In models looking at resistance within individual drug **classes,** resistance to PI was more predictive than resistance to NRTI.

Thank you.

DR. HAMMER: Thank you very much

John?

Summary of Key Points

DR. MELLORS: Thank you, Veronica.

[Slide.]

I would just to like conclude the session by review of prospective studies in progress. There are quite **a** few. They **are** listed here by name, location, design and status. RESA 2026 is closed and will be presented at the retrovirus conference. VIRA 3001 just went through an interim analysis and the FDA has asked me to tantalize you with three slides from that interim analysis.

CERT compares phenotyping/genotyping to standard of care. It is fully enrolled. CTCG 575 is phenotyping versus standard of care. It is enrolling. Nick Hellman from ViroLogics told me this morning he thinks it is fully enrolled. There is NARVAL, SEARCH, HAVANNA, which is a two-by-two factorial design of genotyping versus phenotyping versus standard of care. They are all enrolling.

ERA is a study in the UK, phenotype/genotype versus standard of care. It is opening soon. A5076 is

phenotyping versus genotyping versus both. That is in development. So there are really quite a number of studies that are ongoing or planned.

[Slide.]

I would like to just spend the next few minutes to present some preliminary results from VIRA 3001 which was an open-label, randomized trial comparing the effect of phenotyping, immediate phenotyping, on viral load compared with standard of care.

[Slide.]

Here is the study design. Individuals were screened. There were no therapy change permitted during the screening period. After the baseline individuals received treatment based on the Antivirogram or standard of care, there were 274 patients enrolled and follow up was at 2, 4, 8, 12 and 16 weeks.

[Slide. 1

The patient population studied was prior therapy history of two or more nucleosides and one PI, plasma viral load above 2000 copies, stable regimen for one month prior to screening and no prior phenotypic sensitivity testing.

[Slide.]

Here shows the interim analysis results. This is the HIV RNA response in a modified ITT analysis showing all the observed data.

Let me share with you that there are about 127 patients in this analysis. By modified ITT, is those that dropped out before enrollment and treatment change--in other words, those who dropped out during the screening period were excluded from this analysis. They were balanced between the phenotyping and non-phenotyping arm. This is the median log change in baseline RNA.

What you can see in yellow is the control arm and, in light blue, is the Antivirogram arm and at each of the time points in this analysis, after week 4, there is a statistically significant difference in the viral load response amounting to half a log or greater.

[Slide.]

A more conservative analysis of the last observation carried forward, we see similar results. The control arm and antigram arm, again half a log difference or more statistically significant at each time point. The proportion undetectable at 16 weeks was 62 percent in the Antivirogram arm and 33 percent in the control arm.

The sponsor, Glaxo Wellcome, and Neil Graham are in the audience if individuals want to ask additional questions about this data analysis.

[Slide.]

Let me just end by summarizing some key points. I think the standardized reanalysis of the retrospective

studies has generally confirmed associations between baseline genotype or phenotype and virologic response. The small datasets has led to some variability in the point estimates and broad confidence intervals but, nonetheless, there is consistency after reanalysis.

The placebo intervention-based trials that you have heard about support the clinical value of resistance testing for selection of treatment regimens in experienced patients and, lastly, the data accumulating from ongoing clinical trials of improved and investigational agents will refine the interpretation and improve the predictive value of specific resistance test results.

[Slide.]

I would like to close by acknowledging the tremendous effort that was put forth under very tight time lines by all the investigators involved. I don't have enough slides or time to acknowledge all of them. Some individuals that were not mentioned were Richard Harrigan, Andrew Zolopa and other individuals that you have heard about who contributed datasets as well as their statistical support staff.

Thank you.

DR. HAMMER: Thank you very much. On behalf of the committee, we would like to really thank you and the members of the RCG. It is really a privilege to see these

data first-hand and for the first time.

Why don't we continue with Dr. Aras to wrap up this section. Then we will take a break and then we will have a question period.

Summary of Studies Analyzed by DAP:

Statistical Comments

DR. ARAS: Good afternoon, once again.

[Slide.]

You recall the data analysis plan discussed earlier by Dr. DeGruttola and Dr. Mellors. Briefly, the virological failure, as to type dichotomous variable, is the dependent variable and it is to be explained by independent variables, namely baseline covariates.

This is achieved using a logistic regression model. I will make a few comments here to remind ourselves of the limitations of retrospective analysis. My first set of comments are about odds ratios.

[Slide. 1

These are the covariates that are listed in the data analysis plan. You have already seen them. The extent to which a covariate explains the dependent variable is captured through the odds ratio. Hence, let's remind ourselves what is an odds ratio and its relationship to a related concept, a relative risk.

[Slide.]

Here is a definition of relative risk. Let p-0 be the chance of virological failure when a covariate is at a given level and let p-l be the change of virological failure when a covariate is one unit higher than the previous level. The relative risk is defined as the ratio of p-l and p-0.

For example, if the relative risk is 2, it means that the chance of failure is twice as high when the covariate is increased by one unit. The same interpretation may not be shared by the odds ratio, as we will soon see.

[Slide.]

Relative risk is the quantity of interest but, typically, not estimable from a retrospective study. Hence, a related quantity odds ratio is many times looked. An odds ratio is defined as the ratio of the odds of failure for the covariate at a higher level versus the covariate at the lower level.

The odds ratio can be estimated from a retrospective study but may not be of interest, of itself. When p-l and p-0 are small, relative risk is approximately equal to the odds ratios. In such a situation, an odds ratio can be interpreted as a relative risk.

In the present situation, however, such an interpretation of the odds ratio may not be possible since p-l and p-0 are not small for all of the covariates listed in DAP. For some covariates in some of the studies, the

elative risk can be estimated directly since the
andomization scheme took into account the covariate by
tratification.

Treatment, for *example, is a stratified covariate
n VIRADAPT and GART.

[Slide.]

Here are a couple of graphs that describe the
elationship between odds ratios and relative risk. You can
ee that when p-0 and p-l are small, odds ratio and relative
isk are nearly the same. But, as p-0 and p-l increase and
-1 is much larger than p-0, the gap between the two
oncepts increase.

You can see, as an example, if p-l is at 0.5 and
-0 is at .025, the relative risk is 2 but the odds ratio is
If the gap is even larger, if the p-0 is, say, at 0.025,
-1 and 0.75, then the risk is 3.0 but the odds ratio is
.0.

[Slide. 1

The same information as in the previous slide is
epicted here slightly differently. Here, again, you can
ee that if p-0 is at 0.4 and the relative risk is at 2.0,
nen the odds ratio is at 6.0, so it is exactly three times.
t is the odds ratios and not the relative risks that are
enerated by DAP methodology.

One is cautioned here not to interpret odds ratios

as relative risk.

[Slide.]

Let's move on to discuss the crucial table 1 in DAP. Table 1 in DAP titled Mutations Associated with Resistance to Specific Antiretroviral Drugs is the crucial table there. It is largely based on in vitro data and/or exploratory subgroup analysis. A general consensus is yet to emerge on this table.

[Slide.]

Most of the studies were available to the Resistance Collaborative Group before the DAP was developed. Hence, in principle, a possibility exists that the DAP and, in particular, table 1, was developed interactively to fit the studies. We recognize that these are useful and valid exploratory practices to generate specific hypotheses that could be tested in confirmatory studies later.

With this caveat, let's look at some of the results of these retrospective studies.

[Slide.]

For illustration, I have chosen only a few of the models from those listed in DAP. Actually, these slides are very similar to Dr. Miller's presentation and reiterate his findings; hence, I will move through them quickly.

Model C has only one covariate, namely genotypic sensitivity score whereas model E is a multivariate model in

which the odds ratios for genotypic sensitivity score are computed in the presence of other covariates; namely, new drug covariates and baseline log HIV-l RNA.

The x axis represents the odds ratio. The y axis lists the retrospective studies. The vertical line is the odds ratio equal to 1. In all these figures, dropouts as failures is depicted. The analysis was similar for dropouts as censored.

The 95 percent confidence interval entirely on one side of the vertical line at 1 indicates statistical significance for that odds ratio. More adjustment for nultiple comparisons are made here. Notice that the nultivariate model, in most cases, has generated slightly wider devices and hence, sometimes, the statistical Significance is lost. But, overall, you see the same consistency that Dr. Miller pointed out.

[Slide.]

Here we have the 95 percent confidence intervals for the number of PI mutations. The number of studies vary across the slides since data was either not available to us on all the covariates on the studies or a specific covariate was not relevant in a given study.

[Slide.]

Here we have confidence intervals for odds ratios with phenotypic sensitivity score as the predictive

covariate, again a univariate setup and a multivariate setup. The SA minimum cutoff is the phenotypic that has been looked at.

[Slide.]

These exploratory analyses suggest a predictive link between virological failure and genotypic and phenotypic measures at the baseline. They will provide insight in generating specific hypotheses to be tested in future confirmatory studies. We look forward to the results of DAP from ongoing prospective studies that Dr. Mellors mentioned.

Thank you.

DR. RAMMER: Thank you very much.

This was really very impressive. In order to avoid the risk of the data-overload meter getting into the red zone, I am going to suggest that we take a twenty-minute break and return at 3:30-ish.

[Break.]

DR. RAMMER: I would like to officially reconvene the committee session for this afternoon. We will do this in two parts. The first will be a little bit of a discussion of the data presentations and some clarifications with questions and discussion, and then there are questions specifically for the advisory committee to consider that we will bring to open discussion.

I would like to initiate the second half of this afternoon's session by asking Dr. DeGruttola, perhaps, to comment on some of the statistical issues that were raised in Dr. Aras' presentation: Perhaps we can have some clarification for the group based on some of the issues that we didn't have a chance to discuss.

DR. DeGRUTTOLA: There are a couple of issues that Dr. Aras raised. As he correctly pointed out, an odds ratio is not a risk ratio and shouldn't be interpreted that way. So if you see an odds ratio of 2, you shouldn't say the risk had doubled unless the risk was, in fact, very small.

On the other hand, as Dr. Andrew Phillips was just pointing out, it isn't true that, in this case, we were required to use an odds ratio rather than a risk ratio because this was a retrospective study. If, in fact, we had been confident that we could collect all the data that would allow us to get not only whether or not there was a virologic failure but the time of virologic failure, it would be possible, from the retrospective data, to get a risk ratio.

The reason we didn't was twofold. One was the primary goal. was to get consistent analysis across all the range of studies and keep things as simple as possible. From that perspective, it make more sense to define a yes/no variable of virologic or not rather than trying to find

exactly when those failures would occur, which can get pretty complicated.

The other thing is that it was easier to do the adjustments that we spoke of for baseline covariates, new-treatment covariates, et cetera, by keeping it as a dichotomous variable. And odds ratio is, we thought, the appropriate way to present data when you have a dichotomous outcome. Although you cannot interpret the odds ratio as a risk ratio, as we have said, it is, nonetheless, a valid measure of association between the covariates and the outcome and that is why it was used.

On the second point, to what extent that Dr. Aras mentioned--to what extent was the DAP developed specifically in response to some of the features of the individual datasets regarding the mutations that were important. Was it the case, for example, that investigators knew that the 215 mutation was important in six of these studies, and therefore, we wanted to make sure to get the 215 mutation into the table to confirm what we already knew?

I don't think that describes the process at all. I think that the table of mutations was based on experience of virologists and investigators over a broad range of studies and did not reflect just was important in these individual studies.

So I think that this analysis does move closer to

the kind of confirmatory analysis that Dr. Aras mentioned rather than simply an exploratory analysis although I would agree with Dr. Aras that we need to apply this methodology 'to other studies where the data haven't even been developed yet to try to learn more about the association between these different clusters of mutations and outcome.

But, in this particular case, the tables were not developed based on specific knowledge about these studies. In fact, I think the fact that we get fairly consistent results across the studies even though the mutations that were important in those studies was highly variable, the nature of the drugs being used and so on tends to support the value of these analyses as providing some confirmation.

But, perhaps, Dr. Mellors or Dr. D'Aquila would like to comment further on the development of the table.

DR. MELLORS: I think, Victor, you have described the process well. It had very little to do with the studies to be reanalyzed. This was, really, a collective consensus of, Rich, how many individuals, a dozen or more--a dozen or more individuals some of whom had no, really, knowledge of the studies that were even to be reanalyzed. I think is the important thing.

The other important thing is that, in terms of the phenotypic analysis, we used the assay minimum cutoff that was defined by the company and that seemed to be validated

in this study. That was done without knowledge of the phenotypic information from the reanalyzed studies.

Maybe Rich would like to add some subtlety or major point that I have missed.

DR. D'AQUILA: Richard D'Aquila, Mass General Hospital, Boston. No; I don't think anybody missed anything major. I just wanted to point out another aspect of what was mentioned about the mutation. There were different people involved in these exchange of e-mails, both virologists and clinicians.

I think the bulk of the data that led to defining these mutations came from in vitro studies of site-directed mutants with the specific mutations. I think some of the people who were involved in these discussions were also aware of clinical data that correlated viral load responses with specific mutations, but I think the bulk of the information really came from study of site-directed mutants in vitro and I think that data is relevant and essential.

DR. HAMMER: Thank you.

Dr. Aras, did you have any other questions that you would wish to have clarified?

DR. ARAS: Yes. I do agree with most of you. The data shows tremendous consistency. I did agree with that. My only fear was that, in the epidemiology of rare diseases, typically logistic regression is fitted and there odds ratio

and relative risk are one and the same because the probabilities are very small.

We did see a couple of speakers who did talk about odds ratio as a five-fold increase, three-fold increase. So there is a tendency there. I just wanted to point that out.

DR. HAMMER: It is important.

DR. ARAS: Just to keep that in our mind.

DR. HAMMER: Thank you.

Before we get to the six questions that have been addressed to the committee, I would like to give the committee members a chance to ask some questions. Once again, I would like to thank the speakers that we had this afternoon and the privilege to hear the VIRADAPT, GART and RCG analyses here and particularly some of the new data analyses that have not been presented anywhere else previously. All of that is quite a privilege.

It also created data overload to some extent so it is a bit hard to absorb everything from the RCG analysis in this large group setting and frame questions but I would, perhaps, in order to first have some takeaway point beyond the consistency of the analyses maybe to ask Dr. Mellors one question.

If one could look at some of the resistance analyses that were included in the analysis plan which was on phenotype four-fold and ten-fold cutoff, a genotypic

susceptibility score, numbers of resistance mutations, are you willing to make a broad conclusion as to relative prioritization of the predictability and relationship to outcome for those measures across the studies.

I think some of us can draw conclusions from those slides but they went by quickly and you know the data better than anyone.

DR. MELLORS: I think the GSS and the PSS were the most consistent. I think that we were limited in the individual class genotype or phenotype by the number of studies that could look at that. I did not show anything for NNRTIs because there were only two studies that could analyze that.

I think the take-home message from the DAP analysis is that some score related to genotype or phenotype is very important for determining response. I think it is not just the DAP. It is John Baxter's GART. It is other studies that have looked at it. It is prospective and retrospective. But I would say that those two fell out most importantly.

DR. YOGEV: Can I follow up with that. Did you make a correlation between the GSS and the PSS?

DR. MELLORS: No; not in this analysis. There was no analysis of the relation between GSS and PSS.

DR. YOGEV: Are there any plans to do that? The

reason I ask is it may be more practical because phenotype today is a thousand and the genotype is cheaper. Should one just go to one, two? This is a practical issue more than scientific.

DR. MELLORS: There is no reason we can't do correlations. You saw some data earlier today from ACTG 372. I think Doug Richman summarized it very nicely. The correlation will never be 1.0 because the virus is always a little bit ahead of us. But, as databases build, we can improve the correlation. This is one dataset that can be used but it is a relatively small dataset.

DR. HAMMER: Clarifying questions, again on the afternoon's presentations.

DR. WONG: John, I was very impressed that the data really seem to be very consistent. But I am concerned that they are aggregate data, kind of all drugs at a time analyzed together, study by study.

Can you tease out from all of the studies that were analyzed in the metaanalysis any conclusions about individual drugs by crossing through studies and looking at individual drugs.

DR. MELLORS: The one study that looked at an individual drug was the abacavir pooled data. Otherwise, I think we are limited again by what is the lack of frequency or the low frequency and number of studies in which a drug

was used consistently. So it makes it different.

We could look at classes and, depending on the prior treatment experience and what was used in the new regimen, we saw, in the example of ACTG 333, a very important relation between protease and response to saquinavir or indinavir.

So there are, within the aggregate data, examples of individual response, abacavir, saquinavir, indinavir. The problem gets into the fact that new therapies, particularly in experienced patients, are not going to look at individual drugs. They are going to look at combinations. We talked about that this morning and it is a problem.

So there are bits and pieces but we can't make statements about each of the fourteen antiretrovirals; no.

DR. HAMMER: I guess it is fair to say that, even with certain studies like 372 where individual drugs were looked at in relation to virologic response in the study-specific analysis, no relationship was found although a summary measure of susceptibility, there was a correlation. So, in fact, within a study you can look at the question you raised. In fact, it creates the complexity of needing a summary measure of susceptibility in combination regimens or at least looking to that as helpful.

DR. MELLORS: I think Scott raised an excellent

point that I missed. In 372, the relationship between individual drug susceptibility and response was not apparent. It is only in aggregate that we see a correlation with response because the response is dependent not on single drugs but multiple drugs.

DR. POMERANTZ: I am going to let John sit down, finally. Two of the trials gave some very surprising findings. The one was led by Dr. Khaled and the other one by Veronica Miller, Dr. Miller. The one by Khaled showed that when you looked at phenotypic analysis for virologic failure that the four-fold rather than ten-fold difference was statistically significant.

Then, when you looked at the one by Miller, either four or ten-fold phenotypic differences were statistically significant. Now, based on this morning's discussion, I am sort of surprised at that because part of that is trying to dissect out what is a level of phenotypic resistance that has clinical correlates. This is some of the first that showed a true cutoff.

Maybe Doug or Dr. Little would like to comment on these two trials.

DR. RICHMAN: Those conclusions are, to some extent, a function of the population on whom--the data that Dr. Ait-Khaled showed was that the ten-fold cutoff, this 80 or 90 percent of the patients, didn't fit about that

threshold so that there is no discriminatory power to show a difference while the data of Dr. Miller, that breakpoint might have been different so you have had the discriminatory power.

So I don't think that that observation that his data showed a difference between four and ten addresses the question about individual drugs. I think we know from the abacavir monotherapy data that probably four or six-fold is enough to kill it.

With a number of drugs, I think you just need a Eew-fold to probably make it inadequate but there may be others in which that threshold, you have more margin of error and you can have a higher cutoff.

So there are two points. I think, in general, four-fold has more discriminatory power and clearly was as rood or better in all of these studies but for making decisions about individual drugs, we still need more data except we have the benefit of the abacavir which had a lot of monotherapy experience that they could look at retrospectively.

That is not a privilege or an opportunity we are going to have with future drugs, I hope.

DR. POMERANTZ: Maybe we will talk about it more omorrow, but then why pick tenfold as a cutoff for true esistance rather than four-fold or three-fold.

DR. RICHMAN: No; I wasn't. The reason the analysis had two different cutoffs is because we didn't know what the right one was and we were interested in finding out. I think one other point that has been mentioned is the pharmacologic issue where, with, for example, a number of the protease inhibitors that have significant pharmacologic enhancement by ritonavir, as monotherapy, a low level may be enough to bump them off where you may be able to get away with a higher level with ritonavir.

But that is the type of information that is going to have to be generated in study designs in order to convince all of us that that is true.

DR. POMERANTZ: I still remain a little bit confused. If you were to look now at phenotypic analyses and pick it, this is a resistant virus, this is not a resistant virus, would you pick it at four-fold for most drugs?

DR. RICHMAN: If I were managing a patient and got some phenotype results, I would try to pick a combination of drugs that were all less than four-fold. But, if I had no choices--

DR. POMERANTZ: Thank you.

DR. HAMMER: Perhaps Dr. Ait-Khaled and Miller might want to comment since their studies were mentioned.

DR. MELLORS: Roger, I think we have to be careful

in this analysis. The important thing I tried to stress was that the point estimates are not that different between four- and ten-fold. The confidence interval is greater because the proportion of individuals who have greater than ten-fold resistance is smaller.

When we are dealing with tiny datasets of 50 and 70 patients, the confidence interval can result in it intersection 1 and the holy grail of the p-value is lost. But I don't think we should be too swayed by that. I think if you look in the 2007 study that the point estimates were very similar between four- and ten-fold.

So I really think--maybe Victor would like to comment on that, agree or disagree. The other thing--I think Doug said it. For many of the approved drugs, particularly the approved PIs, a four-fold change in susceptibility may be critical because, at trough, there may be three- to five-fold trough to IC50 ratio.

With a newer generation of protease inhibitors, ten-fold may not be the cutoff. It may be thirty.

DR. DeGRUTTOLA: Just to comment briefly. I agree with what John said. I wanted to go back to what the objective of the DAP was, too, which was just to say that you could find these summary scores, either phenotype or genotype, that did show an association between the testing and the response to the drug.

They are obviously not intended to try and resolve the question of what is the cutoff, either the best cutoff overall or the best cutoff for individual drugs, which is going to take a lot more intensive research.

DR. HAMMER: Other clarifying questions?

DR. GULICK: A question for John Mellors. The GSS relies a lot of table 1. You implied in your presentation that this table engendered a lot of controversy to settle on what it was. One of the critiques from the agency was that there is no general consensus on the mutations in table 1.

How do we reach consensus on this and how close are we to reaching consensus on a chart like table 1?

DR. MELLORS: I think it depends on who you are trying to reach consensus with. At home, it might be relatively easy but, in a complicated field like HIV, I think it is very, very difficult.

And we will never reach consensus on every mutation in a mutation table probably while we are still breathing. I think, though, that we will hammer down--no pun intended--mutations that have been shown in studies to be related to outcome.

They will be black-and-white ones, as Doug said. They will be grey ones and they will be highly contested ones. So I don't think we are close to reaching a consensus.

DR. HAMMER: Wouldn't it be fair to say that the controversy does not extend to the whole table, that 80 percent of that table was probably easily reached and that it was the fine points, if you will, of particular mutations and particular drugs that led to the endless debate, or near endless debate, that ended today. Is that fair?

DR. GULICK: That was my next question. so you think 80 percent is pretty solid?

DR. MELLORS: I think that probably 80--yes; 80 percent or higher.

DR. RICHMAN: A lot of the mutations were, as a participant in this 500 to 1000 e-mail exchange over a couple-of-week-period, these are talmudic debates over whether something is a primary or secondary mutation in protease, and it is the seventh or eighth most important--you know. Where do you put it on the table because the whole analysis plan, if something goes up or down one point, if it is on one side or the other.

But most of the debates were over three or four Little things here or there, should 3TC resistance--how do you deal with that with adefovir. In general, we agreed on almost everything.

DR. HAMMER: We should give Dr. D'Aquila a chance to make a comment.

DR. D'AQUILA: The way I think we will reach

consensus is over time to build large databases that look at genotypes and phenotypes on the one hand and viral-load responses on the other.

I think we are just beginning to build those kinds of databases. What we need are correlations now with the in vitro behavior of the virus but how well we can treat it with our current regimen. Those databases will take a long time to develop and they will never be static because our drugs will keep changing.

DR. JACKSON: Just a couple of questions on the VIRADAPT and the GART study in terms of the turnaround time of the genotype or phenotypic assays, how long that actually took. Clearly, this virus evolves over days and a lot over weeks. With the better assays, maybe we are underestimating the effect if we get these back fast.

I don't know how long it took to get these back. And then one other question was that, on the other hand, against that, there seems to be no difference, as you pointed out, in CD4 levels between the two arms later which makes you wonder whether there is going to be a clinical benefit of those.

DR. BAXTER: The average turnaround time was about twenty-one days and the average time to randomization was four weeks. In terms of the CD4-count issue, we found that patients benefitted from GART regardless of their CD4 count,

whether they were between 50 and 200, and 200 to 500.

DR. JACKSON: When you say the "benefitted--"

DR. HAMMER: I think he was talking about the CD4 response in the two arms, not the baseline stratification.

DR. BAXTER: Oh, right. Part of that effect, in the GART arm--there was more hydroxyurea used in the GART arm which could have blended the CD4 response for those patients, but also it was a very short-term study. It was twelve weeks.

DR. HAMMER: Short-term pilot proof of hypothesis, if you will; right?

DR. BAXTER: Right. We were really looking at viral-load outcome.

DR. CLEVENBERGH: For the VIRADAPT, the turnover time at the beginning of the study was quite long, about four weeks. But now, we get the results in less than two weeks. About the CD4, there was a wide range of variation between the patients and also all the patients are already on the PI regimen which has been shown to increase already the CD4 even if the viral load was raised, so maybe there is a confounding effect.

DR. PETTINELLI: I have a question for Victor. I notice that in the different models, baseline CD4 was not one of the covariates. Is there a reason for that?

DR. DeGRUTTOLA: I think it might have been

interesting to look at baseline CD4 and to look at other possible confounders as well, but the main goal here was to try and get an analysis plan that was as simple as possible that really could be conducted in **a** fairly short period of time consistently rather than to try and do analyses of everything that was potentially of importance.

Given that we wanted to include one primary indicator of patient status, baseline HIV RNA was selected. Obviously, with our treatment, virologic failure of treatment being the primary endpoint that was expected to be the most important.

DR. HAMMER: Thank you. I think we should move on to the questions that are before the committee.

Questions to the Advisory Committee

There are six questions, **a** couple of which we, perhaps, can consider together. I will read them for the record.

The first question is, for use in drug development, what are some important principles for developing algorithms that classify mutational constellations as "drug resistant?" How should mutational algorithms and/or phenotypic breakpoints be developed in the Euture?

This is quite a question, or two questions. Who would like to tackle this? Again, not everyone at once, nere.

DR. MAYERS: I would agree with Rich D'Aquila. I think, for the short term, what genotypes you like, phenotypes, is what we can do. But what we ultimately want is what correlates with failure to respond to the drug or to loss of response to a drug.

So I think that the ultimate goal would be to get genotypes and clinical and phenotypic breakpoints that correlate with loss of antiviral activity of your drug, if you can acquire that. The issue is going to be how to do that in the context of combination therapies.

The other issue, which I think is going to come into here is that all failure is not resistance and how you discriminate the patients who have decided that you drug is too toxic to live with from the patients who are failing virologically is not necessarily easy.

So some level of compliance or looking at drug exposure is going to have to be factored in when you make those breakpoints. Otherwise, you can, for many drugs, mix in patients who have stopped taking the drug because it is too difficult with patients who are losing the activity of the drug because of drug resistance.

Somehow, the company is going to have to factor both of those into their analyses of breakpoint determinations.

DR. HAMMER: You put your finger on the key issue.

How do you develop these algorithms in the current context of combination therapy and the inability, and appropriate inability, to add a single agent to background therapy and tease this out.

DR. MAYERS: I think that, probably, you are going to get more data from the patients who are failing, who have been experienced, where you can, perhaps, see the mutations that you have lost with your drug because the other drugs may not be contributing much.

But you are going to have to analyze it in the context of the other drugs the patients are receiving. That is not going to be easy, as you said. I think that, looking at your naive patients who are on the drug and failing will have some use, but the problem may be--for example, with a protease inhibitor, what you will find is that all your patients are failing with 3TC resistance which might make the company happy but isn't going to answer the question as to what causes you to lose the drug.

DR. HAMMER: I think one point that comes out is the first failure in naive subjects who have wild-type virus at baseline will, depending upon which drug you are testing, you might be able to tease out where the resistance is and it might be the test drug, in fact, because it depends on the genetic barrier. So the first failure may give you a fair amount of information, depending, of course, on what is

the drug of interest.

Other comments?

DR. YOGEV: I just wonder if you take a naive patient who starts breaking through and then starts to level off at a certain level, isn't that a population for resistance because there is some selection, maybe by the virus, to go with one resistant and not for all the three or four drugs which you are using.

So I would suggest that, if you have--you can tease out who is taking the drug for toxicity. Hopefully, we develop better mechanisms for compliance to see those out and then you have to take those who are just coming out of begin--suppose the drug is effective--and then start moving away from it. The is maybe the time to tease out.

We know, by consensus, hopefully we will develop sometime what the mutations are for the drugs which are not new, to look into the other mutations and see if there is any repetition of one. That is how we develop the resistance now, knowledge. So there is no reason not to do that.

Also, I thought that, in vitro, you can develop the resistance in vitro by increasing different factors to at least identify what part of the 1500 or 2000, whatever you are going to screen, would be of interest to look in that variation.

So I think the company should maybe supply, in the developing of the drug, what is the resistant development in vitro to then look later in vivo for this variation.

DR. HAMMER: Other comments?

DR. MATHEWS: There are some modeling things that I am sure the investigators probably dealt with but didn't have time to discuss. There is an assumption that there is a linear relationship, the way the variables were coded and the creating of the GSS and the PSS. So you don't really get a sense of the explanatory power of that compared to other ways of categorizing the variable, the cutpoints and whether there is a ceiling effect and so on.

And model discrimination and calibration; were those things looked at?

DR. DeGRUTTOLA: I think that that is a very important point. I think what we have to keep in mind is that, in developing these algorithms, there are two approaches. There is sort of an exploratory approach where you are trying to find out are there any new mutations that you don't know about, how are these mutations interacting when they affect the response to a drug, and confirmatory analyses where you just trying to show that, with some kind of metric you have developed, there actually is the relationship that you think should exist.

As Dr. Mathews just pointed out, the analyses that

we did here were quite straightforward. We weren't trying to model the specifics of the relationship between the development or the presence of certain mutations or the phenotype and the response.

In order to do that kind of modeling using something that really might be appropriate, look for nonlinearities, look for threshold effects, ceiling effects, and so on, you would need a lot more information. I think, to proceed with two different kinds of analyses, the more exploratory one in which you don't set up all the rules of the analysis in advance but you really try and let the data sort of speak to you about what is going on to try and understand these relationships.

Then, when you do understand them, you sort of iterate and then say we are going to take what we now believe is the appropriate model and apply it to another dataset.

In this particular case, once again, these analyses are done very quickly in the course of a month and the size of the datasets were small. So we really couldn't investigate in that period of time with this information, the question that Dr. Mathews raised.

But, as Richard D'Aquila mentioned, only by developing large databases and really trying to mine them for this information, can that be addressed.

DR. MAYERS: Victor, before you sit down, do you think, with the datasets and the analysis that you did, it would be possible, over time, to develop parameters for each of the individual drugs in the model for response models such that when a new drug came long, you could potentially use all the data you had previously accumulated for predictions for the individual drugs and the regimen to try and then get the parameters for the new drugs that you are evaluating?

DR. DeGRUTTOLA: I think there are two parts to that question. The first part is it appropriate to tease out the mutations that are associated with each drug and to view them separately. As long as those mutations don't interact in the way that they affect the response to treatment, that is a reasonable thing to do.

But, as I understood this morning's presentations and others, sometimes they do interact. For example, the 184V mutation has different effects on different drugs. So it may not be appropriate to try and look at the relationship between each drug and each mutation but to look at the relationship between patterns of mutations and the combinations of drugs.

To the extent that that is not true, that you can tease them out and allocate the mutations to specific drugs, then you certainly can use the information in these data and

other data sources to try and do that. That should provide a big help. Even if it is not perfectly true that they are completely independent should provide a big help towards understanding what is going on with a new drug.

DR. HAMMER: I don't think we have been particularly helpful, but let me try to summarize question 1 for the agency. This may stimulate discussion or additions, but we also need to move on.

The issue of developing algorithms and other important principles, I think one thing to keep in mind is the history here and that the table that we have been discussing and other similar tables have benefitted by the history of drug development and some of the limitations of that because, during serial areas of monotherapy or addition of single drugs that allowed the mutational patterns to be evolved and defined, and it is has been iterated several times but needs to be reiterated that that is going to become more and more difficult.

However, it would seem to me that the basic principles derived from what we talked about this morning from a number of different ways. First, the in vitro selection studies that will done for new drugs and development and the mutations that are seen in vitro and their proof of what they do by site-directed mutagenesis, those basic virology assays are important.

The testing of new drugs against panels of viruses with characterized mutational backgrounds including multidrug resistant profiles is important. Obviously, the good characterization of isolates from patients who fail therapy and taking good looks at those viruses in vitro and testing them, Looking at the mutational patterns, testing those mutations in site-directed mutagenesis assays and phenotyping them are all important.

I think it is also important to mention that relational databases are also going to help us, organizations. Both commercial and academic are involved in developing large relational databases that will help us determine, or at least make hypotheses about what certain patterns will look like for resistance to the drug of interest and for cross-resistance.

We didn't really talk about breakpoint issues, and that is part of this question. But I think that essentially has been answered that nobody knows the breakpoints. What we know, at least for phenotypic assays, what individual assay performance can tell us, is what is the cutoff level at which there is a clear, statistically proven difference against a control isolate.

But how that translates to activity in vivo is a complex issue related to the particular susceptibility of the drug, the susceptibility of the isolate and the drug

exposure. I think one thing, perhaps, to reiterate, is that we do need more and more data about achievable levels in vivo and be able to relate IC50s or IC90s or IC95s to that and then make some determination of how many fold above that we want to be.

So I think no simple answer about determination of breakpoints can come but just as MICs are different for various bacterial agents in relation to achievable levels and activity of drugs, et cetera, the same thing holds here.

Does anyone want to add to that modest summary? Dr. Murray is not happy.

DR. MURRAY: No; I am happy with it. I just wanted to know logistically if anybody would like to comment who and how could this be accomplished. Should it be left up to the sponsors to develop their own individual breakpoint for their drugs? Ideally, would it be nice to have a committee?

DR. HAMMER: I wasn't really being facetious earlier when I mentioned the NCCLS because I actually think that, at some point in our development, we need to bring a consensus body together to try to actually put a dataset together that can be used and built upon.

That would be my own feeling.

DR. CHARACHE: I am also on the NCCLS direction at this point. They have just been working, as you know, on

susceptibility testing for other viruses. Some of the thoughts might be helpful putting in here.

But I did want to add just the practical consideration that it really will be key to standardize the methodology that is used to determine phenotypic breakpoints or whatever of the IC50s and IC90s because I am not certain, in some of the various reports we have heard today, whether all are using the same methods. It is the details that would make a difference between the results and whether four times or ten times makes a difference.

DR. HAMMER: I would just say that, of the two recombinant methodologies that are out there, the first-level cutoff and the second-level cutoff vary with the manufacturer but they are clearly definable and, I think, well characterized on an assay-performance basis. They differ slightly one to the other, but they are pretty clear cut.

Does anyone else want to comment on Dr. Murray's question?

DR. IACONO-CONNORS: Before we move on, I just wanted to ask the committee sort of a basic question. It is a little bit backwards but with respect to all the genotypic data we have been hearing about today, I am curious as to whether we should be thinking about any mutations in terms of their linkages.

Most of the data that we see, it is all population-based sequencing so we really don't have a sense of their linkage. I wonder if any of the speakers have an opinion on whether we should be concerned about this.

DR. HAMMER: Since the question was first directed to the speakers, does anyone want to talk about that?

DR. D'AQUILA: I do have a few slides about that in the talk for tomorrow morning, but just to directly address it, to date, there is good evidence that population sequencing is adequate for detecting linkage.

DR. HAMMER: Other questions or comments on No. 1 before we move on?

Question 2, which I think we can deal with reasonably quickly as far as targeted responses are, in addition to the covariates and methodology for assessing treatment response included in the DAP, what additional zovariates or methodologies should be considered when avaluating resistance data.

I think some statistical input and other clinical research input is important here. Dr. Pettinelli already raised the issue of the CD4 count. Who would like to add to this?

DR. MASUR: One of the correlates that I guess I am not clear about in terms of pharmacology is we seem to be relaying everything to Cmin in terms of the likelihood of

response. Is it unequivocal that Cmin, rather than area under the curve or some other pharmacologic variable, is the correlate that we want when we are looking for correlating drug levels to the likelihood of response?

DR. HAMMER: I will turn to Dr. Fletcher as the resident expert on this committee.

DR. FLETCHER: No; I don't think it is. I think, for the three classes of drugs that are available, I wouldn't be surprised that it is not different for each class of drug. I am persuaded that, for the PIs, troughs are important. For the non-nucleosides, I am not. There is some data on troughs but also area under the curve.

For the nucleosides, where the active moiety is the intracellular triphosphate, I am still really not sure what it might be. I think this is one of the real challenges, as you try to move forward.

Virology and pharmacology here, I think, are fundamentally intertwined. That doesn't mean that you have to absolutely understand exactly how to move forward with virology, but I think if you want to have a vision for optimal patient management and being able to understand everything that contributes to it, we are going to have to lay out a plan so that we can quantitatively understand what does pharmacology contribute, what does adherence contribute, immunology and so on.

So back the pharmacology, I think one of the real challenges is trying to understand what metric is it that is going to tell you the most about response.

DR. MASUR: I guess, then, the corollary to this is it is just going to be too complicated, as we sort out the correlation of genotyping and phenotyping to failure, to make sure that patients are, in fact, failing in the presence of adequate drug or whether they are failing for other reasons.

There is going to be a lot of background noise here when the patients are failing for all the reasons that a number of speakers have pointed out other than lack of activity of the drug, but are we just going to have to accept all that as background noise that we are going to have to surmount with these assays?

DR. FLETCHER: I think initially you might because just the virologic data alone, I am fairly persuaded that phenotype and genotype information contributes to the information on why you have success or failure. I think that question is can you do better if you add additional information into that.

I don't know to what degree--for example, take the CNAA2007 study where there is an adverse interaction between amprenavir and efavirenz. In the back of my mind, is that, somehow, contributing to this difference between a fourfold

at

resistance falling out and a ten-fold not.

I can't come up with how it might be but I keep wondering about what confounders that may be adding. I think you have to lay out a plan to start somewhere.

DR. HAMILTON: It strikes me that there is a variable that we haven't heard much of anything about concerning the durability of the benefit conferred by genotypic testing. What we have are relatively short-term studies and a limited number of tricks in our bag.

I wonder if there are not some projections that have been made about what happens next. Do we go back to time zero and the same series of events happen again? Or are we in a better position or what? I am not even now talking necessarily, although I could be persuaded to talk about it, what the long-term implications are here. I am talking just about recurrence of viral loads at relatively higher levels.

DR. JOLSON: It is a very important point. I don't think we have the data to answer that. It all relates to what the armamentarium is going to be at the second and third and fourth round. But a hint of that, of the utility of resistance testing on the second round, at least, is from the open phase of the VIRADAPT study where the deferred use of genotyping allowed a response in the second six months.

Now, that is an open phase of that study but,

again, it says that something is up even when you use genotyping in a deferred fashion. So, to me, that is encouraging although obviously not proof in a randomized fashion over a strategic trial over three or four years.

DR. MAYERS: I think an issue, as we try and relate pharmacology to resistance is going to be that we are going to have an agreed-upon methodology for protein-adjusted IC95s or IC50s because it is clear that these drugs have very different protein binding and it is free drug that is important so that if you are going to try and relate and IC95 in a relatively protein-free culture media with the patient's blood, you are going to have to do the adjustment and it would be nice to actually get an agreed-upon method of adjusting drugs for their protein binding and then to be able to make the correlation between IC90 in a test tube and drug levels attained in the patient.

DR. HAMMER: Let me ask Dr. Fletcher a follow up to that. It is obviously important but even it is more complicated. It is not just the free drug but it is the off rate and the protein binding, the tightness of the binding that may influence it, not just the proportion of free drug versus bound drug. Is that correct or not correct?

DR. FLETCHER: It is. When you get into some of the multiple, let's say, PI-PI combination or a PI-non-new combination where you now have two highly protein-bound

drugs, I don't think we understand can there be, then--if they are bound to the same binding site, can there be competition.

I suspect that the final best way to sort that out will actually be quantitation of the unbound concentrations in plasma.

DR. RAMMER: Let me ask Dr. Murray a question related to the question that has been put to us because the question includes the specifics of the DAP and the issues of covariates and methodologies, particularly statistical methodologies.

Are you referring specifically to the RCG analysis that we have seen and how those data should be looked at more thoroughly or in the next round or is this a more generic statement about other covariates that should be looked at because the RCG and the DAP is a very specific project?

I don't know if this is a project-specific question or a more generic question that is being placed to us.

DR. MURRAY: I think it was a more generic question for analyzing data that might come to the committee in the form of maybe even an NDA for a new drug where they want to include resistance data, just standardizing ways for evaluating resistance data in the future.

DR. HAMMER: I would just mention a few things. Again, it depends on the dataset. I think Dr. DeGruttola made it very clear as to the streamlined nature of this for the purposes at hand. To bring all these datasets together, streamlining was critical.

Other baseline covariates besides what Dr. Pettinelli mentioned; obviously, gender, age, some of the normal demographics would have to be looked at even if one doesn't suspect that there is going to be much of an issue. But I think subtype and other virologic characteristics as we move into more spread of viruses beyond group B around the world into areas where there is drug development going on are going to be important to look at.

But I don't know if there are any statistical issues that Dr. Woolson, maybe you want to comment on?

DR. WOOLSON: I think one thing is, as Dr. DeGruttola had mentioned, that the intent with the Data Analysis Plan was to have an overall summary the would work across a number of studies. If we are talking specifically about projects that would come before our committee for review, I think it really is good for us not to lose sight of the fact that the continuous variables, like the log of the HIV RNA, actually provide very valuable information.

We saw a number of the presentations today where we had a dichotomy based on failure--for example, the GART

summary--and also had the actual log of the HIV RNA. I think that information, in terms of those changes, is very valuable. The Data Analysis Plan, as I understand it now, doesn't speak directly to hanging on to those values, but I would imagine that, as they go through future iterations, they will probably look at other kinds of endpoints.

But I think that is one issue, that we would need to, I think, continue to focus on the continuous variables. At least, I think that is very important.

The other issue I wanted to raise about the data analysis, and maybe this is more of a specific question to Victor, is the very last slide that was up, and I think that does deal with question 2, as well, for us. Your very last slide indicated how you would apply these, this Data Analysis Plan, to intervention studies.

You didn't mention specifically that you would Look at a treatment by these covariate interactions yet, in some of the studies that were presented today, there were some very clear treatments by covariate interactions. Again, I was assuming that that was something would be worked out at a later point.

DR. DeGRUTTOLA: I think those issues are very important. But I think it would have to be left to the investigators of GART and VIRADAPT to work out those interactions and present them for those studies. Again, the

purpose of this was to just come up with an analysis that could be applied consistently across studies with a very pointed question for the issues that you raised.

I think the investigators could probably best do that on their own and, as you mentioned, some of those points were addressed in their presentations.

DR. PETTINELLI: I might just say the obvious, but when you were looking at covariates in modeling, I think we also have to make sure that we have the right number of patients to look at because I was a little bit concerned, for example, for some of the statements that were made following the analysis of the Frankfort cohort which stated that when you are looking at--like, resistance to PI was more predictive than resistance than NRTI, to look at the confidence interval at least when we are looking for fold differences.

That is the difference. I don't know if it is because there are less patients, there is variability. So, what I am saying is we have to be careful in terms of what we are working with.

DR. MATHEWS: Jeff's question reminded me that, very often, the committee and the division have to look at several studies in different populations to evaluate a product for licensure. In the future, if claims are going to be made that a product is effective in the setting of a

certain mutational pattern, you are going to have to be able to make cross-study comparisons even within the same NDA application.

That means, based on some of the VIRADAPT data and other studies we haven't talked about yet, that there has to be comparability in variables like adherence and drug levels and those kinds of things which are increasingly being incorporated into clinical trials but are not uniformly there.

Now, for example, there is data already that adherence, itself, can make over a log and a half difference, a population at three or six months which is larger than the effects that were described for genotyping testing.

So I think that these things are covariates that need to be looked at and incorporated up front in some standardized fashion into clinical trials. It also affects the issue of poolability for large cross-study comparisons like what was presented this afternoon because, unless those things are standardized in some way, they will just increase the overall variability and the quality of the studies is not the same.

DR. YOGEV: I just wondered, from the modeling, we are almost accepting now that less than 400 is the cutpoint. The clinical studies suggested there is a negative effect,

even if you have about 50. So shouldn't the modeling be changed to go to 50 as a prediction of failure to try to find that resistance that maybe there are earlier in that system than waiting for the 400 which is a convenience which I think we should leave at point.

DR. HAMMER: Just on your comment on the RCG, I think 400 was taken because it was an endpoint that could be seen across all of these studies and it wasn't making a statement that that is the way that all of these analyses should be done.

DR. YOGEV: I have no quarrel with what was done, I am saying, for the future, what covariate to look at. I was wondering should we just also lower the modeling for future presentation to go to 50. We know that when we say 50, we are still in the billions in the body. So maybe we need to go even lower. But, at this point, I think that already 50 for the future study, not for what was done.

DR. HAMMER: Any other comments on question 2 as far as far a covariates are concerned and resistance data? I am not sure I can really summarize this for the agency. I think some of the issues have been obvious, that the populations need to be well defined. RNA has been mentioned and was already included. Obviously, CD4, basic demographics, gender, age, viral subtype is important where the populations expand and, I think, most importantly, the

covariates of drug exposure or measures of drug exposure, adherence and PK.

I am also thinking, taking Dr. Mathews statement one step further in what he said earlier today, whether one is looking at resistance retrospectively in a study, particularly in a substudy, or in a prospective fashion for drug that is being developed specifically to go after resistant virus.

I think some of the issues of defining the populations and powering the studies appropriately are important when these are done in a prospective fashion to prove a drug's validity and efficacy in a drug-resistant population rather than in retrospective studies.

More and more we will see that as drugs that look particularly good as salvage agents come through.

Any additions to that?

MR. HARRINGTON: Scott, just one. I want to just go back to what Dr. Hamilton said how the patients are doing now. Ideally, if you have these huge, large relational databases, it would nice to know how the patients are actually doing physically, clinically. That would be nice.

DR. HAMMER: I don't think anyone would argue with that. At least one of the relational databases that I am aware of doesn't really have that kind of clinical follow-up data but certain groups can develop that and should develop

it, certainly, in the academic environment where people are being entered through clinical trials and long-term follow-up is being attempted.

DR. STANLEY: I just want to add to the things that can contribute to resistance that are not intrinsic to the virus. We have to keep in mind it hasn't been a big issue for us so far in HIV virology, but if we go back to other infectious diseases, thinking about things like Dr. Richman mentioned this morning, proton pumps and other things that may be intrinsic to the cells or limiting nutrients in the cells or other things like that.

Presumably, that would try to be controlled for in the in vitro preclinical stage, but those are other non-viral intrinsic resistance factors.

DR. HAMMER: Thank you. I think we are going to see a wave of information about that in the Nature and Medicine report.

We are going to move on to questions 3 and 4 which I think we can combine since they are identical except for one word; do the data presented support the clinical utility of HIV genotypic testing for use in drug development. That is question 3. Question 4 is identical replacing genotyping testing with phenotypic testing.

This is an easier question, at least to me, it is. Who would like to start this discussion?

DR. YOGEV: Yes.

DR. JACKSON: Just to play the devil's advocate here, though, based on the prospective studies of the VIRADAPT and the GART study, one could argue, getting back to this adherence issue, these were open-label studies. I assume that the patients knew whether they were on the study arm or the control arm and whether that may have made a difference in adherence in a claim, this difference.

There were also differences in the study group. Not only did they have the benefit of GART but they had the benefit of expert opinion that was based, not just one the GART results, but on the antiretroviral history.

If the control arm had had the benefit of expert opinion, just based on the antiretroviral history, would that have made a potential difference. There are also no differences in CD4 counts by the different arms, as we have already mentioned.

These were relatively sick populations with about 200 CD4 counts. Whether this really will clinically benefit this relatively sick population, I am not sure the data is really there to say that. My personal opinion is I think it probably will, but the data are the data and there are these potential confounders that I think we can't ignore.

DR. RAMMER: Just to add to that, I think the per-proportion below the copy number limit chosen in VIRADAPT

and GART at 12 weeks actually is not dissimilar than salvage-regimen studies whether it be 359 or 372 that were doing it without prospective testing. But I think, again, these were pilot groups of hypotheses studies and it relates, again, to the armamentarium used.

But the data are consistent. I think this question, I would say, although potentially primarily driven by the prospective studies, our thinking needs to be influenced by the new prospective study on phenotyping that we got a glimpse of here but also the retrospective analyses that are pretty consistent.

You have to ask yourself has any study shown the negative that the application of resistance testing leads to a worse outcome.

The people who wouldn't speak into the microphone were murmuring budgetarily and financially.

DR. GULICK: Just to follow up on what you just said, Scott. At first glance, it is disappointing that the best arms in VIRADAPT and GART have rates that are very similar to the early salvage studies where we didn't have the benefit of resistance testing.

But then when you go back and look at the studies that you mentioned, 372, 359 even CNAA 2007, what they have in common, as a big success rate, are people that were non-nuke naive and then got a new nucleoside. So you wonder if

that what it balances it out.

DR. HAMMER: It also brings up that it is more than just having a resistance test and someone to interpret it. It is intervening at the right moment at the lower viral load rather than a higher viral load, choosing the right drugs, et cetera. It is complex, as everything in this disease is.

DR. MAYERS: I don't think the clinical management question is what the FDA asked here. What they asked is what is useful in drug development and understanding how your drug succeeds and fails and, perhaps, defining populations that your drug would be better suited to and not suited to.

I think, for that area, the answer is yes and yes. And then, for how we use the test and who should use them and when, I think the answer is a bit more open. But for the drug-development issue, I think the answers are fairly clearly "yes" and "no."

DR. HAMMER: Although I think they are interlinked because the drug development process is going to use the application of resistance testing to either do randomizations up front or strategically use the drug in a population that fails. The immediate derivative of that, again, which we will talk about tomorrow, is the clinical application. So I think these are somewhat inextricably

linked.

But the question is, specifically, in drug development which, if anything, should make Dr. Jackson more happy about saying "yes" without the words in his mouth.

DR. PETTINELLI: I also agree that we are ready to use the genotypic and phenotypic testing drug development. I really refer to our discussion this morning which was presented by Doug in terms of where and when. Again, the characterization of post-treatment isolates, is important, most important, to try the specific patients that were failing, particularly if the drug is presenting some very positive effect over the specific use and population.

DR. HAMMER: Any other additions? We did touch a little bit, as Carla mentioned, on this question earlier this morning looking at applications in phase I and phase II-III.

I think this is easy to summarize. The questions 3 and 4 are "yes" and "yes."

Question 5. What are the limitations of the data correlating mutational patterns with clinical income? Clinical outcome; thank you. It brings us back, full circle, to the psychologic issues that were raised this morning that I am quite suggestible when someone mentions financial matters in my ear next to me, here. Both are true, however.

Who would like to talk about this, clinical outcome and mutational patterns.

DR. PETTINELLI: To me, the major limitation today is that we have only short-term data and I think it is very important to develop a long-term database. I don't know if the several studies that are ongoing that were discussed earlier, indeed, there was the duration. But I think that would be important.

DR. JOLSON: I just had a quick follow-up question to Carla's to direct to Dr. Mellors. In terms of the prospective ongoing studies that you showed, can you just give us a general sense of how large they are and how long the follow-up is of patients so we can kind of expect what might be coming with additional data?

DR. MELLORS: They range in size from a couple of hundred to over 500. Follow up is as far as 48 weeks. But that is the initial protocol. It could be extended.

DR. HAMMER: Can I ask a follow-up question, John? RCG, phase II, **as** far as analyzing the database further but, also, with the number of prospective trials that you have, in follow-up to Dr. Jolson's question, of bringing analyses together for prospective trials looks interesting, as well as longer-term follow up in a cooperative fashion.

DR. MELLORS: The RCG is holding out for a retirement package, a pension, similar to what you get in

the military. So, we haven't talked about phase II. We are not quite done with phase I until tomorrow at 5:30. As Doug said, we were hoping to declare victory and disband.

DR. HAMMER: But you can be reborn with another acronym.

DR. MELLORS: It will take more than that.

DR. MASUR: Just to follow up on what Mark Harrington said before. This question asks what are the limitations regarding mutational patterns with clinical outcome. I guess, at this point during the day, we have accepted the fact that clinical outcome equates to virologic status whereas, in fact, clinical outcome is obviously a much more complex issue.

When we talk about the limitations of data concerning the correlation with clinical outcome, I guess the question is do we really have some way of looking at whether or not the patient is better off far down the road in ways 'other than simply virologic load, or in terms of a long-term strategy.

DR. POMERANTZ: I think that is a very important point. We talked about it a little bit yesterday as well as this morning is that clearly it would be great if we could everyone to undetectable or to profoundly decrease their viral load. I think everyone, now, would agree with that. But with a lot of these salvage therapies where the

genotypic and phenotypic analyses are going to be used, you are not going to be able to do that, at least in this first generation or second generation of drugs that we have.

So there are these patients that are being more and more described, some anecdotally, some non-anecdotally, that do not get a great benefit in the CD4 count, do not have profound changes in their viral load, yet seem to do fairly well and may do better than those that did not have changes in their antiretrovirals.

I think that was the point is that we are not only going to follow viral load but, as this goes out, I think you are going to see this disconnect in certain patients where some will do better than you might think just looking at the viral load and CD4 count.

Understanding those parameters on a basic biological level is important but understanding them clinically with these correlations is going to be very important because a lot of these people will not have great changes in the viral load as these studies sometimes have shown even this afternoon.

So I would just reiterate that and say that I think that you are going to have more of that problem rather than less of it.

DR. MAYERS: I want to stay away from the issue of long-term benefit of resistance testing. On the other hand,

there is another resource. Both the ACTG and, I think, the CTCRA have large mutation datasets now. We are going to have over 1200 patients fairly soon in CTCRA and every one of them is enrolled in the clinical-endpoint long-term monitoring study.

So I think the issue of other mutational patterns that would be better or worse against the true clinical data, virological data, will be emerging over the next few years in both the ACTG and the CTCRA and, if we are truly smart, we will make the datasets meldable.

DR. KUMAR: I want to follow up what Dr. Pomerantz just said. We continue to see in clinical practice a number of patients who clearly have failed therapy by all measurements that they are set for but continue to do clinically well.

I am not sure taking the genotype testing and altering the antiretroviral treatment without looking into toxicity data, quality-of-life measurements, that we are really doing them a service. Without looking into those type of data whether we can say yes. On a short-term basis, there is clearly a decline in viral load. Whether that includes the long-term benefits, I really don't think we know as yet.

DR. HAMMER: Although I would just make a plea. I agree we need cohorts and long-term follow up and to see the

fruits of what we are doing over the next five to seven years in a more cohesive fashion, I don't think we are suggesting, or can go back or should go back to clinical endpoint issues here.

At least that is my own feeling. I think what is being raised is, in fact, the fact of discordance and viral fitness issues and replicative capacity when viruses are impaired. But even some of the groups that have reported discordant responses and hanging up of CD4 counts and clinical benefit, clinical status patients that have remained well over time despite virologic failure are now beginning to demonstrate that those patients do go on to failure immunologically as well.

So those data will emerge. But I think it is complicated because viral load doesn't tell you everything pathogenetically that is going on and you need to look at other factors. I think that is what we are saying and it is getting more complex than some of the issues about whether certain mutations impair the virus more than others is clearly important and one of the limitations, I think, in response to this question.

Hopefully, we will stimulate some discussion.

DR. MAYERS: I think there has been a confusion between the strategy of using a resistance test--let's just make it generic--a resistance test to manage patients, the

strategy of using a resistance test to manage patients every time a viral load becomes detectable.

I think that is a very different issue. I think that you have two--it looks like three prospective studies that show that when you make the decision that it is time to change drugs, the testing will help you to get your patient--your patient is twice as likely to go below the limits of detection with the test than without out.

But that is very different than how you choose the test and at what time you choose the test so that it is quite rational to let people float through for two years with discordance in viral load and do nothing. But when you decide you want to change therapies, I think, at least in the short term, you are twice as likely to get undetectable with genotyping or phenotyping, per your preference, at that point that you choose to change.

So it is going to be hard to address that in any serious way for long-term clinical benefit.

DR. HAMMER: Personally, I agree. I think we are falling a little into the trap, which is obviously easy in discussions like this and is the same thing we did with viral load. Everyone focussed on a number. We can't focus on a resistance result and think that you are going to manage an HIV patient properly if that is the only thing you are thinking about.

DR. HAMILTON: This just actually emphasizes the point I wanted to make which was that to extrapolate the data that was presented today, as compelling and as interesting and as important as it is--extrapolating that information to the population at large, I think, would be a big mistake at this moment.

I think we need a lot more information in terms of how this actually benefits patient outcome in very real terms. You can call them clinical endpoints if you like. I call them life. I would like to think that what we are doing is going to have a long-term impact.

How we go about measuring that, assessing it, implementing the current set of guidelines that undoubtedly will appear as a result of this impressive dataset is another matter. But I think we do not want to lose sight of the larger picture here.

DR. CHARACHE: I think, in terms of new-drug development, a very key issue here is to define what is meant by clinical outcome. Is it CD4 count? Is it other parameters? I think this will have to be agreed upon.

DR. HAMMER: I am not sure, in this question, whether it is marker RNA outcome, CD4 outcome or ultimate clinical outcome. Some of us reserve clinical outcome for--purely what is truly meant by that. I thought that was the implication of the question.

DR. KAPLAN: One thing I have always been impressed with in the GART data is, in a sense, it represents kind of a real-life situation in that about half of the providers in the GART arm didn't do exactly what they were asked to do.

This happens in clinical practice. They make their own judgments and so forth. In the study, which only went out to week 12, the curves start to come together and the percentage of patients that were undetectable becomes insignificantly different in twelve weeks. So I think, again, it raises the issue--here you have, perhaps, even a short-term virologic benefit.

Now, VIRADAPT showed a longer term, but what exactly does that short-term virologic benefit, just for three months or so, mean in terms of a long-term endpoint for the patient. I think we are hearing that over and over. I wouldn't argue about the value of phenotypic or genotypic testing and producing some virologic benefit, but question No. 5 specifically says what are the limitations.

I think we are all owning up to what the limitations are.

DR. MATHEWS: I think, also, and it is really not a limitation since these are very preliminary analyses, but I think the kind of confirmatory studies that Dr. DeGruttola was talking about are really essential if this kind of data

is going to be useful. Develop a model in one study, apply it to another dataset and see if it is equally predictive, look at alternative parameterization of the measurements to the scores that are developed and just, really, see if you can get the biggest bang for the buck out of the data that is already there.

That is going to take a lot of time and energy and I certainly appreciate all the work that has already been put into it but you can milk it for a great deal more.

DR. HAMMER: Let me just try to summarize question 5. The limitations of the data correlating mutation patterns with clinical income--clinical outcome. Something else comes to mind in my new role outside of this committee. Clinical divisional income is probably the thing we deal with on the most frequent basis. It sounds like I need to talk to John about his counseling.

The limitations, I think, in what we have seen today, the data summarized, et cetera, relate to, I think, some of the numbers we were--some of the studies don't sort of have enough numbers within certain drug-class mutational patterns to help us. So I think there are a number of issues. Dr. Pettinelli raised the issue of follow up.

Obviously, resistance mutational interactions are difficult, sometimes difficult, to infer from just looking at mutational patterns and that is where phenotypic testing

comes in. We talked about that there may be impaired--what these mutations are doing to the functional replicate of capacity of the virus which there may be mutations which, if just scored as mutations, may actually be helpful to the patient by impairing the virus.

There are, obviously, issues of what we don't know. It was only two-plus years ago that we didn't really know--or at least I didn't know about the 69 insertion. And then there was a wave of presentations at the Resistance meeting, two meetings ago, and then just everywhere.

So there are other multidrug patterns that we are going to need to know about. So there will always be limitations on a mutational pattern as well as the fact that we do not discover everything we need to discover unless we start doing multiple clones on patients.

So there are many limitations but it doesn't, I think, obviate the importance of defining the--it emphasizes long-term follow up, it emphasizes the importance of relational databases.

Other additions to that?

The last question is another classic ending question for the day; what additional clinical studies are needed to further define the clinical utility of resistance testing? Dr. Mathews talked about that in relation to datasets that currently exist and particularly in relation

to the RCG analysis, but would anyone like to add--

DR. STANLEY: I just to add the caveat, define the clinical utility of resistance testing in drug development which is, I think, what we are here for today. I think that in studies that are planned that we need to take into account the nonvirologic factors. I don't think it is too much to ask of a drug company, if they are going to use resistance testing to show data about their new drug, that they show that they have accounted for absorption and the pharmacokinetics and adherence and other issues.

DR. GULICK: One thing we haven't mentioned today; every study that we have heard today, and this is a bit tangential, but every study we have heard today took people who were failing their regimen, did resistance testing and then switched them to a new regimen immediately.

We have also been hearing, and I guess Veronica Miller has presented some interesting data, about stopping the so-called course of treatment interruption and using resistance testing in the context of treatment interruption to try to maximize future options.

I guess it is more of a clinical management, the drug development question, but it directly speaks to the question, I think.

MR. HARRINGTON: Just to follow up slightly on what Trip said, if you can restore sensitivity to some drugs

during SGIs and you can also maybe get a more potent effect on the salvage regimen which might actually make appealing for companies to study the new agent in this population, which is something that has been hard for to persuade industry to do until now.

So, using the resistance test pre- and post-treatment interruption would actually be very important.

DR. KAPLAN: In terms of what the question specifically says, and what we have been talking about today, drug-experienced patients and salvage therapy, obviously, we need to study, to look at, other uses such as in antiretroviral-naive persons. I presume we are going to be getting into that tomorrow morning from what I am seeing on the agenda, that there are such things going on but we haven't even gotten into that area, the utility of resistance testing in that arena.

DR. PETTINELLI: I think it would be important to establish the relative utility of phenotypic versus genotypic resistance assay. That is a little bit different from drug development but I think, again, it would be important to compare each other how they perform in terms of outcome of the patient.

DR. HAMMER: Thank you. It looked like there were some studies, some that I am aware of and others that I wasn't aware of, that are listed on Dr. Mellors' slide of

prospective studies that are enrolling or have enrolled or will enroll.

Other comments? I think we have really touched on this. As far as further studies, I think the important point that Dr. Stanley made about not just looking at resistance but I think taking the VIRADAPT pharmacologic study further that other cofactors are really important. I think the pharmacology study within VIRADAPT is very helpful, in fact, in showing that there are independent predictive power to resistance testing and to optimal levels, how they were defined in that study.

So other factors need to be prospectively defined within studies. Within the context of defining the clinical utility, we certainly have more retrospective analyses, but I think the key will be in the prospective studies and with the relation of drug development, it is going to be in defined populations depending upon the drug and its characteristics.

I think there are two ways--more than two ways--to do this and one is, at baseline, as far as the population that one recruits and one could, prospectively, do resistance testing to see, as it evolves--one could randomize on the basis of resistance testing if one has the drug.

We talked about this before, that may be useful,

particularly with a particular genotype or phenotype. And then one could use it strategically in a clinical trial as far as management which gets into how long the benefit and the continued benefit of resistance testing can be in first failure, multiple failure and over time.

So I think the studies that we have seen so far, retrospective and prospective, are really the tip of the iceberg. As to how to apply them, it really is better-defined populations, up-front randomizations, proper stratifications, and use in a strategic fashion that will tell us where these will ultimately be placed.

I think it is also worth stating philosophically that, like every other assay in HIV care, it is here, it is being used, it is out in front of how we know how to use it, and that will be true of the next major assay in HIV as well.

So I don't think we will actually ever catch up. The technology is ahead of us. It has taught us a lot and it will continue to tell us a lot, so I don't think we will ever actually catch up with our clinical trials with the advance in technology, and maybe that is a good thing.

Other statements or additions to question 6?

We are twenty-one minutes ahead of schedule. I would ask Dr. Jolson or Dr. Murray what else you would like to--

DR. JOLSON: I just wanted to thank a moment to thank all the invited speakers today and, also, in particular to thank all of the people who helped put together the DAP presentation. We recognize it is an extraordinary amount of work. It was all done as part of a cooperative effort.

It has been enormously helpful for us to see these analyses. I think we have learned a lot from it. I really can't thank you enough for the time and effort that went into it.

I would also like to make just a personal appeal not to disband the RCG after this because I think, particularly as tomorrow's discussion evolves, it will become clear that the issues are going to get more and more complex as we start thinking of different scenarios to apply to resistance testing and drug development.

It would be impossible for the agency to think through these issues unilaterally. We really need comment and participation by industry and academics and others working in the field to help further the thinking. So that would just be a personal appeal that there be some sort of an ongoing effort to work collaboratively on that.

DR. HAMMER: I would echo that. I think that the presentations were fantastic this afternoon, and impressive, but, in no way, can they be absorbed by this committee or by

the agency within a couple of hours. I think that the data need to be mulled over and thought about and there are whole issues of analysis and interpretation that can help understand resistance testing and its application on the drug-development process.

So I would hope that there would be further communications between the agency and the RCG and other groups that have datasets that have datasets that have an impact on this analysis. It has really been impressive.

So I would, in closing, thank the participants, the speakers, the agency, our guests and consultants of the committee and, really, everyone that put a lot of effort into today's event. It is not over. Tomorrow is part 2 and we welcome you back here at 8:30 tomorrow morning.

[Whereupon, at 5:15 p.m., the proceedings were recessed, to be resumed at 8:30 a.m., Wednesday, November 3, 1999.1

- - -

C E R T I F I C A T E

I, **ALICE TOIGO,** the Official Court Reporter for Miller Reporting Company, Inc., hereby certify that I recorded the foregoing proceedings; that the proceedings have been reduced to typewriting by me, or under my direction and that the foregoing transcript is a correct and accurate record of the proceedings to the best of my knowledge, ability and belief.

Alice Toigo

ALICE TOIGO

B

C

D

E

G

H

I

J

K

L

M

N

P

Q

R