a comment on the clinical drug-development side, at least as you being Phase I studies, there it would seem to me that the opportunities for use of phenotypic assays could become very important in terms of trying to integrate, as Dr. Mellors and others said, that information on how susceptible the virus is with what concentrations are actually going to be achievable in the body.

To state the obvious, you can have the most sensitive virus but, if you can't reach effective concentrations, it is not really going to matter. I think a converse could be plausible as well--that is, you could have an intermediately susceptible virus but the concentrations that you can achieve in the body are so several-fold higher that you could have activity in that setting.

So I think there, in the Phase I environment, that integration between phenotype and pharmacology could be a real role for that. Just lastly, what will need to be sorted out is what is really important quantitatively about

the phenotype. Is it the IC50? Is it IC95? Is it multiples of the IC95 that one needs to use there?

DR. HAMILTON: A nuts-and-bolts question that I think has some implications for preclinical-trials design and that has to do with several points that I think Doug Richman raised having to do with what seemed pretty sensitive levels of detection of mutants for both the genotypic and phenotypic assay.

If I understood him correctly, it was between 95 and 99 percent for phenotypic versus genotypic assays. Does that mean, actually, that above a thousand copies, let's call it, that the level of detection is the same at all levels or would one need, actually, to stratify based upon viral load when using this methodology?

DR. RICHMAN: If I could clarify that a bit. The assays that are based on amplification of RNA from plasma, and this applies both to the genotype and the phenotype assays--what they are doing is analyzing the population of the amplified material so that, at most concentrations of plasma RNA, it really doesn't make a difference if you are starting with 10,000 or 100,000; you are looking at sort of the consensus or majority population.

The issues; it depends upon the specificity of the assay. I showed examples of both phenotype and genotype, that maybe some situations for a particular drug or a

particular codon, 20 percent of resistant is enough to show the difference. Sometimes it maybe 50 or 60. That depends on the drug and the codon.

That is really independent of the virus load. Where an exception occurs is at the very lowest concentrations. There you can have what we call a founder effect. If you happen to get a positive amplification from someone with a hundred copies, the source for that amplification could, just by chance, be a single RNA molecule.

So, at the very lowest concentrations of RNA, you can get founder effects and stochastic-effect data. But, in general, from a practical point of view, what we do is we do RNA assays to define whether someone has succeeded or failed the treatment and then someone who has failed is someone who then gets a drug-resistance assay for either evaluating a drug or for managing a patient.

For most of those patients, we don't start doing these assays when they have got 200 copies. But that is a potential risk.

DR. HAMILTON: Doesn't that, then, make quantification difficult over time when you are evaluating longitudinally the emergence of resistance if, in fact, what you are looking at is an amplicon and not a total load of resistant virus?

DR. RICHMAN: I think what I tried to say is once you are up in the thousands or more, what you are looking at is reflective of this complex mixed population. Now, that doesn't deal with the fact that that complex mixed population is changing over time, too. That is true as well.

So you may get a different result when someone is at 2,000. If you keep treating them and they go up to 200,000, the result may change because the population has shifted as well.

DR. STANLEY: I want to go back and build a little bit on what Dr. Gulick said and break this again, as you suggested, Dr. Hammer, into the preclinical and the clinical. Clearly, resistance testing and drug interactions are not going to predict in the individual patient but I think we need to decide, preclinically, what do we want to know, in vitro, about these drugs before we then go to Phase I trials and can go better pharmacokinetic studies and look at interactions.

As Dr. Gulick said, if, in your in vitro studies, your early preclinical development, you are able to pinpoint some advantages that this agent may have over certain resistant strains or whatever, then, as you move into Phase I, you can start to do the pharmacokinetics to understand the drug interactions and also any resistance

advantage.

So I think I still would advocate for some drug studies on panels and resistance and combinations in the preclinical in vitro setting.

DR. HAMMER: We need to move on to the second question. Are there any other comments? Just for the agency's sake, maybe just to help summarize as we go question-by-question, first I think, as was stated earlier, two of the last slides that Dr. Richman showed really answer the first question pretty much as was summarized by the committee.

When asked about the relative strengths and limitations of genotype versus phenotype, I think we saw them--for the purposes of drug development, both are important. The genotypic profile of each drug that comes through, in vitro needs to be classified.

We need to know whether it is multiple step resistance or whether single-step high-level resistance and what the phenotypic correlations in vitro are. We already talked about looking at a new drug that has activity against a defined panel of isolates in vitro.

On the clinical side, one is interested, of course, in what the emergence of resistance will be to that drug, recognizing, again, that this will be in combination with other agents and so there is a challenge there because

of what may be effect of other agents on the combination. But at least one can try to determine the pattern and also, as Dr. Mathews said, I think, on some of the clinical-development side, and this subsumes Dr. Gulick's comments too, an agent that looks good against resistant isolates should be potentially tested prospectively in populations where you define, up front, those populations.

I think, to echo Dr. Mathews, when one knows that isolates are resistant to a particular agent and that will threaten a combination doing prospective studies that eliminate patients who would be potentially harmed by that combination being suboptimal, should be more and more a part of the developmental process.

So I think, basically, on the clinical side, it is resistance emergence which, I think, is part of a lot of profiles but prospective resistance at baseline, et cetera, is important. Then, also, one can think about management and strategic studies along the way but I think that may be best discussed tomorrow.

Anything else you want on this question?

DR. HAMILTON:

DR. MURRAY: I wonder if anybody could comment on, perhaps, using one type of genotypic assay over another early in clinical development when you are screening virologic failures for the presence of mutations. Would you

want to use, like, a hybridization technique where you might detect lower numbers, or a lower percentage of mutant mixture, to have a more sensitive screen for resistance popping up with that drug on what you have predicted preclinically? Or could that data be, actually, misleading?

DR. HAMMER: I think I will tackle that because I think when you are dealing with a new drug in a new combination, as I think Dr. Richman outlined, you don't know what the outcome is going to be and, therefore, I think you need a dideoxy method that sequences the region of interest because, if you are not smart enough to know in advance what to look for, you will miss it with an assay that interrogates certain regions.

My personal feeling is that, in looking for emergency of resistance in vivo, once a drug hits the clinical-research side early and in different combinations, that you really need to look at the regions of interest--RT and protease is what we are interested in today--and make assessments as to the relevance of those mutations.

Then one can go back to the laboratory with what one sees and prove, or disprove, what those mutations may do to an isolate in vitro. However, I think a point in favor of the assays that interrogate certain regions, once you have defined, for example, perhaps a key mutation, those are very helpful assays to look at baseline to see whether

somebody has or has not that particular mutation and then randomize on the basis of that.

Therefore, a more sensitive assay even helps you with your randomization scheme. So I think it depends, really, on the nature of the study design as to which assay is better. But I think when you don't know what the answer is going to be, you should not be doing an assay that goes just after codons of known, previously determined, resistance.

Anybody want to disagree? Dr. Mayers?

DR. MAYERS: I wouldn't disagree. I would just extend that even further to the point that the isolates that have been selected in the preclinical development programs for resistance have, not infrequently, not been the isolates that you actually see in the clinic so that I think you have to look very carefully in'your Phase I program to see what the failure mutations are in your patients before you could even consider doing that.

The problem is going to be, as you start combining these drugs into regimens, you can shift to new patterns that you didn't anticipate before. So I think you are sort of stuck, at least in the understanding of what is happening in your trials, in doing something that screens the whole genomic interest.

DR. HAMMER: We will let Dr. Pomerantz have the

last word on this first question.

DR. POMERANTZ: Thank you. I just wanted to point out something that I had talked about before and that is, if you are going to look at phenotypic resistance and correlate it with genotypic assay as you are asking which assay, I do agree that you should sequence it up front. But I think it is going to be even more complicated than that and that there are groups of viruses that have their resistance that may vary based on context dependency, which is a term that John Condra taught me a while back.

It is not only within a gene but there are a variety of other genes in HIV that affect protease and, in particular, affect reverse transcriptase. So, if you are not going to sequence that gene, you may actually need to look at other genes besides that. Nucleocapsid is important for template switching and reverse transcription.

There are differences. There are differences. There are integrase effects. The integrase gene affects reverse transcriptase. So it is going to get somewhat even more complicated and I think you are going to need the baseline values of sequencing at least that gene of interest.

Protease not only has mutations in protease but, as you know, there are PR cleavage sites that there are mutations in that may become more and more important. so I

would definitely agree with sequencing them up front as a genotypic assay because you are going to need that as you move to the more complex molecular interactions.

I see Dr. Mayers shaking his head, so I would be very strong on that.

DR. RAMMER: Thank you.

We will move on to the second question which we have already sort of broached a bit with Dr. Murray's question; what studies are needed to further define performance characteristics of available assays in order for them to be useful in drug development.

I think no one disagrees with the fact that they are useful in drug development. It should be there. I think this is the issue of where we need to move from the current state of the art as was described this morning.

DR. KUMAR: I have a practical issue to raise regarding this. Most assays currently available need to have a viral load greater than 1,000. So, for clinical trials, right now, the problem that they are facing is **as** soon as a patient fails, we withdraw the patient from the clinical trial. That is usually when we define failure; in most clinical trials, it is greater then 400 copies.

And then we move them on to another regimen. And so we really lose a window of what to do with these patients because we have withdrawn them and we have given them

another antiretroviral regimen. So that is something that I would like to raise here to see what will we do with this information that will clearly be lost.

DR. HAMMER: I would just say that this is, a), a moving target and already the manufacturers are moving down and have some data to suggest that they can get down to 500 copies. It is not as reproducible as a thousand or more copies, but, like everything else in assay development in HIV, this is moving and will be different in six months and twelve months than it is now.

MR. HARRINGTON: That raises something that I wanted to mention before which is there is a difference between fitness and virulence. That is particularly clear in studies like Steven Deeks where those patients are only partially suppressed but there is some virus around their CD4, there is a disconnect between their CD4 and their virus.

So if the clinical trials are just taking everyone off who fails, then we are not really looking at what is happening out in the real world.

One other thing that is, I think, very interesting clinically is what is the relative fitness and virulence of the resistant phenotype because, if the 3TC-resistant virus is replicating very well in people but it doesn't replicate very well in the lab assay that Doug showed, you have to

actually have a lot more 3TC-resistant virus than protease-resistant virus.

So that would something else that might be very useful clinically to know which is the disconnect or what is the relative fitness versus the relative virulence of the resistant phenotype of the given virus.

DR. KUMAR: I want to come back to this issue because this is an issue that we continue to face while we do clinical trials. Invariably, written within these protocols, as it is written right now, for most protocols, is if the viral load goes about 400, we withdraw patients from these protocols.

So I really want to raise that issue here to see what to do with this disconnect that we are seeing.

DR. HAMMER: I think that is a slightly separate topic about the strategic issues of clinical-trial endpoints but I would make a general comment that, although the failure endpoint, whether one chooses 400 copies or whatever, is quite a valid endpoint depending upon the study design. Many studies, although they call that an endpoint, don't necessarily mandate a switch at that point. It is up to a physician and patient to determine whether one moves to a second regimen or not.

It varies study to study, but the mandatory switch is not necessarily a part of every regimen. I think the

point that you are raising for this morning's discussion is the window of knowing, irrespective of what you do strategically at the next step, is defining what the genotype or phenotype of the virus is at the time of the virologic endpoint.

So I think, if I may interpret, the point you are raising is what we need, and what you would like to see, is greater sensitivity at the levels of 400 copies and even above 50 copies to be able, actually, to define the profile of the viruses that patients are failing with and correlating that with the endpoint, whether that is first failure or multiple failure.

We know that one can be surprised, if not intrigued, by what the results are when one looks in those failure populations as to whether it is adherence or drug resistance or whatever.

So I think, if I may infer, what you are saying is we need greater sensitivity at lower copy numbers to eliminate the disconnect between the virologic endpoints and the threshold for amplication that we currently see today.

DR. STANLEY: When I try to answer this question when it comes to the genotypic studies, aside from needing more sensitivity, as you just discussed, I guess if I understand Dr. Richman's data, I have a little bit of difficulty because, when he showed the comparison of the

laboratories and how much they got things right, he made the statement that the success or failure was independent of the kit that was used, that it really depending on the quality of the lab.

So I think that is going to be critical and who is doing the preclinical testing or who is doing the drug development, how do you assess the quality of the lab as opposed to the individual kit, if I am understanding the data.

DR. HAMMER: You can hold that for question No. 3, what quality-control data are needed. You are leading us right into the next question.

DR. MAYERS: An issue that I think is going to come up with the committee is going to be how to interpret results done by different companies for different drugs. There has got to be some effort made to get a comparison of what the expected range of normal would be and what, for certain type isolates, the expected range for an assay done by one company versus an assay done by a second company versus, perhaps, a PBMC-based assay would show you because it isn't clear that the ten-fold difference for one company would be a ten-fold difference for another company or that that would relate in any way to a PBMC assay.

So I think there is going to have to be some effort to define what those ranges are for the different

assays. It may be drug-specific.

DR. HAMMER: One thing to ask in that regard is, certainly when you only have a single isolate from an individual in a trial, you have nothing to relate it to except the control that tht assay uses in a phenotypic assay. But one thing to also think about is having sequential assays from the same patient which provides not only more real baseline information than what is available in the clinical trial but what the sequential changes are related to an individual's isolate.

More and more, that information could be, and perhaps should be, part of packages that one sees where it is definable and would help answer that question that you have got.

DR. YOGEV: Maybe one would like to see resistant in vitro really mean in vivo. We define resistance and in a certain situation, we try to use the genotype getting nice phenotype--those blue marks and everything--and it didn't work in real life.

So I was wondering how to define what we call resistance; for example, a phenotype of eight-fold. What does that mean in **a** certain situation? Just to reflect of what was allowed by Dr. Masur before, it is into microbiology. We know for a fact that what is called resistant for a certain number of bacteria in the blood is

completely irrelevant to what is happening in the CNS let alone the penetration of the drug.

But you need much more of the drug when you have much more of the virus in certain areas. I don't think that we understand what resistance really means defined in vitro. So I would like to see, in the performance, in vivo correlates to that specific definition so that we can accept it.

I think the best example is AZT. If you have one mutation, it is two-fold. If you have two mutations, it is four-fold. This is the gray area that one should define a little better by the range of pharmacological data in a certain area versus the correlate of the in vitro. So, basically, I am asking for an in vivo definition.

DR. HAMMER: That raises the question, again, and I think, also, as Dr. Mellors mentioned, that the drug exposure is the key issue and we need to be able to interpret, at some point, IC50s and IC90s in relation to achievable drug concentrations in vivo and whether it is Cmin, Cmax, area under the curve, whatever, that is really the exposure characteristic that is important.

That may be drug-class specific and even drug specific. But one would hope, I would think, as this process moves forward, that a database will develop of what achievable drug concentrations are and what is important.

We already have some data about trough protease inhibitor levels being predicted, for example, and trying to relate certain achievable drug concentrations to a particular fold change in drug.

Whether this also means therapeutic drug-level monitoring is going to be important for certain drugs and should be available is the topic of another symposium.

DR. YOGEV: I agree with your statement. I just am raising also the issue of the amount of the virus. We all are taught that what you have in the blood is 100 or 1,000 or what is in the lymph node or in other issues. I would like to see a correlation of resistant to the amount of the virus.

So I know, for example, that drug is called resistant when I tested it at 10,000 as well as at 2 million. So, I think, on the top of the pharmacokinetic, I am talking about the virus, itself; the test done in a defined number of viruses is appropriate to the in vivo just on the virus amount.

DR. HAMMER: I think Dr. Richman answered this before and he may want to comment again that at levels of, certainly, in the several thousand and above range, that the amplification techniques from plasma are relatively representative of the population, at least the lymph-node population that is most immediately reflected in the plasma

and that, as you get to the very low copy numbers, then you are at risk of the population.

DR. MAYERS: I would extend the previous comments even further. The cut points, and that is the real challenge to the industry and the phenotyping companies right now, is what is the upper limit of normal for your drug and where is the break point for your drug where you lack of clinical responsiveness.

They have got to define those break points by drug, not with an arbitrary relation to a one-virus standard. So I think that is going to be the real challenge is what is normal, what is the break point that defines lack of activity in the clinic. And then you have your three ranges. We don't have those for any drug at this point.

DR. HAMMER: Are you suggesting that we need an NCCLS-like apparatus for HIV drug-resistance testing?

DR. MAYERS: I think we need a thought process like NCCLS in this. There has to be a relationship between your break points and the clinic. Right now, those arbitrary break points are not that standard and we don't have the data. The only way we are going to get the data is if the companies acquire it during the drug-development process.

DR. HAMMER: Other comments on question 2?

DR. MATHEWS Again, dealing with clinical trials

and experienced patients, it seems like there needs to be some standardization of the background therapies in relationship to when the baseline resistance assays are done; in other words, a cross-sectional estimation of what the resistance is conditioned on, whether the patient has been unstable or antiretroviral therapy for eight weeks, four weeks, has had a washout period, all those kinds of things.

So if you are trying to correlate response based on a given mutational pattern, it may be completely uninterpretable if there is a great deal of heterogeneity across trials trying to evaluate the same drug or the same regimen and those factors.

DR. GULICK: Just to add on to that, it is a problematic situation when you are in drug development at an early stage and you want to characterize in vivo what resistance mutations are developing with a new drug. However, directly opposed to that is our wish that we can come up with a new regimen that will actually suppress viral load.

How do you come to terms with wanting to do both those things at the same time? Just to echo something Dr. Mathews said before, we don't want to design unethical trials in order to gain resistance information. It is a bit problematic.

DR. HAMMER: It is very problematic. I think one thing it raises, which, I think, is a little bit beyond the scope of today is what the role of certain animal models might be in allowing you to investigate regimens and drugs that would be unethical to use in humans and where animal models may or may not help you define, at least generate hypotheses and give you some look at what may happen in an in vivo situation--Skid-U mice, for example, have been used, primate models, et cetera.

DR. MELLORS: Just to follow up on what Chris and Trip said, we are not always going to be successful in our salvage regimens. Hopefully, we will be more successful in the future but it points to the critical need to not let failures escape or follow up and our detailed evaluation.

There has to be a shift from presenting the proportion that are successful to the proportion that failed and why and what resistance occurred as a consequence. Right now, we are in an era where dropouts are treated as failure and they are kind of swept under the rug.

But the agency and the research community has to really focus on characterizing well the failures. That is the way we will build a database. It won't be a comprehensive database. If we do our job well, there will be few failures. But we really have to hang on to them and learn as much as we can from them. No counseling jokes.

DR. HAMMER: Any other questions on question 2? We need to move on to question 3. Again, just to summarize for the agency, what studies are needed to further define performance characteristics of available assays in order for them to be useful in drug development. I think the processes that are going on now are defining them. I think we have raised a couple of issues here.

Everyone is interested in greater sensitivity of these assays. Personally, I think that is coming. I think the issue of picking up the mixtures is also important and some assays are better at this than others. We will never be able to be perfect because we know that even less than 1 percent of a viral subpopulation can emerge under the right conditions and will likely not get quite there.

But I think the two areas for further improving the performance characteristics of available assays are lower limits of sensitivity on amplification and the ability to detect mixtures.

Anyone want to add to that?

Question 3; it is already sort of introduced a bit by Dr. Stanley. What quality-control data are needed to support use of these assays in clinical trials and drug development. What additional studies should be conducted with respect to reproducibility and quality control

Among others, I would like to hear from Dr.

Jackson on this question. Sorry to put you on the spot, Brooks.

DR. JACKSON: I think Doug presented a number of the performance characteristics that need to be quality-controlled such as sensitivity, specificity, the precision. I think, in judging from the ENVA panel, there is still a real need for this type of data and standardization of this sort of quality-control data that needs to be developed both for the kit performance characteristics as well as for the performance of the laboratories.

I guess the latter is really not part of FDA's thinking on this at this point in terms of this discussion.

DR. HAMMER: Although, I think we should separate two things. One is kit approval issues through CBER and drug application packages that come to this division and this committee. That is different. That is not an approval process. That is sort of what should the data package look like.

Of course, the assays underneath that have to have a degree of validation but I think we can make suggestions about what a data package should look like in the drug-development process which I think is what we are being asked to do.

DR. JACKSON: Just like any diagnostic assay, I think the same sort of things that are required when data

are submitted in terms of sensitivity, precision, the specificity, all those sorts of things in different populations, different sets of populations that are looked at in terms of how these assays perform in those respects are going to be needed.

I can't think of anything other than those specific topics that Doug brought up but clearly there is a need for that.

DR. HAMMER: Other comments? This could be a short discussion.

DR. HAMILTON: I have a question of the agency which may not be answerable but I wonder if, in the preclinical mode and prelicensing mode, the agency provides any advice to applicants, to sponsor, relative to the testing procedures they are proposing to use in a subsequent study.

DR. IACONO-CONNORS: What we normally tell sponsors when they propose a series of studies, we basically ask them to identify what assays they would like to use to support their clinical studies. If those assays are unlicensed or unapproved, as would be the case with these types of studies, then we ask them to provide us with additional performance information that would allow us to be able to interpret the data that those assays would generate.

So we don't give them any direct advice. We wait

for them to come in with their proposals and then see what kind of performance information packages that the drug companies would need to get from the kit companies in order for us to understand the data.

So it becomes a bit of a problem.

DR. POMERANTZ: I have a question for Doug. You presented a nice picture of how the ViroLogic assay goes for phenotypic analysis. My question is, as you are thinking about generalizing things, you pointed out that they are in-house. That is basically an HIV-based pseudotyped vector.

If we are going to start using phenotypic assays as we are talking about now more globally, that is going to be an interesting way of straining the ability to use what is still a complicated assay. How do you see it developing or how would you see the company developing it so that it could come to use in all these different studies and non-studies for patients that we are talking about now.

DR. RICHMAN: I can't speak for the two companies but I have talked to them. I think they think they can or hope to expand to accommodate the phenotype needs of investigators and practitioners. Now, if a third or fourth company comes along with performance characteristics that are similar, that is fine.

I think I can say they have been talking about trying to document that they come up with similar results on

the same specimens as Doug Mayers tried to raise. I am not concerned with the numbers that they come up with in terms of the fold change because I think I showed slides from both companies showing how tight their values from wild-type isolates were around the control value.

The issue of what value is important is of practical concern. You will see data this afternoon giving some indications that for many of the drugs we are using now, four is a better number than ten. But, for each drug, we do need to know some data.

There are, in fact, I think good data regarding abacavir, retrospective data with monotherapy, indicating that somewhere about four- or six-fold is probably enough to be enough to make it not work.

That is probably true with a number of the drugs, but the pharmacologic issues with some of the protease inhibitors raised a question about more moderate levels. I think there are really no alternatives because the data from the PBMC-based assays--blinded comparisons have been done with these other assays and they certainly can outperform speed, cost, precision and everything else.

There are a lot of internal problems that we have all observed with the data that we have generated with our own assay. So, for these two targets, I don't see an alternative from a practical point of view. I am not sure

if that answers your question.

DR. POMERANTZ: It does. The main point is that you think that one company in the United States can, in one site, handle this?

DR. RICHMAN: You will have to ask--

DR. POMERANTZ: Because we are talking about doing a lot of trials and moving phenotypic analysis--

DR. RICHMAN: There are representatives from both Virco and ViroLogic.

DR. POMERANTZ: Would you think about having subsites? How are you going to do this? This is not a trivial--it may not be the PBMC assay but it is still not a trivial process.

DR. RICHMAN These are not assays that can go out of house.

DR. POMERANTZ: Exactly.

DR. HAMMER: I think it is fair to say that we are talking about market forces here. If the demand is there, I think that it will be met. I don't know that we need a specific--since we were asked, really, not to be product specific, I would rather not ask representatives of ViroLogic or Virco to answer this directly, officially, but during a break, I think this question can be asked.

Given entrepreneurial skill and market forces, one would think the need could be met.

DR. JACKSON: Just to follow up on some of my earlier comments that Dr. Dayton brought up I thought would be relevant that things we have found with viral-load testing, for example, in terms of these assays being used in clinical trials over time, I think it is important to see quality-control data regarding lot-to-lot variation, that sort of thing.

We have seen, in past trials, various deteriorations sometimes with the stability of reagents. These sorts of things will probably need to be shown as well as validation panels that are done showing the performance of the laboratory personnel using the assays in these trials. These are fairly standard procedures but I think they are important.

DR. HAMMER: I think there are a couple of things. One is we should probably think about phenotype and genotype a little bit separately here. There are two commercial groups in the phenotype field at the moment. That may change in the future, but I think the data we have seen about the descriptions of their assay performance and at least the approvals that have come from some regulatory groups gives a fair amount of credence, I think, to the data that are generated and help in any submission in a packet.

But those are the kinds of things we need to see. I think, as Brooks outlined and as was outlined by the

earlier--Dr. Dayton's talk, that is sort of the standard things. On a genotyping side, there are the classic kit issues, reproducibility and lot-to-lot variations.

I think one other thing just to mention with regard to PCR and sequencing, as Betty Korper has, more than once, made us aware of the issues even in very good laboratories of contamination and that when large datasets or even moderate datasets are sent in of sequencing in a single laboratory, what procedures that laboratory goes through to be sure they have not had a problem with contamination--that is, sort of checking their sequences across their isolates and against the isolates that are commonly used in the laboratory is a reasonable question to ask of any laboratory, whether that be in something that is submitted for publication or in something that is submitted for drug approvals because that problem is one even the best labs have occasionally.

Are there any other comments on the quality-control issues and issues of reproducibility?

DR. WONG: I guess I just want to make one comment and that is that I am a little concerned that phenotypic testing seems to be going to route of adopting a heterologous expression technology exclusively. I understand why this is necessary and that it is easier and cheaper and more reproducible for now, but I guess I would

.ope that we not abandon completely the idea that the virus, n the context of the virus itself, has to be assessed as 'ell.

Just taking out a gene, expressing the protein in tester strain and then testing that strain is not, .ecessarily, the whole story. Although practically it may e the way to do it today, I wouldn't want us to get locked nto that sort of system.

DR. MATHEWS: I had a question about the assays or reproducibility. It seemed that most of the ones that oug mentioned were assay reproducibility and not really ooking at sources of biological variability besides drug ressure. When the viral-load assays were looked at, when D4 assays were developed, biological variability was one of he big issues that had to be grappled with.

To what extent has this been looked at, ariability in patients not on drug at different time points nd under different conditions?

DR. HAMMER: You mean multiple sample from the ame patient, either off drug or on the stable regimen and he reproducibility--

DR. MATHEWS: If they are on drug, that is a roblematic issue because, obviously, if they are not uppressed, they may evolve. Is there a sort of steady-tate background biological variability that needs to be

taken into account?

DR. RICHMAN: In the interest of time, I took out slides. Richard Haubrich, whom you may know--there is a strategic study looking at the utility of phenotypic testing versus not done by the California Collaborative Treatment Group. He presented at ICAAC the results of two consecutive tests taken on each of those people who entered at that time.

I could show you the slides, if you like, or he could--but, basically, the reproducibility on two consecutive assays was quite impressive.

DR. HAMMER: What was the interval between the sampling?

DR. RICHMAN: It was 30 days.

DR. HAMMER: Thank you.

I think, on that note, we will break for lunch. We will return in an hour at 1:05.

[Whereupon, at 12:05 p.m., the proceedings were recessed to be resumed at 1:05 p.m.1

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

[1:10 p.m.]

DR. HAMMER: We will begin this afternoon with Session 2 entitled the Evaluate of Relationships Between Genotype, Phenotype and Treatment Outcome. I would like to introduce Dr. Girish Aras who will introduce the afternoon session.

SESSION 2

Evaluation of Relationships Between Genotype, Phenotype and Treatment Outcome

Introduction

DR. ARAS: Good afternoon.

[Slide.]

I am Girish Aras from the Division of Biometrics 3 and the Division of Antiviral Drug Products, CDER, FDA. I am going to introduce Session 2 that is devoted to the evaluation of relationships between genotype, phenotype and treatment outcome.

The session objectives are: one, to discuss approaches for categorizing mutational patterns for assessing their prognostic value on treatment outcome; two to discuss approaches for categorizing susceptibility profiles for assessing their prognostic value on treatment outcome; three, to determine whether available evidence supports the clinical utility of HIV genotyping in drug

development and to determine what additional information is needed; four, to determine whether available evidence supports the clinical utility of phenotypic testing in drug development and what additional information is needed.

Prospective studies will be presented first followed by retrospective studies. The prospective studies are VIRADAPT and GART and will be presented by Drs. Clevenbergh and Baxter, respectively.

The Resistance Collaborative Group developed a Data Analysis Plan abbreviated as DAP to statistical model the link between treatment outcome at the end of the trial to the baseline covariates including genotypic and phenotypic baseline covariates.

Dr. John Mellors will present an introduction to the Resistance Collaborative Group's reanalysis of selected studies using the DAP methodology. That will be followed by Dr. DeGruttola's presentation of the DAP methodology. Dr. John Mellors will then present an overview of retrospective and prospective studies reanalyzed using DAP.

Drs. Para, Ait-Khaled and Miller will present some of the retrospective studies analyzed according to DAP followed by a summary by Dr. Mellors. Finally, I will make a few statistical comments on retrospective analysis.

We look forward to an interesting and productive discussion. I will introduce the first speaker and then

turn the session over to our committee chair, Dr. Scott Hammer. Dr. Phillipe Clevenbergh will now present the VIRADAPT study.

Thank you.

Prospective Studies

VIRADAPT Study

DR. CLEVENBERGH: Good afternoon.

[Slide.]

This is an overview of the VIRADAPT study as it has been published and presented at the San Diego meeting regarding the long-term follow up and the pharmacological data.

[Slide.]

Numerous retrospective studies have linked the presence of resistance mutations with drug failure. Two prospective pilot studies have shown short-term virological benefit of using genotyping technology to manage patients failing therapy.

[Slide.]

In VIRADAPT, heavily pretreated patients failing therapy were randomized into two arms, a control arm and a genotypic arm in which the antiretroviral treatment changes were based on the detection of resistance mutations. Entry criteria were **a** plasma HIV RNA remaining over 10,000 copies/ml despite treatment with at least six months of a

nucleoside analogue and at least three months of a PI. Clinical status, CD4 cell count, HIV RNA and genotype were assessed every three months.

If viral load failed to decline at least 0.5 log after treatment change, therapy could be changed again based on the latest available genotype in the study arm or standard of care in the control arm.

From March, 1997 until March, 1998, 106 patients were enrolled in the study, 43 in the control arm and 65 in the study arm. The study protocol was approved by the institutional ethical committee in our hospital and informed consent was obtained from all patients.

[Slide.]

We report here the 12-month follow up. After the month 6 interim analysis, a statistically significant advantage was formed for the genotypic group and, therefore, we decided to allow genotyping for all patients.

[Slide. 1

Therefore, the first six months was a randomized study in which the study arm received genotypic guided treatment, yellow, and the control arm received standard of care, blue, whereas, during the second six months, both arms received genotypic guided treatment changes in an open label fashion.

[Slide.]

The primary endpoint of the study was the variation of HIV RNA from baseline at months 3 and 6. Analysis was performed by intent-to-treat.

[Slide.]

For the 6 to 12 months, open-label follow-up study, the primary endpoint was the evolution of HIV RNA from baseline at months 9 and 12. All data collected were included in the analysis which was performed by on-treatment analysis.

Of note, since some patients in the control arm had already progressed beyond six months, genotypic-guided treatment change was performed only at month 9 and for those who were already completing the first year, no genotyping could be performed. Genotyping was performed for 30 of the 43 patients in the control arm--that is, 69 percent of the patients.

[Slide.]

Complete sequencing of the major part of the reverse transcriptase gene and the entire protease gene was performed in plasma HIV RNA. During the first six months of the study, a number of sequencing technologies were used including, initially, a home brew ABI-based sequencing followed by the prereleased version of the VGI TrueGene assay and, ultimately, the VGU TrueGene kit.

After the first few months, technology was

standardized and the 6-to-12-month follow-up study was performed using the VGI True Gene HIV system as depicted on this slide.

[Slide. 1

Decisions for genotypic guided therapeutic changes were based on correlations linking specific mutational patterns to decreased activity of specific drugs. This table is rapidly evolving over time as knowledge about resistance pattern increases. Shown here is the actual version of that table used to guide our decisions regarding RT inhibitors and protease inhibitors during the study.

[Slide. 1

As can be seen on the slide, baseline characteristics were similar between the two arms. Mean HIV RNA about 4.8 log.

[Slide. 1

All patients were heavily pretreated with antiretroviral drugs. They were exposed to a mean of 3.9 nucleoside analogues and to a mean of 1.8 protease inhibitors. This exposure was similar in both groups.

[Slide.]

The overall prevalence of primary mutations for the reverse-transcriptase gene was 90 percent. All patients presented at least one or more secondary mutations. We found one strain with the 151 and one strain with 69S

insertion mutation conferring multidrug resistance. Mutations at positions 67, 69 and 215 on the RT gene were statistical more prevalent at baseline in the study arm than in the control arm.

[Slide.]

The overall prevalence of primary mutations in the protease gene was 48 percent. All patients had at least one secondary mutation in the protease gene resulting in an average of 6.2 primary or secondary mutations.

[Slide.]

103, 99 and 92 of 108 patients were evaluable at months 3, 6 and 12, respectively and included in the analysis.

[Slide.]

Depicted are the changes in HIV RNA from baseline for patients in the genotypic arm, in red, and for the control arm, green. Shown are both, the first six months presenting the randomized study and the second six months in which both arms received genotypic-guided treatment.

In the genotyping arm, reduction in HIV RNA **was** maintained throughout the twelve months of study with a mean drop in HIV RNA of 1.15 log. In the control arm, at the completion of the randomized study, viral load had dropped 0.67 log. During the second six months, open-label genotyping phase, there was an additional drop of 2.98 but

since this phase was not controlled, no conclusions can be drawn regarding the cause of causes of this additional reduction.

[Slide.]

Shown is the proportion of patients with an HIV RNA below level of detection of 200 copies in both arms. In the genotypic arm, in red, the percentage of patients with an HIV RNA below level of detection remains stable, around 30 percent throughout the 12-month follow up. In the control arm, in green, the proportion of patients with an HIV RNA below the level of detection rose from 14 percent at month 6 to 30 percent at month 12.

As noted earlier, some patients in the control arm had a delay in receiving genotyping-guided treatment to month 9 which may partly explain the delayed increase in percent of patients undetectable.

[Slide.]

We performed additional analysis to determine predictive factors effecting HIV RNA responses. The presence or absence of primary protease gene mutations at baseline were correlated with reduction in viral load at three and six months. Shown are the changes in HIV RNA from baseline at three and six months according to the treatment arm and the presence or absence of primary protease mutations.

The greatest reduction was seen in patients who did not have primary protease mutations and received genotypic-guided treatment, yellow symbol, with a drop of 1.5 log. The poorest response was seen in those in whom primary protease mutations were present and received standard of care, pink. Intermediate results were seen in groups in whom protease mutations were absent and received standard of care, orange or patients in whom primary protease mutations were present and received genotypic-guided treatment, green.

[Slide.]

In this heavily pretreated patient population, genotypic-guided therapy resulted in a sustained reduction in HIV RNA of greater than one log throughout one year follow up. Performance of genotypic-guided therapy may have contributed to additional viral-load reduction seen in control patients. Multivariate analysis showed that the presence of primary protease mutations and performance of genotypic-guided treatment both independently affect virological response.

[Slide.]

Even with the use of genotypic guided treatment, 70 percent of the patients in the VIRADAPT study did not achieve complete viral suppression. Multiple parameters determine the response to antiretroviral therapy and many

causes other than development of drug resistance should be considered when analyzing drug failure. Among these factors, suboptimal concentrations of antiretroviral drugs may play a major role.

[Slide.]

In contrast to the nucleoside reverse-transcriptase inhibitors, significant correlation has been found between antiviral activity and plasma concentrations of protease inhibitors and low plasma levels of PI have been linked to rebounds in HIV RNA.

[Slide.]

Our objective was to correlate protease inhibitors plasma levels with the changes in HIV RNA and to determine the multiple factors contributing to the efficacy of antiretroviral therapy in treatment-experienced patients.

[Slide.]

Of the 108 patients included in the study, the 87 patients from both arms participating at the Nice center were included in the pharmacological substudy. Several PI plasma trough levels were performed in these patients during the twelve months study periods.

[Slide.]

Plasma PI concentrations were measured by HPLC. Patients were instructed not to take the morning dose until after drug levels were drawn. The analysis was performed on

batched frozen samples and levels were determined for the four PI utilized in the study. Data were analyzed only for patients having at least three samples obtained.

[Slide.]

Eighty-one patients were included in the substudy. Similar to the parent study, the two groups, according to genotype utilizations, were comparable for risk factor, age, sex, previous treatment, CD4 cell count and HIV RNA at baseline.

[Slide.]

Shown are the mean and median plasma levels obtained for each protease inhibitor from patients in the standard of care and the genotypic-guided treatment groups. Concentrations are expressed in micrograms per ml. The median plasma levels were not significantly different between the two randomization arms.

An efficacy threshold level, as will be explained later, was defined, permitting to categorized patients with optimal or suboptimal concentrations. Between 27 percent for nelfinavir and 43 percent for ritonavir of the patients had suboptimal drug concentrations following our criteria. Among these patients with suboptimal concentrations, the majority even had no detectable PI plasma levels.

[Slide.]

Shown is the correlation between PI plasma levels

and HIV RNA for all individual time points where both measurements were available. Linear regression analysis showed a statistically significant correlation between plasma concentrations and HIV RNA for each PI. Higher drug cons correlated with lower HIV RNA levels.

[Slide.]

A wide range performance pharmacokinetic data is available in the literature for the various PIs. We did not find a consensus statement standardizing these values. Presented in this table are values for IC95, Cmax, Cmin and efficacy threshold concentrations we used for the various PI. The threshold value of twice the IC95 was used as a cutoff for optimal PI level.

[Slide.]

Patients were divided into two groups; suboptimal concentrations for the patients having had a PI plasma concentration below the defined threshold on at least two occasions during the study period, optimal concentrations for patients having at no more than one level below two times the IC95.

According to our efficacy threshold, 32 percent of the patients had suboptimal concentrations and 68 percent had optimal concentrations. The distribution was similar in both groups.

[Slide.]

Shown are the changes in HIV RNA from baseline for the groups of patients segregated based on optimal or suboptimal PI concentrations regardless of randomization arm. HIV RNA decreased 1.2 log at month 12 in patients in the optimal concentration group, in pink, versus 0.36 log at month 12 for patients in the suboptimal concentration group, orange.

[Slide.]

Patients were categorized based on randomization arm and drug level: group 1, suboptimal PI concentrations in the control group; group 2, optimal PI concentrations in the control arm; group 3, suboptimal PI concentrations managed with genotypic guided treatment; and group 4, optimal PI concentrations managed with genotypic-guided treatment.

[Slide.]

Shown are the changes in HIV RNA according to PI concentration and genotype usage at six months. The smallest reduction in was obtained in patients with suboptimal PI concentrations managed on the control group. The greatest reduction in HIV RNA was obtained for patients with optimal PI concentration and genotypic-guided treatment. Intermediate results were seen in patients with suboptimal PI concentrations managed with genotype, light blue, and in patients with optimal concentrations in the control arm, orange.

at

[Slide.]

Shown are the proportion of patients with an HIV RNA below detection limit according to PI concentration and genotype usage at six months. The highest proportion is seen in patients with optimal concentrations managed with genotype, yellow. No patients with suboptimal PI concentrations managed with standard of care reached undetectability, orange symbol. Intermediate results were seen in patients with optimal concentrations and standard of care, in red, and suboptimal concentrations managed with genotype, in blue.

[Slide.]

To summarize our data, multivariate analysis of the predictive factors of virological response showed that optimal PI concentrations, genotypic-guided treatment and the absence of primary protease mutations were all independent significant predictors of favorable virological outcome. The use of different cutoffs for optimal drug levels and definition for primary protease mutations may have influenced these results.

[Slide.]

In conclusion, exposure to drugs was inversely correlated with plasma HIV RNA levels for all PI. Genotypic-guided therapy, drug concentrations and the presence of primary protease mutations are all factors which

independently affect the response to therapy in experienced patients. Therefore, assays to determine both drug levels and resistance mutations may be useful in improving therapeutic responses in drug-experienced patients.

[Slide. 1

This is the list of the contributors, particularly Dr. Gerome, Dr. Garrafo, Professor Dellamonico, Professor Schapiro. The VIRADAPT st was sponsored by Roche, Abbott and Visible Genetics.

DR. HAMMER: Thank you.

Are there clarification questions for Dr. Clevenbergh?

DR. KUMAR: Would you clarify for me, in your control group, who chose the salvage regimen? Was it the same group of physicians who chose the salvage regimen?

DR. CLEVENBERGH: The study was conducted on three sites and for patients, as it is written in The Lancet, considered difficult to treat, there was a discussion between the physician in charge of the patients and the team running the study to decide which was the best treatment to give to these patients.

DR. HAMMER: Other clarification questions?

DR. WOOLSON: Even under the best of circumstances, it is difficult to interpret subgroup analyses that are done in clinical trials. I was wondering

if you had done any modeling of those variables to try to see which of those might be more important, in particular with some of the comparisons you were representing a moment ago with the optimal and suboptimal concentration in those groups.

DR. CLEVENBERGH: I think that the most important of the variables analyzed is the pharmacological effect on the genotypic effect.

DR. WOOLSON: Let me try to ask a follow-up question, actually a somewhat different question. You mentioned, at twelve months, there are some 85 percent of the individuals for whom you have evaluable data. The 15 percent of individuals who are not evaluable, can you describe those individuals in terms of their outcomes?

DR. CLEVENBERGH: I know that there were six deaths. Six patients were lost to follow up. Six patients died and four did not complete the study. I have no idea of what happened to these last four patients.

DR. WONG: Can you say something about the numbers of drugs and the classes of drugs that the patients in the two groups received? Were they comparable, except that on one group, they were chosen specifically based on their resistance pattern?

DR. CLEVENBERGH: The number of new drugs given to the patients is similar and, also, the new classes, because,

at the time the study was done, there were not a lot of new classes. There was just nelfinavir arriving on the market and NNRTI appeared at the end of the study. So the distribution of new classes of drugs is similar between the two groups.

DR. HAMILTON: Since the endpoints that you have chosen to analyze were virologic, were statistical tests done to compare the differences, actually, in these lines chat you generated? It appears to me that, in this one table, at least, or this one graph, for the person who actually doesn't take his drug optimally or at least has suboptimal levels, genotypic analysis doesn't add that much, if anything.

Were there, in fact, statistical differences between these lines?

DR. CLEVENBERGH: I didn't get the question; I'm sorry.

DR. HAMILTON: This graph here which shows--

DR. CLEVENBERGH: Ah, yes; the impact of the pharmacological concentrations.

DR. HAMILTON: Right. Were there statistically significant differences between these lines?

DR. CLEVENBERGH: I don't think that we performed a p-value on this.

DR. HAMILTON: A follow-up question, and it is a

very nice study. I am very interested in the outcome here. It would be a shame not to collect clinical events in the course of this even though they might be small in number. Are there clinical events that are occurring during this period of time that are related to HIV?

DR. CLEVENBERGH: Yes; these are depicted in The Lancet. I don't remember them that we defined. I think that four new defining events in each arm.

DR. HAMILTON: Each arm being--which arms, now? Genotype or--

DR. CLEVENBERGH: Genotype and control; yes.

DR. GULICK: It looks like very few patients had taken prior non-nucleosides. Could you comment on the number of patients who used a non-nucleoside in their salvage regimen?

DR. CLEVENBERGH: There were only a few patients who received a non-nucleoside because it arrives only at the end of the study. The number is written exactly in The Lancet, also.

DR. KAPLAN: This is not included in the presentation but I am wondering if there is any phenotypic resistance data done for some of the protease inhibitors used. Based on some discussion we heard this morning, it would be very interesting to see if there was a correlation between phenotype resistance and the virologic responses.

DR. CLEVENBERGH: We have baseline phenotype for about 90 patients of the 108 and we will present this data on a later meeting.

DR. HAMMER: Thank you very much.

I think we need to move on. The next speaker is John Baxter. Dr. Baxter will speak and describe the GART study,

GART Study

DR. BAXTER: Thank you.

[Slide.]

CPCRA 046 was a randomized study of antiretroviral nanagement based on plasma genotypic antiretroviral resistance testing, or GART, in patients failing antiretroviral therapy.

[Slide.]

To enter the study, patients had to have at least a three-fold rise in viral load while taking at least 16 weeks of two nucleosides and a protease inhibitor. Patients also had to have a CD4 count between 50 and 500 and a total nistory of at least twelve months of antiretroviral therapy.

Patients came in for a baseline visit at which time plasma was collected for viral load testing and genotyping. The genotype results were then reviewed by the protocol virologists and the protocol virologists also had the treatment history. They then prepared a GART report

which included the mutations as well as the interpretation and up to four treatment suggestions.

Patients were randomized between July of 1997 and December of 1998. There were 153 patients randomized, 78 to the GART group, 75 to the no-GART group. Prior to randomization, we required clinicians to prespecify what salvage regimen they would prescribe in the absence of the GART report. For the no-GART patients, these regimens were prescribed.

For the GART patients, in contrast, those clinicians prescribed a regimen after reviewing the GART report. Patients then had follow-up visits at 4, 8 and 12 weeks. At 8 weeks, the clinicians were permitted to change the salvage regimen. Patients were encouraged to stay on their salvage regimen at least until 8 weeks unless toxicity occurred. The clinician could change at 8 weeks based on the 4-week viral-load result.

[Slide.]

This is the algorithm that the protocol virologists used for interpreting the genotypic mutations. That you can see here is the major mutations in yellow, and for the YD4 mutation, for example, we defined this as conferring full resistance to 3TC, ddI and ddC.

For the non-nucleoside mutations, they are listed here. However, only about 10 percent of patients entering

AZT and 3TC or d4T and 3TC in combination with either indinavir or nelfinavir. There were a small number of patients entering failing ritonavir or saquinavir and about half of the patients were'failing on their first protease inhibitor.

[Slide.]

This shows the distribution of major mutations. Overall, 73 percent of patients had both a major RT and major protease mutation. 20 percent had a major RT mutation only. 4.6 percent had no major mutations and 2 percent had a major protease mutation only.

The major mutations are shown here on the left side of the slide showing that the most common failure genotype was 82 percent having the 184V mutation. 61 percent had either 215 F or Y. For protease, 14 percent had 30N, 31 percent had 461 or L. 34 percent had 82F or The and 31 percent had L90M.

[Slide.]

This slides shows the impact of the GART report for patients randomized to the genotype arm of the study showing that the GART report influenced the choice of the salvage regimen for 83 percent of the patients. So, in other words, the report altered what the clinician would have done in the absence of the report.

However, for only 54 percent of the patients did

at

the clinician actually prescribe one of the suggested treatment regimens. The reason why some clinicians chose not to follow one of the suggested regimens included concerns about drug toxicity, patient preference and also saving classes of drugs such as non-nukes for subsequent regimens.

[Slide.]

The primary endpoint of the study is shown here which was the change in viral load from baseline to the average of the 4 and 8-week visits showing that the in GART arm, the mean viral-load reduction was 1.19 logs versus 0.61 Logs for the no-GART arm with a highly significant difference. The 95 percent confidence interval ranged from --0.29 to -0.77.

[Slide. 1

This slide shows the viral-load change by study week showing that, at each time point, there was a significant difference favoring the GART arm. Out at 12 weeks, the difference was 0.44 logs with a p-value of Cr.003.

[Slide.]

This slide shows the percent of patients undetectable by study week showing a significant difference between the two arms at the 4 and 8-week visit. By 12 wreeks, the p-value was 0.1. The difference here in the GART

arm, 34 percent were undetectable versus 22 percent in the no-GART arm.

[Slide.]

One of the things we looked at was changes in viral load by the number of active drugs prescribed. Active drugs here were defined as drugs which the study virologists interpreted as either sensitive or possibly resistant. What we found is that for each additional active drug prescribed, when we combined both treatment groups, there was a 0.37-log reduction in viral load for each additional active drug.

So, for example, those patients who received one active drug in their salvage regimen, the viral-load reduction was 0.1 and, for those who received four active drugs, the viral-load reduction was 1.25 logs.

What you can see here is that the GART group received a greater--there were actually a greater number of GART patients who received three or more active drugs, 86 percent, and in the GART arm received three or more active drugs as compared to 45 percent in the no-GART arm. We believe that this explains part of the treatment difference we saw between the two groups.

The other interesting finding is that, for each additional inactive drug, there was a 0.17-log reduction in viral load.

[Slide.]

We also looked at patients in both arms who received the same number of drugs in their salvage regimen. What we found is that patients in the GART arm were more likely to receive three or more active drugs. So, for example, if you compare patients in both arms who received four drugs, 91 percent in the GART group received three or more active drugs as compared to 62 percent in the no-GART arm. This was similar for those that received three or five or more drugs.

[Slide.]

The other part of the study that we looked at was baseline predefined subgroups and the effect of GART versus no-GART. We found that there was a favorable effect of GART across all of the baseline predefined subgroups so, for example, looking at patients who had CD4 counts above and below 200, the effect of GART was significant.

Also looking at patients that were above and below the median viral load, 25,000 copies, there was a favorable effect for the GART arm.

[Slide.]

Also, looking at patients who were failing their first protease inhibitor versus those who had failed two or more proteases, there was a similar favorable effect of GART. Also when looking at the individual proteases the patients were failing on at study entry, there was also a

favorable effect of GART for each protease inhibitor. The error bars are wide here for ritonavir and saquinavir because there were small numbers entering on those drugs.

[Slide.]

When looking at the presence of baseline RT and protease mutations, we looked at those who had both major RT and PI mutations and found a favorable effect for GART and also for those who did not have both a major RT and PI mutation.

Also, in looking at those patients where the clinician proposed using a non-nuke and a salvage regimen, we found there was a favorable effect for GART. So, in this group of patients, essentially all of them in both arms received a non-nucleoside but you can see that there is quite a difference between the response favoring the GART patients.

For those where the site clinician did not propose to use a non-nuke, there was also a favorable benefit from GART. What this suggests is that when you are adding a new class of drugs to a patient failing a protease-containing regimen, genotyping helps you choose drugs to use along with that new class of drug, in this class, the non-nuke.

[Slide.]

We also looked at the impact of baseline mutations on virologic outcome and found something that was rather

interesting. We found that those patients that were entering the study that had the 30N mutation versus those who did not have the 30N mutation in protease had an average viral load change of -0.41 logs. In contrast, the 90M mutation had an adverse impact on virologic response with those patients with the 90M mutation having a 0.31-log higher viral load as compared with those patients without the 90M mutation.

We found that 46, 82 and 84 did not have an impact on virologist response.

[Slide.]

So, in conclusion, GART with expert advice in patients failing therapy was superior to no-GART as measured by short-term viral-load responses.

[Slide.]

The greater viral-load reduction with GART is attributed to the greater number of active drugs prescribed in the GART arm and we found that those patients who received three or more active drugs had a better outcome. Furthermore, the impact of GART was similar for patients failing their first protease inhibitor than for those who had received multiple proteases. Baseline genotype was associated with virologist response to salvage therapy.

[Slide.]

I would like to acknowledge the protocol team

members, especially the protocol co-chairs, Doug Mayers and Tom Merigan as well as the protocol statistician, Jim Neaton and the protocol manager, Debbie Wentworth.

Thank you.

DR. HAMMER: Thank you very much.

Let me start off the questioning by asking about the clinicians who were influenced by the GART report but didn't prescribe one of the specific regimens. Have you looked at is there any consistency in how those clinicians were influenced, because I think that is a very interesting potentially practical derivative of this study.

DR. BAXTER: It was interesting. We found that the report impacted 86 percent of the clinicians. It changed what they would have done in the absence of the report, but, apparently, 54 percent did they actually follow one of our treatment suggestions.

The other thing that we found is that clinicians who were more likely to follow one of our suggested treatment regimens, those patients tended to have a better outcome virologically. Although it was not significant, there was a trend suggesting that those clinicians that followed our treatment suggestions, those patients did better.

DR. HAMMER: The ones who were influenced but, again, didn't specifically use one of the recommended

regimens, did they tend to use more drugs or a different class or go along with the general results of the study that suggest the more active drugs you use, the better one is--I am just trying to sort of think about how they interpreted this.

They may have avoided a drug because of potential toxicity or a drug interaction but then they would substitute what the physician might have thought would be a comparable agent along the lines of the suggestion of the committee or is it just not possible to really take a look at that.

DR. BAXTER: We have looked at that. For four patients, the regimens that were prescribed were very similar except there was a substitution for the non-nuke, for example. For the rest of them, they tended to use less drugs than what we had prescribed but I can't give you the specific details.

DR. HAMMER: Thank you.

Other clarifying questions?

DR. POMERANTZ: A couple of quick questions. When you showed that the level of undetectability became insignificant at twelve weeks, that was a p-value of 0.1, I think it was.

DR. BAXTER: Correct.

DR. POMERANTZ: 400 copies? 50 copies?

DR. BAXTER: We used the bDNA second-generation assay, 500 copies. But even though there was a loss of significance for undetectable, that wasn't the primary endpoint.

DR. POMERANTZ: No; I understand.

DR. BAXTER: But there was still a significant difference overall between the two groups.

DR. POMERANTZ: Did you see any difference in effects on CD4 counts?

DR. BAXTER: CD4 counts; no, there was not a significant difference over twelve weeks in CD4 count. I think, by twelve weeks, it was +25 CD4 cells for the GART group and +18 for the no-GART group.

DR. POMERANTZ: Do you have any longer-term follow up than twelve weeks?

DR. BAXTER: We are actually rolling--we have rolled patients over into a long-term monitoring study so we will be able to get some additional follow up on these patients. But this was really a short-term look. Clinicians could change therapy at eight weeks and we felt that going much further than twelve weeks probably wouldn't yield much.

DR. YOGEV: In the slide which you are showing HIV RNA changes by number of active drugs prescribed, was GART suggesting one drug only in two? Those are the physicians

who did not follow GART and are still called GART?

DR. BAXTER: For that slide, what we did was to combine both treatment groups. So we looked at all patients and we looked to see how many drugs prescribed in a salvage regimen would have been sensitive or possible resistant according to the algorithm that the protocol virologists used.

So those were for drugs that were either considered sensitive or possibly resistant in the salvage regimen. The greater number of active drugs, the better the virologic response for both groups overall.

DR. YOGEV: So the GART would suggest, let's say, three but the physician would use one which is sensitive? It will be in the category of GART?

DR. BAXTER: That is combining both the GART and the no-GART group so that for the GART group, those patients received a greater number of active drugs. So 86 percent received three or more active drugs in the GART group versus 45 percent in the no-GART group.

But if you are asking how many drugs we recommended in the treatment suggestions, it varied per patient. In retrospect, looking back, what the three protocol virologists were actually doing was trying to recommend at least two or three active drugs in that regimen. So that is how we, in retrospect, were making our

treatment suggestions.

DR. YOGEV: It is a little bit puzzling that when two drugs in the GART were used, were less effective percentagewise than--

DR. BAXTER: That is active drugs, So, for example, in that group where we are looking at two active drugs, those patients could have been prescribed four drugs but only two of them were active.

Does that make sense? That is just looking at active drugs. That is not looking at total drugs.

DR. YOGEV: It is just surprising that, through the GART which supposedly by genotype is more sensitive, there was less percentage in that group.

DR. BAXTER: Yes. The GART patients were less likely to receive fewer active drugs. They were more likely to get more active drugs in their salvage regimen.

DR. HAMMER: Not to go into two many specifics, but how did the expert panel deal with multidrug pain resistance?

DR. BAXTER: Those were the most difficult patients. Those were the longest conference calls. We tried to come up with regimens where there were at least two active drugs but there were some patients who really had very few options and, when we specifically looked at the outcome for those patients, they didn't do well regardless

of which treatment arm they were in.

DR. MATHEWS: It looks like about 25 percent of the patients either had only RT mutations or no mutations. Did you look at whether GART was less effective in that subgroup at all?

DR. BAXTER: We looked not specifically in that subgroup. We looked at those who had both RT and protease major mutations versus those that did not and we found that for those that did not have both a major RT and protease mutation that there was a favorable effect for GART, favoring GART.

DR. HAMMER: So is that similar to VIRADAPT? Would genotyping help even when the virus was wild type at baseline? Essentially, is that sort of what you are asking, Chris?

DR. MATHEWS: Yes.

DR. BAXTER: There were so few patients. There were only 4.6 percent of patients who were really wild types so the numbers are too small to really say anything about those individuals.

DR. KAPLAN: Sorry if I missed this but I wonder about the people in the GART arm who did not follow the advice regardless of the number of active drugs. When you showed the slide of mean HIV RNA by study week, where would that group fall into that slide? Do they become, then,

nondifferentiable from the no-GART arm?

DR. BAXTER: I don't know that we specifically looked at that group's virologic response. We basically just looked at the GART treatment group. I think that, from looking at it the other way which is how well clinicians adhered and their response, we found that those clinicians that adhered less well to the treatment suggestions, their patients had a weaker response or a less vigorous response as compared to the clinicians that followed the treatment suggestions. So it did dampen the response when they did not follow treatment suggestions.

What is interesting is that when we looked at which patients the clinicians were more likely to accept one of our treatment suggestions, it **was** typically patients who were more advanced and had lower CD4 counts and higher viral loads.

DR. HAMMER: Thanks again.

I think we should move on. We are now going to move into the Resistance Collaborative Group's analysis. I would to just introduce Dr. John Mellors.

Retrospective Studies

Introduction to Resistance Collaborative Group's Reanalysis of Selected Studies Using the RCG Data

DR. MELLORS: Thank you, Scott.

[Slide. 1

at

-- It is a pleasure to present the work of the Resistance Collaborative Group. What I would like to do is just give you a little history to start out with, to tell a story. What we are going to do in the next hour or so is review the goals of the Clinical Validation Subcommittee, review the subcommittee workshop that occurred in April of this year, work through the development of a standardized data-analysis plan. Victor DeGruttola will present that to you.

I will then review the mutation table used for the DAP and I will describe the reanalyzed studies and then present the composite data from the reanalyzed studies. Then there will individual presentations of representative studies, ACTG 33 by Mike Para, CNAA2007 by Mounir Ait-Khaled and data from the Frankfort cohort by Veronica Miller.

I will then review prospective studies in progress and make some summary points about the analysis of the data presented.

[Slide.]

The Resistance Collaborative Group Clinical Validation Subcommittee members are shown here; Rich D'Aquila, who had a key role in developing the DAP table, Veronica Miller, Louise Pedneault, Amy Patick, Victor, Andrew Phillips, Lynn Dix, Dan Holder and Jeff Murray from FDA.

Our progress was rapid and steady. There was a temporary slowing of progress while the Boston Red Sox were losing to the Yankees. There was some anger and depression that we had to work through but we persevered.

[Slide.]

The goals of the Clinical Validation Subcommittee were simply to compile and evaluate existing data on validation of resistance tests and to review issues relevant to that clinical validation including study design, patient populations studied, definitions of drug sensitivity and resistance as we heard this morning are critical, definitions of the virologic endpoints used in these studies and, importantly, the methods of analysis and how important covariates such as HIV RNA and new drugs were controlled for in the analysis.

[Slide.]

Let's review the workshop that occurred in April of this year. The goal of that workshop was to review existing clinical data on the relationship between genotype or phenotype and virologic response. We identified studies by review of meeting abstracts. There were presentations by lead investigators that flew in from all over the globe.

There were questions and clarifications by the subcommittee similar to what you all have been doing for the last 48 hours. The studies presented and reviewed were

:hirteen retrospective studies and two prospective ntervention-based studies.

[Slide.]

The impressions of the subcommittee were that :here were consistent associations presented between)aseline genotype or phenotype and virologic response but :here were highly variable methods of analysis including lifferent definitions of resistance, different types of nutations analyzed, different cutoffs for phenotypic assays, lifferent virologic endpoints, methods of analysis and :ontrol for key covariates.

It was clear at the end of the subcommittee neeting that there was a need for a standardized data analysis of these studies to make sense out of them collectively. So the action item from the committee was to develop a data-analysis plan for standardized reanalysis of the retrospective studies and the prospective studies presented.

[Slide.]

So we developed a standardized analysis plan and it is a royal "we."

[Slide.]

Because it was really Victor DeGruttola, the chair, Dan Holder, Andrew Phillips and Lynn Dix who are all in the audience.

Now I would like Victor to come up and present the data analysis plan.

DAP Statistical Methodology

DR. DeGRUTTOLA: Thank you. If we could have the first slide which gives the objective of the data-analysis plan.

[Slide. 1

The primary objective of the analysis plan **was** to specify an analysis that could be applied to a wide variety of clinical HIV resistance studies to relate genotypic and phenotypic testing to response to treatment. The goal was not to try and come up with the most subtle or scientifically powerful analysis of each individual study but to try and come up with an analysis plan that would address the fundamental question and, most importantly, be applied consistently and could be applied consistently across studies so that we could put all of the information together.

Both intervention and non-intervention studies were included but the objectives of the intervention studies that we have just seen presented are different from those of the non-intervention studies so they would be analyzed and presented separately.

Of course, the focus was on resistance data in the experienced population. Where studies may have followed

naive patients, they would be analyzed separately.

[Slide. 1

The non-intervention studies; the first request was to characterize the study population in terms of summarizing baseline RNA and CD4, prior ART experience, and summary of duration of therapy on a HART regimen.

[Slide.]

The most important purposes of the DAP was to try and get a consistently applied definition of an endpoint. Our primary endpoint was virologic failure by week 24 which meant an RNA value about threshold. For studies that were of duration 8 to 16 weeks, change from HIV RNA could be used as an endpoint. Once again, those would be handled separately.

In determining virologic failure, we excluded or asked the individual analysts to exclude transient increases in HIV RNA. If a week-24 value was greater than threshold but the next determination was below, despite no change in therapy, we would not call that a failure.

Our week-24 window was defined to be week 16 to week 32 to try to keep things as inclusive as possible but, for studies whose duration was 16 to 24 weeks, we used the last week with sufficient data. The failure threshold should have been taken to be whatever was specified in the original protocol but if it was not specified, we used

400 copies/ml.

[Slide.]

Patients who withdrew early from study medication were handled in two different ways. The goal here was to handle them in different wavs statistically to try and get some sense of how much study results are affected by the withdrawals. So we proposed two different analyses, a DAC, dropout as censored, and the DAF, dropout as failure, analysis.

In the dropout or censored, patients who withdrew early from study medication without evidence of virologic failure were treated as censored and then excluded from these analyses, but there were provisions given to define an early virologic failure for patients who did not make it out to 24 weeks.

For example, if two of the RNA values were above baseline, they would be scored as a failure even in this DAC analysis. If the reduction from baseline was less than half a log between weeks 4 and 8, they were failure or if the nadir was below threshold.

Patients who were on a study regimen for at least 16 weeks but who had no RNA values in the week-24 window were excluded from all analyses. Finally, we get to the dropout as failure analysis where we had patients who withdrew early from treatment were counted as failures.

[Slide.]

There were two measures of genotypic information that was explored. The first one was a genotypic sensitivity score which, combined across all the drugs in a study regimen, to which a patient had genotypic sensitivity. For each drug in the regimen, the genotypic sensitivity was generally 1, if the patient genotypic has no resistance mutations and 0 otherwise.

I say generally because there were a few exceptions that John Mellors will describe in more detail in a minute. So the genotypic sensitivity score is the sum of the genotypic sensitivities over all the drugs in the regimen, basically the number of drugs to which a patient is genotypically sensitive.

In addition to that summary score, genotypic information was used as three separate variables. The numbers of mutations in each class, protease, NRTI and NNRTI, for classes of drugs represented in this study regimen but not limited to drugs in the patient's actual regimen.

[Slide.]

For phenotypic resistance, we first used two different metrics of phenotypic resistance, the first using the minimum cutoff for the assay, sensitive if the value was less than or equal to the minimum cutoff and resistant

otherwise. In addition to that, a ten-fold cutoff was used. For each drug in the regimen, the patient's phenotypic resistance score was defined as 1 if sensitive or 0 if resistant.

[Slide.]

Once again, there were two measures of phenotypic sensitivity. The first was aggregated against across all the drugs in the study regimen. The PSS is defined as the sum of the phenotypic sensitivities over all drugs in the regimen. The second one disaggregated this and looked at phenotypic sensitivity in each one of the drug classes for classes of drugs that were represented in the study.

[Slide.]

Our analyses also included investigation of the effect of confounding variables measured prior to initiation of therapy. Those included baseline HIV RNA, whether or not the patient was on the potent PI or NNRTI. This would be a yes if the patient was naive to protease and there was protease in the regimen or if they were naive to NNRTIs and there was an NNRTI in the regimen.

So this was just a yes/no variable. Finally, the number of new drugs in the regimen. Once again, nothing is totally straightforward here because after much discussion, it was decided that in ddI-naive subjects, ddI plus hydroxyurea would count as a new drug, but if patients were

ddI experienced and added hydroxyurea, that counts as half a new drug. A mini-dose of ritonavir did not count at all.

Once again, the goal here is to try and get something which is reasonable and consistently applied across all studies although some might argue about where the 0.5 is the perfect choice.

[Slide. 1

Method of analysis was logistic regression of the binary response, just failure versus success, by week 24 with a common set of covariates. There were six models for studies with genotypic data. The first one is just looking at the effective baseline HIV-l RNA; second, looking at new drug covariates, whether there is a potent PI or NNRTI and number of new drugs; C, the one, of course, of most interest, the genotypic sensitivity score; D, the number of mutations in each one of the drug classes; E looks at the GSS, genotypic sensitivity score adjusted for the baseline RNA and for the new drug covariates so it combined the variables in models A, B and C.

F looks at the number of mutations in each one of the drug classes and also includes baseline HIV RNA and the new drug covariates. So F includes all the variables from models A, B and D.

[Slide.]

For the phenotypic data, there were ten models

because of the fact that we were interested in both the assay minimum cutoff and the ten-fold cutoff. Models A and B are the same as before. C is now the phenotypic sensitivity score using an assay minimum cutoff. D is the phenotypic sensitivity score using the ten-fold cutoff.

E is the disaggregated phenotypic sensitivity score for each one of the classes of drugs separately using assay minimum cutoff. F is using the ten-fold cutoff.

[Slide.]

G is the phenotypic sensitivity score adjusted for or including baseline RNA and the new drug covariates, the covariates from models A, B and C put together.

H is the same thing using the ten-fold cutoff for the phenotypic sensitivity. I is just the phenotypic sensitivity scores disaggregated to look in each drug class adjusted for RNA new drug covariates and J is the same as I out using the ten-fold cutoff.

[Slide. 1

In order to summarize results, the summary statistic was the odds ratio for a unit change in each covariate and the 95 percent confidence interval for all the covariates in each model. P-values for testing each covariate individually were also developed. P-values were calculated either using likelihood ratio tests from nested nodels or from the wild statistic.

[Slide. 1

The final slide, for the intervention studies that we just heard presented, we requested that, first of all, the studies examine the effect of the treatment, the randomization to receiving genotypic information versus not--obviously, studies would have already done that--then to do all the models which I just went through and, finally, go through **all** these models again with the addition of the treatment covariate.

That **was** the DAP analysis plan which was sent out to everyone who participated in the Resistance Collaborative Group.

DR. HAMMER: Thank you.

Just for the committee sake, I think what we will do is go through all of the presentations and then take questions **as** a group so that we can get through the retrospective studies.

DR. MELLORS: Thank you, Victor.

Overview of Retrospective and Prospective Studies Reanalyzed Using the DAP

DR. MELLORS: What I would like to do next is review the mutation table used for the DAP.

[Slide.]

The **goal** of the mutation table is a standardization of genotype analysis. It is really not

intended for patient management. It was a consensus, and a hard-fought consensus, of the RCG members and it **was** based on clinical data and opinion. I think the collective number of e-mails was probably over 500 and approached 1,000 surrounding the development of the DAP.

Let me stress that the table was developed before there was any reanalysis of study so it was a priori. We a focussed on primary mutations for each drug that would be expected to markedly reduce response to that drug. The table that is before you is not inclusive of all possible mutations that could influence susceptibility.

The table was used to calculate genotypic sensitivity score and the number of mutations present for each drug class.

[Slide.]

You can review this table. It is your materials. It is self-explanatory. I won't read through the individual mutations because that is even more lethal than the presentations of the data analysis plan.

[Slide.]

We had a group for nucleoside--1 hesitate to show this drug because I think that we have heard enough about adefovir for a while, but, suffice it to say, 65, 70E and the second pathway to multinucleoside resistance was scored as a 0 for adefovir but the 184V was scored as 1.5 because

it increased susceptibility.

There are two multinucleoside resistance pathways, one involving the 151M and the second involving the 69 insert complex. Here are the non-nucleoside reverse-transcriptase inhibitor mutations.

[Slide.]

And, finally, the protease inhibitor mutations.

[Slide.]

Let me just review the calculation of GSS. As Victor said, when a mutation was present for a drug received, the score was 0. When a mutation was not present for a drug received, the score was given as 1. When a mutation was not present for a drug received, the score was given as 1. The exceptions, as I mentioned, 184V for adefovir was given a score of 1.5. Because AZT resistance mutations can reduce the response to subsequent nucleoside therapy, when these were present, d4T, ddI, or ddC were given a score of 0.75. The total score is the sum of the individual drug scores.

[Slide.]

For the calculation of PSS, similar; resistance present, score 0; resistance absent, score 1. The total score was the sum of the individual drug scores that the patient received and resistance was defined as either greater than four-fold or greater than ten-fold decrease in

susceptibility or increase in IC50. Separate analyses were done for each of these cutoffs.

[Slide.]

I would like, in the next two slides, to describe the studies that were reanalyzed. First, the criteria that were used to select studies for reanalysis were that the study was completed, that there was adequate size for a multivariate analysis and twelve of the fifteen studies we reviewed qualified ten retrospective, two prospective, intervention-based.

These tables are difficult. These are summary tables of the develop studies including abacavir pooled data, ACTG 333, ACTG 364, 372, CNAA2007. These are studies with treatment defined by the protocol. The investigators are listed and the number of patients with genotype and phenotype is shown. Note that the numbers, overall, are relatively low.

The treatment experience was varied from nucleoside experience only to heavily nucleoside experience and PI experience with 42 percent non-nucleoside experience. The resistance technology was ABI sequencing or, for these studies, phenotyping was done by Virco. Median baseline RNA shows a broad range as does CD4 cell count.

[Slide. 1

Here are the cohort studies in the GS408 study

that you heard about yesterday. The cohort studies were relatively small in size, experience ranging but generally heavily pretreated. Resistance technology is shown and baseline RNA and CD4 showing a broad range as with the protocols that had treatments specified in them.

You can look at these at your leisure.

[Slide.]

Now let's move on to the composite data from the reanalyzed studies. What I am going to present is first the eight retrospective studies that were reanalyzed and then the two prospective studies. The GS408 and Swiss studies will not be included because HIV RNA change from baseline was modeled rather than the failure endpoint, and this is difficult to show as a metric with the other studies.

I will show you the dropout as failure analyses. The models that I will show you, both unadjusted and adjusted, are the odds ratio of failure for HIV RNA, GSS, number of mutations by drug class and PSS. The other analyses and models have been provided in the documents.

[Slide.]

To put these data that I am about to show in perspective, let's look at the results of the Surrogate Marker Working Group, their metaanalysis of RNA relationship with outcome. These were studies of dual nucleoside therapy, familiar old names, NUCA 3001, 2, ACTG 116a, VA 298

and other studies.

This metaanalysis looked at the reduction in risk of progression to AIDS and death per log reduction in RNA. Here we see the point estimates for the reduction in risk across this metaanalysis. Two things are obvious. First, most of the points are less than 1 showing that there is a lower risk of progression with a log reduction in RNA, per a log reduction in RNA, but not all the studies show the point estimates less than 1.

The second thing that is obvious is that there are broad confidence intervals around these estimates. The one thing I would like to point out is here is a composite estimate showing a narrower confidence interval of all the studies combined. We did not do that because we did not have time nor did we have the logistics to do that quickly.

So that is background for the data that I am about to present.

[Slide.]

Let's warm up to the data by looking at a tried-and-true variable, baseline HIV RNA, as a predictor of failure. What we have here is the odds ratio on a log scale, the points estimates and the 95 percent confidence intervals for baseline RNA. This is the odds ratio per 1-lot unit increase in RNA unadjusted. This is the retrospective studies with the dropouts as failures.

What you can see, and the reason we showed this, is that we showed a consistent finding that has been reported in the literature that baseline RNA is highly predictive of treatment failure with an odds ratio above 1 in all the studies examined.

The confidence intervals are broad but, with one exception, they do not intersect 1 unless they are statistically significant. So this tells us that there is internal validity to the data set.

[Slide.]

This shows the same variable, HIV RNA, adjusted for genotypic sensitivity score and the new drug covariates, whether or not they received a potent PI and NNRTI and the number of new drugs.

What we can see is the unadjusted and, beneath it, the adjusted point estimate and confidence interval. We see that there is some shift in some of the studies to a lower power of the variable but, in general, there are not major shifts. The majority of the studies, the confidence band did not intersect 1 and, thus, show a statistically significant association of RNA with failure after adjusting for the GSS and new drug covariates.

[Slide.]

Let's move on to the baseline genotypic sensitivity score. Let me say that this is a rather crude

estimate of the activity of drugs and the odds ratio is shown per one unit increase. This is unadjusted, again retrospective, analysis, dropouts as failures.

What you can see, again, is the point estimates with one exception, ACTG 372, fall less than 1, meaning the higher the GSS score, the lower the risk of failure, the opposite of RNA. So the more sensitive drugs, the lower the risk of failure and the confidence intervals do not intersect 1 in five out of the eight studies.

[Slide.]

When we adjust for baseline RNA and new drug covariates, we find that there is a general shift in the point estimates towards 1 and that the confidence intervals broaden, but there are no major differences after adjusting for RNA and new drug covariates, and the majority of the point estimates, again with the exception of 372, group b, fall less than 1.

[Slide.]

Lets shift to the baseline number of nucleoside mutations. The odds ratio is shown for additional mutation. This is adjusted for other mutations in classes used in the regimen. In this dataset, we see that the point estimates generally are above 1, meaning that for each additional mutation, nucleoside resistance mutation, the risk of failure is increased.

The confidence intervals intersect 1 in several of the studies but the point estimates are consistently greater than 1.

[Slide.]

This is after adjustment for HIV RNA and new drug covariates, again adjusting for mutations in other classes. After all this adjustment, we still see that the point estimates for the majority of studies are greater than 1 although the magnitude is not as great as the GSS score.

[Slide.]

Let's move on to number of PI mutations. Similar analysis. Point estimates are all greater than 1, unadjusted.

[Slide.]

And adjusted, we see some diminution of the effect of the number of PI mutations but, still, there is consistency among the data in that the number of mutations in protease is associated with a greater risk of failure.

[Slide.]

When we looked at phenotypic, the first analysis we did was the four-fold phenotypic sensitivity score meaning that if the IC50 was less than four-fold, there was one point, if greater than or equal to four-fold zero points. This looks at the four-fold PSS score odds ratio per one unit increase adjusted for baseline HIV RNA and new

drug covariates, unadjusted in white, adjusted in yellow.

What you see here is consistency, again, in the relationship between the PSS and the risk of failure. Each of these point estimates substantially less than 1, meaning the higher the PSS, the lower the risk of failure and, before adjustment, one of the analyses intersects 1, after adjustment 2.

But I think the important point is not how many confidence intervals intersect 1 but the consistency of the point estimates of the odds ratio.

[Slide.]

When we look at the ten-fold phenotypic sensitivity score adjusted for RNA and new drug covariates, we see a little less effect of that variable in the model. The point estimates are generally unadjusted, less than 1, and there is some movement after adjusting for RNA and new drug covariates.

But, in general, the magnitude of the effect is less for a ten-fold cutoff than a four-fold cutoff.

[Slide.]

I am going to show you just a few slides looking at the reanalysis of the prospective studies, VIRADAPT and GART. Each of these studies is broken down by whether there was genotyping or no genotyping. Here, GART, no genotyping, VIRADAPT, no genotyping; GART genotyping and VIRADAPT

genotyping.

This looks at baseline HIV RNA unadjusted and adjusted for GSS and new drug covariates. The dropout is failure analysis. What we see is that for GART, it is consistent with our previous datasets that the higher the RNA, the greater the risk of failure for VIRADAPT. In the no-genotyping arm, we don't see that effect.

For GART, in the genotyping arm, there is a curious effect in that the baseline RNA is no longer predictive of failure and suggests that the intervention altered that effect, meaning that the resistance test interpretation and therapies prescribed negated the effect of baseline RNA on outcome.

In VIRADAPT, in the genotyping arm, we see an effect similar to previous studies that I presented showing the greater risk of failure with higher RNA.

[Slide.]

When we look at the baseline genotypic sensitivity score, again for the same studies, divided by treatment arm, we see that, with the exception of the VIRADAPT no-genotyping arm, the point estimates are less than 1 suggesting, again, that the higher scores associated with a lower risk of failure, broad confidence intervals and some intersect 1.

[Slide.]

So that is a composite data analysis of the retrospective studies and prospective studies that we did. Now I would like to turn it over to the individual presenters.

[Slide.]

There will be presentations first by Mike Para and then the'other individuals listed.

Thank you.

DR. HAMMER: Thank you.

Mike?

Key Retrospective Studies **Analyzed in a Standard Fashion**

ACTG 333

DR. PARA: Good afternoon. Thank you for the opportunity of presenting briefly the results of ACTG 333 which was the study titled The Antiviral Effect of Switching from Saquinavir, Hard Capsule, to the New Formulation of Saquinavir Versus Switching to Indinavir After Greater than One Year of Prior Use of the Saquinavir Hard Capsule. This will be the DAP analysis that you just heard about.

[Slide.]

The objectives were to determine if, after prolonged use of the saquinavir hard capsule, Invirase, there was a fall in viral load upon switching to either indinavir or saquinavir, soft gel, the Fortovase. The

objective of the DAP was to determine if the amino-acid substitutions at the protease positions associated with in vitro saquinavir or indinavir resistance at baseline were predictive of the RNA response.

[Slide.]

This is the protocol design. Eligible patients had been on saquinavir--that is, Invirase--for more than 48 weeks prior to study entry. Of note, they had no antiretroviral change for the two months prior to admission and they were naive to the other protease inhibitors.

Upon entry, tney were randomized in an open-label fashion, to continue their saquinavir hard capsule or have a single drug switch, the protease switch, to the saquinavir soft gel, or a single-drug switch from the hard capsule to indinavir. Viral loads were then followed for the next six months.

For the purpose of the DAP, a success was being suppressed below the level of 500 at 24 weeks. Within the DAP, because these individuals were only placed on this drug for sixteen weeks, they were not further analyzed. In the study, patients who failed this were allowed to cross arms. Those failures were considered in the DAP analysis as failures.

[Slide.]

Baseline characteristics of the population are

here. Note that these are extensively treated patients with over two years of prior saquinavir, hard capsule. They had been on greater than or equal to, at the time of entry, to nucleosides. 85 percent were on two or more nucleosides. Their baseline viral load was 4.1 log-base 10. Their baseline CD4 count was 240.

[Slide.]

The next slide shows the virologic response for the first eight weeks. You can see with a single drug switch from the hard capsule to the saquinavir soft gel, viral load fell about 0.25 logs at eight weeks. With a single-drug switch to indinavir, viral load fell about 0.6 log at eight weeks.

[Slide.]

The purpose was to look at the predictive power of the baseline genotype. We did genotyping by two methods. One was population-based sequencing performed by Charles Bouchet with ABI sequencer. The second was a multiple PCR independent clonal sequencing also with ABI which was performed in the laboratory of John Condra.

The amino-acid positions that were analyzed were those described in the DAP and are shown here. Of note, if people were found to have a mixture of both wild type and resistant at a particular position, they were considered to be mutant at that position.

Differences by the two methods were resolved by repeat sequencing.

[Slide.]

The next slide shows the distribution of the number of protease mutations for the 89 subjects enrolled in the study. As you can see, in spite of the fact that these patients had a median of two years of prior saquinavir hard capsule, 31 percent had none of the mutations that were selected to be analyzed.

[Slide.]

The next slide shows baseline covariates and virologic response. This is the baseline RNA. As you can see, with increasing baseline RNA, there was a greater likelihood of virologic failure, as John just said. That is not surprising. But you can also see with increasing number of protease mutations, there was also an increasing likelihood that these individuals would fail their new drug regimen.

[Slide.]

Statistical analysis of this shows that these relationships were highly statistically significant in that a one-log change of baseline RNA was associated with a seven-fold increased risk of failure. Also, for every additional protease mutation, there was a two-fold increased risk that they would not be successfully suppressed.

In the multivariate analysis, it was found that these two were independent of each other; that is, the more mutations, the less likely they would be suppressed.

[Slide.]

The next slide is my conclusion. RNA change **was** greater for indinavir than saquinavir, soft gel, and saquinavir, soft gel, than the hard capsule. I didn't show this data but we found that there was a trend between increasing number of protease mutations in greater saquinavir prior use, more mutations the higher the baseline RNA, more mutations the lower the baseline CD4 count.

What we showed here was the number of protease mutations and the viral load at baseline were independently predictive of RNA response. Patients with 0 or 1 protease mutation had a greater RNA decrease. Patients with 2 or more protease mutations had a blunted RNA response.

[Slide.]

Lastly is the team, shown here. Ann Collier was the vice chair. Bob Coombs was the protocol virologist, and then our virologists and collaborators.

Thank you.

DR. HAMMER: Thank you.

Dr Ait-Kahled will present the CNAA2007 study.

CNAA2007 Study

DR. AIT-KAHLED: Good afternoon.

[Slide.]

I am going to present the retrospective analysis of the baseline resistance in CNAA2007 to investigate the predictive value and the predictive quality of genotype and phenotype data on biological response to salvage therapy in HIV-infected patients.

[Slide.]

The objective of that study was to evaluate the treatment with abacavir, amprenavir and efavirenz as salvage therapy in heavily experienced patients who experienced viral rebound on the PI-containing HART.

Ninety-nine subjects were exposed to the study drugs and patients in this trial were heavily pretreated with 44 percent NNRTI-experienced. This reflects the inclusion of an NNRTI-naive patient in the protocol. 50 percent of patients had experience with three or four PIs and over 70 percent had prior experience to four to five NRTI.

[Slide.]

The methods used in that study were as follows; the plasma HIV RNA was measured by the Roche Amplicor. Genotyping was carried out by a plasma-population sequencing using the ABI technology. Phenotyping was obtained by the recombinant virus assay using the antiviral grown at Virco. The statistical analysis was carried out according to the

RCG data-analysis plan using the virological endpoint of HIV RNA greater than 400 copies/ml in the week 24 time window.

[Slide.]

The baseline characteristics of the study population is shown here. Genotyping was obtained for 94 patients and phenotype values were obtained for 64 patients. Both populations had high baseline HIV RNA of about five logs or 100,000 copies/ml, a low baseline median CD4 count of around 100,000 cells/microL. The prior NNRTI usage was similar in both populations with 40 percent being NNRTI-experienced prior to study entry.

[Slide. 1

In accordance with the numerous prior therapies of these heavily pretreated patients, a large number of mutations were found in reverse transcriptase and protease. 65 percent of the baseline isolates had more than four nucleoside reverse-transcriptase-inhibitor-associated mutation. 45 percent had more than one non-nucleoside reverse-transcriptase-inhibitor mutation with 21 percent having more than two. 81 percent of viruses had more than four protease-inhibitor-associated mutations with 50 percent having more than five PI-associated mutations.

Looking at the baseline phenotypic data for the study drug in CNAA2007 at the two different cutoffs, at the four-fold cutoff, 45 percent of the isolates were sensitive

to abacavir, 58 percent sensitive to amprenavir and 67 percent sensitive to efavirenz.

Looking at the ten-fold cutoff, over 70 percent of the baseline isolates were susceptible to all study drug according to that cutoff, so 94 percent to abacavir, 86 to amprenavir and 73 to efavirenz. In that particular study, the ten-fold cutoff had less discriminatory power to discriminate the baseline isolates according to phenotype.

[Slide.]

I am now going to present the results of the various univariate and multivariate analyses described in the data-analysis plan for both the dropouts-as-failure and dropouts-as- censored population. This graph--luckily, Dr. Mellors had made an excellent job by describing the way this data is presented.

The graphs here show the odds ratio or the risk of virological failure and the X axis represents the variable used in the model presented. The risk of 1 is marked for comparison and the 95 percent confidence intervals are also shown.

So here in the univariate analysis, you can see that a higher baseline HIV RNA is, as stated, with an increased risk of failure. The introduction of a potent new drug and, in this case, efavirenz in the NNRTI-naive patient is associated with a lower risk of failure, about five-fold,

and the higher genotypic sensitivity score, again, associated with a decreased risk of virological failure.

[Slide. 1

The multivariate analysis using those three variables has shown that the genotypic sensitivity score remained the only factor associated with virological failure.

[Slide.]

Looking at the genotype analysis using the number of mutations for each drug class, the mutations for NRTI, NNRTI and PIs, after adjustment for baseline RNA, potent new drug, one can see that the number of NRTI mutations and NNRTI mutations remained the only factor associated with an increased risk of virological failure.

[Slide.]

Looking now at the phenotypic analysis, this is the univariate analysis of each of the variables with the baseline HIV RNA, potent and the four-fold cutoff phenotypic sensitivity score and the ten-fold cutoff sensitivity score. What is apparent here is that increased baseline HIV RNA was associated with increased risk of virological failure and a higher phenotypic sensitivity score at the four-fold cutoff was associated with a decreased risk of failure but not the ten-fold cutoff.

[Slide.]

In the multivariate analysis, if you concentrate just on the left-hand side of the graph, looking at the phenotypic sensitivity score based on RNA and potent new drug, the baseline HIV RNA and phenotypic sensitivity scores remained the only factors associated with virological response.

[Slide.]

In this multivariate analysis, looking at the phenotypic sensitivities for each of the drug classes, the baseline HIV RNA, an increased baseline HIV RNA, was associated with an increased risk of about six-fold of virological failure. NRTI sensitivity score at the four-fold cutoff and the NNRTI sensitivity score were--sorry; in this multivariate analysis, the only sensitivity score that remained associated with a decreased risk of failure was the sensitivity score for the nucleoside reverse-transcriptase inhibitors.

The possible explanation for the loss of association between the NNRTI sensitivity score and virological failure is probably the use of the presence of confounding variables in the model such as the use of potent new drug which was the use of the efavirenz in NNRTI-naive patients, so, in other words, looking at two different variables of two identical effects and of two different variable measures.

[Slide.]

In conclusion, this retrospective study has shown that the genotypic sensitivity score was the only predictor of virological response in the multivariate analysis using baseline HIV RNA and new drug covariate as the variables. The genotypic analysis has also shown that the number of nucleoside reverse-transcriptase inhibitor and non-nucleoside reverse-transcriptase inhibitor mutations were also associated with biological response after adjustment for baseline HIV RNA and new drug covariate.

[Slide.]

The phenotypic analysis has shown that, in the univariate analysis, the four-fold cutoff was more predictive of virological response than the ten-fold cutoff but here, again, it was population-dependent.

Finally, the study seems to suggest that, in heavily pretreated patients, the phenotypic sensitivity score using the four-fold cutoff was the best predictor of virological response independently of baseline HIV RNA and new drug covariates.

[Slide.]

Finally, I would like to acknowledge all the investigators, team members and patients who participated in this study.

Thank, you for your attention.

DR. HAMMER: Thank you very much.

Again, we will hold questions. Dr. Veronica Miller will now present the study results from the Frankfort Cohort analysis.

Frankfort Cohort

DR. MILLER: Hi.

[Slide.]

What I will be talking today about is the association between phenotypic resistance and virologic response to mega-HART regimens in patients from the Frankfort HIV cohort. What I want to point out here is that this is going to be different from the other trials that you have just heard about because here we have a total mixture of the kinds of pretreatments these patients have seen and, also, complete freedom **as** to the types of drugs that the patients could receive in their treatment.

[Slide.]

So the Frankfort HIV cohort consists of all patients attending the clinic who had presented at least once since January 1 of 1995. Data pertaining to antiretroviral history, immune status, HIV-l viral load, HIV-associated clinical events and demographics are in this database.

Patients with multiple treatment failures were treated with a salvage regimen consisting of six or more

antiretroviral drugs and phenotypic resistance tests were performed from stored plasma samples.

[Slide.]

So, in terms of the patients and methods, we had 50 patients who met the definition criteria and the DAP requirements. Viral load was assessed in the Amplicor system. Resistance was assessed using the Antivirogram. Virologic failure was defined as having more than 400 copes at the week-24 window. And the statistical analysis basically followed the DAP requirements very closely.

[Slide.]

Here are the baseline characteristics of these patients. As you can see, they were relatively advanced with the virus meeting viral load of 5.52 and a CD4 cell count of 95. They had been on HART for a median time of 18.3 months.

[Slide.]

This is the summary of previous ART divided as to less than one week, one week to one year in yellow and more than one year in red. So you can see drugs like ZDV, 3TC these patients had seen quite a bit and, also, 90 percent of these patients had been on HART for more than one year.

[Slide.]

The susceptibility status at baseline looked like this. As you can see, there was a lot of resistance to most

of the drugs, a lot of resistance to the nucleosides, especially ZDV. 3TC; most of the patients did have resistance to 3TC, cross-resistance to abacavir although this patient population hadn't seen this drug very much, extensive resistance to the non-nucleosides as we presented here as well as to the protease inhibitors.

[Slide.]

These were the drugs that were used on treatment. About 90 percent of the patients used six or seven drugs and 10 percent of the patients used eight or nine drugs. You can see that all of the drugs were represented and 100 percent of the patients did receive protease inhibitors, most of them two and, in some cases, even three.

[Slide.]

This is the univariate analysis where we looked at the variables of RNA, number of new drugs, sensitivity of all the drugs with the cutoff of 4, the same with a cutoff of 10, NNRTI sensitivity cutoff of 4, PI sensitivity cutoff of 4 and the same with the cutoff of 10.

But we see what comes out statistically significant is the RNA level using both types of analysis, the sensitivity cutoff score of 4 as well as a 10. Then, when we get into the individual drugs, the PI sensitivity of 4 and of 10 remains statistically significant. So we see about a two-fold increase in the OR here and about a

50 percent reduction in the OR for the amount of sensitive drug that the patients were receiving.

[Slide.]

Here we have the different multivariate analyses. We first looked at an RNA level at baseline, the number of new drugs, sensitivity score of 4 of sensitivity score using the ten-fold cutoff assay. What remained significant was the RNA with about a four-fold increase in the OR and the sensitivity score at 4 as well as at ten with about a 60 percent reduction in the OR.

[Slide.]

Looking at the multivariate analysis 3 and 4, in this case we also kept RNA, number of new drugs, but now looked at the sensitivities to NRTIs and 2PIs using a cutoff of 4 or of 10. In this analysis, the only thing that really remained significant was the PI sensitivity either at 4 or at 10.

[Slide.]

So, in conclusion, in this retrospective analysis, based on a small group of extensively experienced patients, phenotypic resistance was a significant predictor of virologic response independent of baseline virus load and previous antiretroviral history. Either a four-fold or a ten-fold cutoff provided significant independent information.