DEPARTMENT OF HEALTH AND HUMAN SERVICES

FOOD AND DRUG ADMINISTRATION

CENTER FOR DRUG EVALUATION AND RESEARCH

ANTIVIRAL DRUGS ADVISORY COMMITTEE

OPEN SESSION

Monday, July 14, 1997

8:30 a.m.

Armory Place
925 Wayne Avenue
Silver Spring, Maryland 20901

MILLER REPORTING COMPANY, INC.
507 C Street, N.E.
Washington, D.C. 20002
(202) 546-6666

P A R T I C I P A N T S

Scott M. Hammer, M.D., Chairman
Ermona McGoodwin, Acting Executive Secretary

COMMITTEE MEMBERS:

Wm. Christopher Mathews, M.D., M.S.P.H.
Judith Feinberg, M.D.
Pamela S. Diaz, M.D.
James J. Lipsky, M.D.
Wafaa El-Sadr, M.D., M.P.H.

CONSULTANTS:

Vernon Chinchilli, Ph.D. (Guest Statistician)
Joel I. Verter, Ph.D. (Guest Statistician)
John Modlin, M.D.
Fred T. Valentine, M.D.

GUESTS:

Mark Harrington (Patient Representative)
Brenda Lein (Patient Representative)
Donald J. Brambilla, Ph.D.
Winston Cavert, M.D.
Lynne M. Mofenson, M.D.
George Johnson, M.D.
Paul Palumbo, M.D.
Sharilyn Stanley, M.D.

FDA STAFF:

Lauren Iacono-Connors, Ph.D.
Jeffrey Murray, M.D.
Michael Elashoff, Ph.D.
Paul Flyer, Ph.D.
Donna Freaman, M.D.
David Feigal, M.D., M.P.H.

C O N T E N T S

Clinical Confirmation of HIV-RNA Changes

C O N T E N T S (Continued)

Company Presentations:

P R O C E E D I N G S

Call to Order

DR. HAMMER: Good morning. Welcome, everybody, to what portends to be an interesting two-day discussion by Antiviral Drugs Advisory Committee on using RNA as a primary endpoint in HIV trials.

I would like to start the meeting by having the people at the table introduce themselves, and I will begin on the left with Dr. Feigal.

DR. FEIGAL: Good morning. I am David Feigal, FDA.

DR. FREEMAN: Donna Freeman, Acting Division Director, Antiviral Drugs.

DR. FLYER: Paul Flyer, FDA.

DR. ELASHOFF: Michael Elashoff, FDA.

DR. MURRAY: Jeff Murray, FDA.

DR. IACONO-CONNORS: Lauren Iacono-Connors, FDA.

DR. VALENTINE: Fred Valentine, NYU, Bellevue Hospital.

DR. DIAZ: Pamela Diaz, Chicago Department of Public Health.

DR. MATHEWS: Chris Mathews, U.C., San Diego.

DR. HAMMER: Scott Hammer from the Beth Israel Deaconess Medical Center and Harvard Medical School in

Boston.

MS. McGOODWIN: Ermona McGoodwin, FDA.

DR. LIPSKY: Jim Lipsky, Mayo Clinic.

DR. EL-SADR: Wafaa El-Sadr, Harlem Hospital and Columbia University, New York.

DR. CHINCHILLI: Vernon Chinchilli, Penn State, Hershey Medical Center.

DR. VERTER: Joel Verter, George Washington University.

DR. MODLIN: John Modlin, Dartmouth Medical School.

MS. LEIN: Brenda Lein, Project Inform.

DR. HAMMER: Thank you. I would like to turn now to Ermona McGoodwin who will read the conflict of interest statement.

Conflict of Interest Statement

MS. MCGOODWIN: Thank you, Dr. Hammer. The following announcement addresses the issue of conflict of interest with regard to this meeting, and is made part of the record to preclude even the appearance of such at this meeting.

In accordance with 18 USC 208, general matters waivers have been granted to all Committee participants who have interests in companies or organizations which could be

affected by the Committee's discussion of plasma HIV-RNA measurement as an endpoint in clinical trials for drugs to treat HIV infection. A copy of these waiver statements may be obtained by sending a written request to the Agency's Freedom of Information Office, Room 12A-30, the Parklawn Building.

In the event that the discussions involve any other products of firms not already on the agenda for which an FDA participant has a financial interest, the participants are aware of the need to exclude themselves from such involvement and their exclusion will be noted for the record.

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

DR. HAMMER: Thank you. I would like to turn now to Dr. Feigal, who will introduce today's session.

Introductory Comments, David Feigal, M.D. M.P.H.

DR. FEIGAL: Good morning. In 1991 this Committee met to consider an application by Bristol-Myers Squibb to approve didanosine. At that time they had evidence which consisted of a control trial, showing an average of about a ten-cell increase in the CD4 count in patients with very low

CD4's.

It was wondered at that time whether or not that would be the basis to approve this drug. It had been about three and a half years since the approval of zidovudine. At the high doses that zidovudine was usually taken, the average duration that people could take zidovudine was less than a year. So there were many patients with HIV infection who really had no therapeutic options.

The question really was sort of what did ten cells mean, and was that an adequate basis to approve a drug product? The regulation for accelerated approval had not yet been written, although ideas of how to do such a regulation for conditional type of approval had been discussed by Commissioner David Kessler.

ddI went on to be approved based on those small changes but always with the understanding that the surrogate markers that the CD4 count and now later viral load with regards to the use of identifying promising drugs that would be followed up with clinical trials would show what the real benefit was.

One of the challenges for the Division right from the start was trying to decide what to do with the surrogate marker in the labels. What was it fair to tell people what really made sense, what really was the basis for clinical

information. Certainly, with the early drugs we didn't feel, and I don't think clinicians felt that individualized therapy based on small, transient changes in CD4 counts was a way of telling whether a drug was active or not.

As the therapies improved and as we moved into patient combination regimens, we could demonstrate that groups on average would have higher counts with new agents added to a new regimen. But it still really was not much basis for individualizing therapy, and although we described these CD4 results in the study section of the labeling and that information was available through the promotional literature and through the educational materials the companies had, it still was not clinically very satisfying.

Where we are now, however, is that we appear to have a measure of disease activity that is very sensitive in real time, is available commercially and has become a goal of therapy per se. Treatment panels have met to make recommendations about the optimal way to use the currently approved drugs, and have made recommendations about how to follow the load, and how to assess when someone has had good response and when a response is lost.

We began planning this meeting probably over a year ago as we began looking for trials to help us find a way to bring the data about viral load more systematically

into product labeling. The goal isn't simply to describe the studies that have led to an accelerated approval as to give the clinicians a sense of what the evidence is that the drug is active. The goal is really to describe the number of important features about the way that a drug performs and the way that individuals respond.

If you look over time in terms of how the therapies have been introduced for initial use when there was very little data available, we started with ddI with the rationale was, well, it appears to be an active drug and it should be used in patients with few alternatives, to a time when we realized that there was probably more promise with some of the new combinations than the old regimens where the trials had gone on long enough so that there was pretty uniform eventual failure, to now when there is a real need to be able to individualize therapy and assess response.

So part of the purpose of this meeting is to really take a look at how drugs affect viral load, and how should we describe those effects in the product labeling as a goal of therapy per se. This is not the same thing as a question of saying are we done with clinical endpoints because we still need to study these drugs in patients with clinically active disease. We need to understand the clinical toxicity and any adverse effects that offset the

clinical benefits. But we also feel we need to more systematically approach the way that we study the effects of the virus when the virus meets up with combinations of antiviral drugs.

When we began planning this meeting, we looked to commercial sponsors, we looked to the NIH groups of studies through the ACTG, CPCRA and other cooperative groups for studies that had information that could focus in on some of these questions; could tell us some information about how long should you wait to see response to a drug. How do you define an adequate suppression? How do you detect loss of response? And what is an appropriate evaluation or reasons for that loss of response?

This included studies that had both the measures of virology and immunology, and many of the studies also had the luxury of having measures of disease progression at the same time.

I think we have gotten past the simple question of is viral load validated that. I think we are looking at trying to define the metrics by which we think it will be useful to describe how these drugs work. We have looked forward to this meeting and, in particular, have appreciated the willingness of the study investigators to often break their studies apart and just show us one small focused part

of it to ask a question.

We met with the sponsors and we have asked them to try and follow a relatively uniform format in presenting the data so that it will be relatively easy to jump from study to study, but there will be times too when there are interesting other ways of analyzing these data.

The day will begin actually with some presentations by the FDA on some of our perspectives on these issues and at this point let me introduce the first speaker from the FDA, Dr. Lauren Iacono-Connors, who will talk a little bit about the properties of the viral load tests.

Overview of HIV-RNA Measurements, L. Iacono-Connors, Ph.D.

(Slide)

DR. IACONO-CONNORS: Good morning. I am Lauren Iacono-Connors, Division of Antiviral Drug Products. This morning we are going to hear three presentations on the subject of HIV-RNA quantitative assays. The purpose of this presentation is to generally review what these assays are in terms of their unique methodologies, what they actually measure and, most importantly, what we should keep in mind when reviewing data generated by these types of methods.

First I will present a very general overview of some of the methods currently used to estimate HIV-RNA in

clinical specimens. I will be followed by Dr. Don Brambilla, New England Research Institute, who will discuss data describing certain assay characteristics. In particular, he will focus on data which describes certain aspects of assay variability. Then Dr. Winston Cavert, from the University of Minnesota, will present data on tissue-specific HIV replication dynamics.

The goals of my presentation are to, one, review the general methodologies under development; two, define certain validation parameters which are essential when attempting to interpret HIV-RNA data sets; and, three, briefly discuss the target material for these types of assays, namely, HIV-RNA typically measured in plasma specimens.

Before I begin I would like to acknowledge and thank Dr. Indura Hewlett, from the Division of Transfusion Submitted Diseases from the Center of Biologics, for her contributions to this presentation.

(Slide)

This slide shows my outline. First I will discuss HIV nucleic acid quantification methods, then method validation parameters. I will follow this by a discussion briefly of HIV reservoirs, and then I will mention caveats to both the method and the target HIV specimen, and then I

will summarize.

(Slide)

Subsections of full-length HIV RNA, the virus genome, are the target of the five methods listed on this slide. All five methods require well-preserved HIV material in the clinical specimen. For most of the data which will be discussed today that specimen is plasma.

HIV particles in plasma have a wide range of potential concentration, upwards of seven logs. Several techniques have been developed which can systematically, directly through the RNA target or indirectly through the probe, amplify the HIV-RNA in a given specimen. As a result, the positive detection signal on the amplified specimen falls into a semi-quantitative range which allows for an estimate of the RNA copy number.

(Slide)

A well-documented direct nucleic acid amplification method is the polymerase chain reaction, commonly known as PCR. The majority of HIV-RNA clinical data which will be presented today and tomorrow were generated using this type of technology.

Due the inherent nature of this powerful molecular tool, the capacity for detection sensitivity can be most optimized. However, this technology also has the potential

for a broader range of values when attempting to quantify a single specimen. A greater degree of variability may be expected. Therefore, PCR-based assays may have may have greater sensitivity when attempting to detect small copy numbers but trade that feature for greater variability in all measurements.

Other methods which amplify the virus target are nucleic acid sequence-based amplification and strand-displacement amplification.

(Slide)

Techniques which are designed to indirectly amplify the HIV-RNA target material include branch DNA signal amplification, bDNA for short. Some of the data which will be presented here employed this technology. Since this method amplifies the probe which is directed at the target RNA, instead of amplifying the target RNA itself this technique appears to produce tighter results. Thus, the technique has lower variability characteristics, however, at the current state of development this type of technology appears to be less sensitive when measuring low copy count number material.

Another method which amplifies the probe for HIV-RNA is the ligase chain reaction. In order to interpret HIV-RNA copy numbers generated by a quantitative technique,

certain characteristics of the assay have to be defined. Essentially, the assay's capabilities and the limitations of those capabilities must be described for each specific assay. This is usually achieved through rigorous analysis of numerous analytical and clinical specimens. These assay characteristics are referred to as the validation parameters and, once determined for a given assay, they are unique and specific to that assay.

(Slide)

This slide lists the four basic areas of assay validation. Variability refers to how much wobble is detected when quantifying a single specimen. Two major contributors to assay variability are the assay itself due to its inherent design, its controls, standards, specimen handling and operator error, to name a few, and the biological variability associated with the specimen source.

Sensitivity refers to the lower limit of reasonable quantification of RNA copy number. Specificity simply requires that the assay detect the target material but not other biological components which may be present in plasma. Finally, the assay's linearity. This is described by graphing the expected and observed values in a set of control specimens. As a result, the upper and lower limits of the full range of linear detection are defined.

(Slide)

Now that we have generally identified methods for measuring HIV-RNA and some of their characteristics, the next question is where are we sampling for HIV. There are a number of well-documented tissue reservoirs for HIV. The virus can be found in both intracellular and cell-free compartments of the body. Cellular reservoirs include but are not limited to cells found in hematopoietic, central nervous system, skin and bowel tissues. Cellular subcomponents of these tissues, for instance lymph nodes, harbor actively replicating HIV, while other subcomponents harbor inactive or dormant HIV material.

Newly replicated and infectious HIV virion are shed from cells and either immediately infect an adjacent cell or move into a fluid component of the body, such as interstitial fluid, lymphatic fluid, plasma, cerebral spinal fluid and seminal fluid. Virus in the fluid component will, if unimpeded, cycle back into a competent cellular reservoir to infect new cells.

The HIV-RNA methods which I have already described are predominantly used to measure HIV-RNA in plasma. Therefore, the data to be presented here over the next couple of days describe estimated changes of detectable HIV-RNA by subcomopnent of one tissue reservoir of the body.

(Slide)

There are certain caveats to the methodology of quantitative HIV-RNA analyses to keep in mind when interpreting data. Each method is unique and every variation of a method should be considered unique. Therefore, data comparability between methods, unless rigorously studied, should be considered unclear.

Each method is different, meaning that those assay characteristics, referred to as validation parameters, are assay specific and vary between methods.

Finally, it is most important to remember that RNA copy number reported by an assay is a relative estimate of what may actually be present. In the absence of compelling comparability data, the estimate should also be considered assay specific.

(Slide)

There are also caveats associated with the selected HIV target source. When HIV is evaluated in plasma the copy number reported represents an indirect reflection of whole body HIV replication activity. It doesn't describe or allow predictions on any tissue compartment's contribution to the level of HIV-RNA present in plasma.

HIV nucleic acid material is the measurement target. However, this material is very labile. It's

potential random degradation may allow for inaccurate HIV-RNA estimates if the specimens are not properly managed.

Finally, all HIV particles present in plasma are assessed in these assays. Therefore, both infectious and defective HIV particles are being counted for RNA copy number.

(Slide)

My last slide just summarizes the brief presentation. There is a very wide range of HIV quantitative methodology being developed and used today to support analysis of viral burden in HIV-infected individuals. Because each method is unique, the characteristic validation parameters vary. In addition, due to differences in methods, the degree to which assay-specific data is comparable is not clear.

Although we have a vast amount of data on HIV-RNA copy numbers in plasma, it is important to keep in mind that we are assessing indirectly the replication activity of all competent tissues collectively. We have very little information on the degree of replication activity on individual tissues. The scientific and clinical community should continue to be suspicious of all competent tissue compartments and work towards the capability of HIV quantitative assessments of those tissues.

In closing, I want to point out that all these quantitative HIV methods are molecular-based state-of-the-art technology. We recognize the immeasurable amount of work and effort that has gone into the design and development of these methods. We also understand that these types of technologies are in a dynamic state of improvement and we look forward to each advancement.

This concludes my portion of the morning session, and now it is my pleasure to introduce Dr. Don Brambilla, from the New England Research Institute, who will present data which will further describe assay validation parameters. Thank you.

Assay Characteristics, Don Brambilla, Ph.D.

(Slide)

DR. BRAMBILLA: Good morning. Assay characteristics -- well, people write books about this in detail and have all kinds of characteristics in them. What I am going to talk about today are three basic issues that have to do with comparability of measurements between assay techniques and also reproducibility of measurements from different techniques.

First, the relationship of variation in RNA measurements to RNA concentration, in other words, the viral load.

Second, differences in estimates of HIV-RNA copies per milliliter of RNA concentration among laboratories in which the same kit is used and also among kits. So we are going to look at the comparability of the kits.

Then, lastly, longitudinal variation within patients, and what I want to focus on there is overall variation in series of measurements in the same patient, then the contribution of assay variation and biological variation to the total or the relative contribution.

Then the effect of assay variation, different levels of assay variation, on confidence limits that we can place around a measurement. So I will take each of these in turn.

(Slide)

I am going to talk about three kits, three assays today: The Chiron ES bDNA is the second generation assay that was released last summer. Here I have linear ranges listed. I am more confident about the listing of the lower end of these ranges than I am about the upper end of these ranges. So if anybody sees something they disagree with up here, let me know later please. The Organon Teknika NASBA Assay and then the Roche Amplicor HIV Monitor Assay. Now, I have listed in the linear range two different values for the lower end of the Organon Teknika Assay. That is because the

assay can be run with two different sets of internal calculators, one set being simply a 1:10 dilution of the other, which drops the lower limit from 4000 copies to 1000 copies.

There are two assays I am not going to talk about, the Roche Ultra-Sensitive Assay, which is still under development and characterization at Roche, and I am not going to talk about the NucliSens Assay from Organon Teknika, which was just released, and what that means is that part of my talk is already out of date.

(Slide)

A lot of the data that I am going to show you is derived from the Virology Quality Assurance Program, which is funded by the Division of AIDS, DAIDS in other words, to provide quality assurance for virologic assays in ACTG clinical trials and other NIH-funded studies, not just for RNA but for virologic assays in general.

The VQA laboratory is at Rush Presbyterian St. Luke's Medical Center in Chicago. The statistical center is in the New England Research Institute. That is me. There are several pieces of the program that are important to us today. First, the VQA laboratory provides external standards for RNA assays that are included in the assays that are run in ACTG clinical trials. Second, we conduct

routine proficiency testing of participating laboratories, and I will talk more about that. Third, we conduct studies to characterize and standardize virologic assays and, of course, that is what this talk is all about. The first two will play into what we are going to look at today.

(Slide)

The proficiency testing program, we need to describe that briefly since it is a basis for a lot of what I am going to say. The mandate in developing a proficiency testing program for RNA, the mandate that we were given by DAIDS was to develop a program which would show that a laboratory could maintain the precision needed to detect a five-fold difference between two measurements of RNA in the same assay batch.

We operationalize that by showing that the intra-assay standard deviation for $\log_{10}$ RNA should not be significantly greater than 0.15. The reason for putting it that way is because 0.15 gives you 90 percent power to detect a 5-fold difference. In other words, given the design of our panels, this means that the standard deviation really has to be less than something on the order of 0.19 to 0.21, depending on what panel we are looking at. The actual upper limit depends upon the number of specimens on the panel, number of dilutions.

We also have two other criteria. FDD errors must be less than the cutoff. Now, what is an FDD error? False difference detected. What this really means is that if you have a series of replicates at the same concentration on a coded panel that is sent to a laboratory, you compare the estimates that come back those specimens and if the estimates exceed a cutoff, which is the standard deviation of 0.15, then you have a false difference detected.

The other thing we do is TDU errors, a true difference between two specimens that differ by a factor of 5 in concentration. If the estimates are too close together, again, based on this 0.15 standard deviation, then we declare that there is a true difference that went undetected. In order to be certified under this program, a laboratory must maintain a standard deviation and have counts of FDD errors and TDU errors on the panel below the levels that are, again, specified panel by panel depending on the number of specimens on the panel.

Lastly, we have to have a panel that is free of false positives. The last two rounds of testing have included 6-8 negative specimens. So a certified laboratory meets all of these criteria.

Now, I want to say one other thing. This 0.15 is not out of line with the levels of precision that are

claimed by the manufacturers for the assays that we are looking at.

(Slide)

The proficiency testing program -- let's talk briefly about panel composition because, again, a lot of what I am going to say is based on what is in the panel, the results from the panel, so I would like you to understand what is on them.

First, we have plasma that has been spiked with concentrated HIV, a series of five-fold dilutions. A typical one will have about four different five-fold dilutions on it. So one might have 100,000, 20,000, 4,000 etc. There will typically be at least three replicates in each dilution. The two most recent rounds of tests have also included samples from HIV-infected patients. Again, what we have done, we have had is we have had one to three patients on the panel, so I think it is actually one or two on the panels I am going to talk about today, two to three dilutions of each. So we will take plasma from a patient and send it out neat and then send out a 1:5 dilution of it so we can compare it for TDU as well. Again, about three or more replicates of each one, depending on the panel. Then we are going to have panels that have three to eight samples of HIV-negative plasma, except the most recent panels have

been six to eight. So we have a mixture of prepared specimens and specimens from HIV-infected patients on the panels.

(Slide)

Here are two rounds of testing, round 6 and round 7, which is going to be the basis of a lot of this. The concentrations on round 6 in the HIV-spiked specimens, the 3 replicates of each, range from 2,200 copies/mL up to 1,375,000 copies/mL.

Now, this is above the linear range of the Roche Monitor assay so this was not sent to labs that use the Roche Monitor Assay. The 2,200 copy specimens were. The 2,200 is too low for the original Chiron assay, not for the ES bDNA assay but this panel was prepared before the ES bDNA assay was released. So the Chiron and Organon Teknika NASBA labs, or labs using those assays, received panels that ranged from 11,000 to 1,375,000 copies/mL.

Now, on panel 7 everybody got the same concentrations. You can see that they are slightly different from the previous panel but, again, we have a series of 5-fold dilutions.

(Slide)

The patient specimens -- we had 2 patients from panel 6. Sorry, this is actually incorrect. What we have

on panel 7 is 1 patient neat, 1:5 and 1:25. On panel 6 we had 1:5 and 1:25 dilutions of a specimen from that same patient, and then a 1:5 dilution of a specimen from patient 2. So, sorry about the confusion. It is actually these 3 here that are on panel 6 and then this one, this one and this one are on panel 7.

Here are the median concentrations estimated across all the laboratories participating in the testing, ranging from 14,000 copies/mL up to 320,000 copies/mL. It is the range over which we tested patient specimens in the proficiency program.

(Slide)

Recent results -- well, panel 6 was sent out on 2 rounds. This is because the Chiron ES bDNA assay was actually released between round 6A and 6B. So the data that I present to you that is based on comparisons involving the ES bDNA assay is going to involve 6B and 7A. The data that I present that focuses only on the Roche assay, which will come later when we talk about variability, is going to involve 3.

The bottom line here is that what you see in the numerators is the number of laboratories that met the certification criteria and the denominator is the number of participants in each round. Roughly two-thirds to three-

quarters of labs on a given round meet the certification criteria.

(Slide)

Let's go on to our first topic, the relationship of variation to RNA concentration. Most of the statistical models that we use to analyze data assume that the variation in our measurements is homogeneous. For example, it does not depend upon the mean. In other words, variation is constant. It does not depend upon viral titer. Variation of measurements of titer does not depend on viral load itself.

In assays variation is often correlated with the mean. That is true whether you are looking at the assay variability or the overall variability. The same thing tends to hold. So this violates an assumption of the model. So what we typically do is transform the data prior to data analysis, $\log_{10}$ transformation for example, to stabilize the variation.

(Slide)

Let's look at this. These are data from proficiency panel of the round 6 and 7 for the Roche Monitor assay. What you see here is each symbol on this plot is a standard deviation for 3 replicates. So this is assay variation. This is across all the laboratories that used

the Roche Monitor assay on round 6 and 7. The pluses are the spiked specimens. The circles, if you can see them, are the patient samples.

The only thing I want to point out here is that this is plotted on a log scale but actually the calculations of standard deviations are done without transforming the data. The thing I want to point out here is that the line connects the median standard deviations, simply just point out that the median standard deviation does go up with the mean.

(Slide)

If we transform to the log scale, you can see that what we get now is we transform the data to the log scale and calculate the standard deviations, and now we get a plot where the medians are pretty much constant across this range of concentrations.

(Slide)

If you do the same thing with the Organon Teknika NASBA assay you may actually get a standard deviation that rises somewhat at lower concentrations, but it is a lot flatter than you would see if you plotted standard deviations of untransformed data against the mean.

(Slide)

For the bDNA assay, it is pretty flat down to

about 10,000 copies. We have very limited observation in this range, below 10,000 copies so I am not going to say much about what goes on down there. This is under investigation.

Because of what I have shown you in the last three or four plots, we are going to talk about log transformed data when we do comparisons between kits, between laboratories and when we look at variability.

(Slide)

Next topic, differences in estimates of HIV-RNA concentration. What I am interested in here are two things that I want to show you: to what extent do estimates of HIV-RNA concentration vary among laboratories in which the same kit is used, and to what extent do estimates vary among kits. That is when you pool data across laboratories and prepare kits. So let's take these in turn.

(Slide)

A little bit about statistical methods and differences among laboratories. What I have done is fit linear regressions of log transformed estimates of RNA concentration to log transformed nominal concentrations. For the patient specimens we use those medians as the nominal concentrations. Then we compare the slopes and intercepts among laboratories within a kit. For the

differences among kits we use a random effects model, and what we are doing essentially is assuming that the slopes and intercepts are the regressions in the first part of the analysis are normally distributed around kit means, and we are simply comparing the kit means to get estimates of those.

(Slide)

Here is a comparison of laboratories that use the same kit. This is panel 6B and panel 7A. These are the fitted regressions. The panels that are labeled in yellow are panels in which either the slopes, the intercepts or both varied significantly among laboratories. For the ones labeled in white we didn't find any significant variation. This is on the spiked samples.

One of the interesting things I find about this is that there really is not a lot of difference in the visible spread of these regression lines between labs where you do get a significant difference among slopes and intercepts and where you do not. There are some fairly small differences in the error variance around regressions that seems to account for this. If you want to talk in terms of r-squares, the measure of error variance, as statisticians like to do, all of these models have r-square values of 0.95 or greater. These two are down in the 0.95 range and these

are higher. So the small difference in percent of explained variation in the models seems to account for whether we can detect differences among laboratories or not. All in all though, the regression lines are reasonably close together. You can see that the spread is in all no more than about half a log in any one of these plots, with the exception of Roche where you have a large number of laboratories.

(Slide)

As I said earlier, one part of the VQA program is to send out a set of external standards that are included in assays that are run by the clinical trials group by the ACTG. They are also included on the proficiency panels.

What we can do is take these external standards and adjust the data across laboratories to a common set of standards and see if that reduces any of the variability that we see. We eliminate one significant difference. These lines are actually closer together. We still have some differences here. Although we detect a significant difference in intercepts here, the lines are actually closer together than they were on the previous plots. We still have a significant difference but we have shrunk the difference.

Here we have a bigger spread, and that is largely because of problems with this top regression line, one

laboratory that had problems with the standards themselves. But in general these lines are closer together. That is the effect of adjusting to a common set of external standards.

(Slide)

Now, there are no significant differences among regressions in the patient specimens, which I find interesting. These lines, here, the spread of regression lines in these groupings, is tighter than what we found for the spiked specimens, which is part of the reason we don't find a significant difference.

The other thing is that the r-square values around these regressions tend to be a little bit lower, still up in the 0.9 or greater range. So we may be losing a bit of power to detect differences. Again, for the most part, the spread among regression lines is no more than half a log.

(Slide)

If you correct to a common set of external standards, all of a sudden the effect of that correction seems to be to generate some significant differences, unfortunately. But here, you see, you still have very little difference among the regression lines. Here two have been separated from the others and that is the reason for the significance.

Now I am going to talk about how the kits compare.

What I am really going to do at this point is to talk about if you take these regression lines and compare them, you will sort of find an average regression here, using a random effects model, for an average regression through these to see how they compare.

If you do this, here are some p values. This is for the spiked specimens both for the kit-based estimates and then the estimates that are adjusted to the external standards. What you find is that we have significant differences on panel 6B and on panel 7A among the kids. The intercepts of the regressions differ in both cases. The slopes differ in 7A, and the slopes are close to different on 6B. With the VQA-adjusted estimates we have basically eliminated the significant differences on 6B by adjusting to the external standards. We still have significant differences here on 7A.

(Slide)

However, here is part of the data. Those of you that are familiar with coral snakes will like this color scheme I think. But Roche is in yellow, Chiron is in red, Organon Teknika is in black. It is hard to see, there is a black line here and there is one buried between the two yellow lines. The shorter segments are from 6B, the longer ones from 7A.

The main thing here is that there is a small difference between the sort of average regression for Roche Monitor and the Organon Teknika NASBA assays up here. But the primary reason for the significant differences for both 6B and 7A between kits is because the bDNA assay is providing estimates that are lower than the estimates on the spiked samples and the estimates provided by other kits.

You can see that the spread seems to increase at lower concentrations. In other words, that is the difference in the slopes. For the most part, we are looking at a 0.3 to 0.5 log difference between these average regression lines, a 2- to 3-fold difference between the bDNA estimates and estimates from the other assays.

Now, if you do look at the data adjusted to the external standards, you can see that the regression lines are all on top of each other, which is interesting because it did still have a significant difference among regressions on 7A but it is very hard to see in this plot because the regression lines are so close together.

(Slide)

Here is the same thing for the patient specimens. Here we have differences in slopes on 6B, and that is about it. The only thing we see here is a difference in slopes. However, if you go to the VQA-adjusted estimates we pick up

a difference in intercepts on 6B as well. I will show you why.

(Slide)

There is something funny here. The longer segments are 7A again and you get the same pattern that the NASBA assay is producing estimates, a little bit higher than the Roche assay. Both of these are higher than the bDNA assay.

On 6B we have the slope of this segment, here, which is different from the slope of the Roche and NASBA assay. All the lines tend to collapse together again, even though we still do have a significant difference after adjustment to the VAQ standards, but the lines are much closer together. We have shrunk the difference between specimens.

(Slide)

Let's go on and talk about longitudinal variation within patients, the third topic, and then I will sum up and put everything together.

The three questions that I want to address are what is the overall level of variation of HIV-RNA within a patient? What I am thinking of here is if we track a patient over time during, say, clinical monitoring, quite apart from treatment effect, how much variability can we

expect to see in those numbers?

What are the relative contributions of assay variation and biological variation, overall variation? And how does assay variation affect the confidence limits on a measurement of RNA? These are the three topics that I named earlier.

(Slide)

There are several sources of variation that one can worry about when one talks about variability in RNA measurements. There is specimen handling prior to assay. So the blood is drawn. It has to get to the laboratory and then the assay takes place. But what happened in between? All sorts of things: storage, freeze-thaw, it sits on the shelf too long--lots of issues that can come up.

There is intra-assay variability. We have already talked a little bit about this. You would measure that by looking at differences among replicates in the same batch.

Inter-assay variability. If you took replicate subsamples from a sample of blood and put them in different assay batches what you would be measuring by inter-assay variability is really a combination. Measured in those terms, the difference between batches is really a combination of the variation within the batch and between the batches. So what I am talking about here is the added

component that comes up when you separate two aliquots into different batches.

Intra-kit, which is among kit lots, for example, which may be part of inter-assay variability, which is one way to view it. So if you use the same kit over time within a lab and keep changing kit lots, is that contributing to inter-assay variability?

Then inter-kit variability--we have already talked about this, differences between, say, the Monitor assay and the NASBA assay.

Then biological fluctuation, even if there is no systematic change in the patient and you had zero assay variability, you would not see constant numbers. Things would fluctuate. Those are all sources of variability.

What we are going to do now is to assume a patient is followed over time with the same kit in the same laboratory, with no change in treatment. So what we are going to do is focus on intra-assay variability, inter-assay variability and biological variability.

(Slide)

Measurements of variation is an important point because this tells us how we go about doing the estimates that we need to do to address these issues. Typically, when we talk about variability in an assay or measurements in a

patient, we talk about standard deviation of the measurements or some function of it, which might be a coefficient of variation which is the standard deviation divided by the mean. It might be the confidence limits you produce from a standard deviation. It might be the variance, simply the square of the standard deviation, or you could put the standard error, I suppose, up here as well. We typically discuss variation in terms of standard deviations or one of these functions but variances, not standard deviations, are additive.

(Slide)

Now, this is important. If you want to take apart the components of variation in an assay or in a series of measurements, say, you have to do that in a variance scale. What I am going to do is use this additive relationship here to take apart the overall variation, and then I am going to convert these values by taking the square root, two standard deviations, to present them to you since that is the scale of which we typically talk.

So our little model here is just overall variation in a series of measurements in a patient. It has three components that we are going to talk about. Biological variation may be inter-assay variation between batches and intra-assay variation.

Now, this model applies if you can get a series of clinical measurements in which each time the patient comes in and a sample is taken it is assayed. It doesn't apply if you are looking at samples that are assayed in a batch, typically done in a clinical trial where all the samples from a given patient are saved and assayed in a batch, because in that case the inter-assay component disappears and you end up with a somewhat simpler equation where the total is just the sum of the biological intra-assay components.

(Slide)

As I said, what we want to look at is the overall standard deviation, which is simply the square root of the sum of these three components, and we can define Sb, Se and Sa to be the standard deviations that correspond to what you saw previously from the same components.

(Slide)

Ideally what we do is measure components of variation in a nested study so that we can measure all three components at the same time. What we would do is take a sample from a patient, divide it into four aliquots, assay two in one batch; wait a while and assay the other two in another batch. So we would have both within batch and between batch measurements. We would repeat the same

process with a sample from the same patient, and that way we would be able to look at within batch, between batch and between sample of variability all, as I said, in a nested study.

I don't have data from a nested study to show you. I actually asked the ACTG if it would be possible to do that and when they figured out what it would cost they said no. So we will take a somewhat different approach to doing this instead of a nested design.

(Slide)

What we are going to do is to estimate overall variability from serial measurements in several ACTG trials and in a natural history study. Then we will look at estimating the intra-assay and intra-assay variability using the standards and panels from the Virology Assurance Program. Then we will estimate biological variation by subtraction. Biological variation is simply the square root of the overall variation minus the intra-assay variation. I put it this way because the studies in which I am going to estimate overall variation all involve batched assays so we don't have the intra-assay component in that case.

(Slide)

Here we have estimates of the overall standard deviation. Four clinical trials, ACTG 175 is actually run

in three laboratories, and then the Women and Infant Transmission Study, which is a natural history study of transmission from mother to offspring during pregnancy.

What I have done here for the WITS, I have three or four samples per woman for 57 women. They were collected during pregnancy. For ACTG 076 there are 55 women. This is a subset. They were in the placebo arm and had measurements at study entry, labor and delivery and six months postpartum. They received no AZT. This is a study that demonstrated that AZT reduces the risk of transmission in pregnancy. Then clinical trials 175 and 229 obtained paired baseline measurements. So we can take the differences between the two baseline measurements, the interval between them varying somewhat, two weeks or so on average I think. From those paired baseline measurements we can estimate variability. These are sample sizes and numbers of patients involved. There are 663 total.

We can use linear models, either random effects models, if we have three or more measurements, or basically nested analysis of variance with the patient ID as the predictor if we have two measurements, to estimate overall standard deviations. For the most part, they are in the 0.23 to 0.29 range. There is one exception here, which is quite small.

We can also get robust estimates. What I am getting at here is that the estimates from a model are sensitive to outliers. If you have a few points that are extreme values, they can blow up a standard deviation and give you what some would argue is an inflated estimate of a standard deviation that doesn't really represent what is going on in most of the data. If you take a robust estimate from paired measurements in all of these, in some cases you get pretty much the same thing because you don't have a lot of outliers. In other cases you get, as you do for Lab C in ACTG 175 and also 229, a pretty big drop because there are larger numbers of large differences between measurements in those studies.

But if you combine all the studies the overall estimate of 0.26 is what you get from the linear model; 0.18 is what you get from the robust approach. I am going to work with 0.26 because, although I know it is inflated by the outliers, it represents what is going on in all of the data. As I said, the robust estimate represents what is going on in most of the data but I would rather work with what is in all of the data.

One caveat about this is that the overall estimate is 0.26. It turns out that there is a positive correlation between the time interval between measurements and the

variability that one observes in these studies. The correlation is extremely weak. The correlation coefficient is about 0.05 but it is significant because of the large number of samples that we have. So bear that in mind. I might comment on that further a little later.

All the laboratories involved in this study used the Roche Monitor Assay. So when we go to the VQA standards and the VQA proficiency panels to look at assay variability we are going to look at the standards that are in some of these laboratories or we are going to look at the proficiency panel data from labs that use the Roche assay.

(Slide)

The standards--ACTG 076, the data that I showed you were derived from 48 assay batches, and there were three positive VQA standards in each batch. For 175 there were 26 and 37 batches in the three laboratories, and I think there were two standards included in each batch. So we estimate intra-assay standard deviations and inter-assay standard deviations from those standards. These are made up of plasma spiked with HIV.

The intra-assay standard deviations we have 0.09 for three of them; we have 0.15 for the fourth one. the inter-assay standard deviations are 0.08, ranging from 0.05 up to 0.10. If we put all the data together, all the

studies together, we end up with 0.12 and 0.08 as the estimates.

So the inter-assay component, the added component that one gets by running samples in the same patient in separate batches is a little bit smaller than the variability that one would see running samples repeatedly in the same batch. Of course, what you have to do if you want to get the assay standard deviation if you run samples in separate batches is square these, add them up and take the square root. I will show you a little later on what the effect of that is.

(Slide)

The Proficiency Testing Program, back to what we talked about earlier, let's look at our panels. Twenty-two laboratories participated in both round 6 and round 7 and used the Roche assay. So here we have estimates of intra-assay standard deviation from this. The median estimates from all 22 laboratories are on the order of 0.12 for the spiked samples and patient samples on round 7 to 0.16 for the same laboratories on panel 6. Here are the ranges for the intra-assay standard deviations. You can see that some of them are rather high. Those are laboratories that did not meet the certification criteria that we had set out which I talked about 20 minutes ago.

(Slide)

If you limit this to the laboratories--remember, there were 22 total--that meet the certification criterion just for having a standard deviation that is not significantly greater than 0.15, then you reduce this from 22 to 20 or 17, 19, 18. The standard deviations drop a little bit. Instead of 0.15, they go down to the 0.11 or 0.12 range for the median standard deviation, and you can see that the upper limit has dropped down.

So those are intra-assay standard deviations. There are consistent with what we saw from the standards that were included in the batches for clinical trials for ACTG 076 and 175. That is because the laboratories that participated in ACTG 076 and 175 met the certification criteria.

(Slide)

Biological variation. We have an estimate of intra-assay variability which we can take from the VQA data, and we have an estimate of overall variability. Here is our equation again estimating the biological standard deviation. As I said, if we take our larger value, 0.26, for the total then you can just square it to get the variance. If you take 0.12 ball park estimate of the intra-assay standard deviation, which is a good estimate of the median for the

laboratories that are meeting certification criteria, then what we end up with is that the biological standard deviation ends up being estimated at about 0.23.

The number itself doesn't mean much. What is important here is that if you take the biological variance and divide by the total variance, what you find is that 80 percent of the variability that you would see in a sequence of measurements from a patient is biological rather than assay variability. This is true if you batch the specimens from a patient and assay them altogether. That is the first thing to keep in mind. The second is that it is a bit of a ball park estimate because there is some concern about what the actual assay variability is in a laboratory. You have to take that into account as well. I want to show you that and then I will wind up.

(Slide)

Let's talk about the effect of assay variability on the overall standard deviation. Here is 0.12 where we started. We said we had batched assay so there is no inter-assay component. We are assuming a standard deviation of 0.23 and that gets us back to our 026.

If we simply allow the intra-assay component to go up to the upper limit of what is considered acceptable on the VQA program, then the overall standard deviation drops

from 0.26 to 0.30. If we let it go up a little bit higher, it jumps up to 0.34. If we start throwing in the inter-assay component and now we are looking at what happens if a patient is followed over time and samples are assayed as they are collected, then here we are back to 0.12-plus, the lower estimate of inter-assay variability that we had and we get a very small increase in going from here to here, 0.27.

So if the laboratory is performing well with regard to intra-assay standard deviation and has an inter-assay component that is at the lower of the two levels that we are comparing here, you get a very small increase in overall standard deviation. In other words, you don't lose much in this setting by going to real-time testing.

If the standard deviation creeps up to 0.15 for the inter-assay component, then we are up to 0.3 compared with our 0.26, and now let's just boost the intra-assay component to 0.2 and now we are up to 0.32 and 0.34.

Finally down here, at the bottom, we have a laboratory that is not meeting the performance criteria for intra-assay standard deviation set up by the VQA and also by the manufacturers, at least according to the package inserts. Then we have the two inter-assay standard deviations and we are up in the 0.35-0.37 range. So we have a substantial increase in overall standard deviation here.

(Slide)

What is the impact of this on our actual numbers? Here is the same chart and I have added 95 percent confidence limits. Here, at 0.26 it is just 95 percent confidence limits plus/minus half a log. You can expect the confidence limits that you could put around a measurement on a patient in that setting wold be plus/minus half a log. Then you can just count through here and see how the confidence limits would increase. Actually, the one I wanted to point out is right here, again, a small increase in standard deviation from going to real-time testing translates into a small increase in the confidence limits, from 0.51 to 0.53 to 0.59. However, if you got problems with the intra-assay standard deviation you are up to around plus/minus 7/10 of a log in terms of confidence limits.

We did all of this in terms of log scale because of the variability issues that I talked about, but we can now take plus/minus half a log to 7/10 of a log and translate back to real numbers, real RNA measurements, and say, okay, what is the confidence limit that we would place around a measurement of 10,000 copies/mL, just to pick a number that we can work with?

Here you have from 3,100 to 32,000. So if you had a measurement of 10,000 copies/mL in a laboratory that was

performing with this intra-assay standard deviation you might say that you got about a 10-fold range in which you would park that number. It lies somewhere between around 3,000 and 30,000.

If you go to real-time testing the limits, again, don't increase very much. They are still in the range of 3,000 to 34,000. But when you get down here where you are having intra-assay problems and you have a larger inter-assay standard deviation, now you are down to 1,900 to 54,000 for the confidence limits on a single measurement.

This translates directly, although I don't have it on a slide here, into talking about the confidence limits on the difference in two measurements, in other words, in measurement of change within patients. As you increase the assay variability, the confidence limits for measurements of change also increase. I am sorry I don't have a slide to demonstrate that.

(Slide)

Summary. I covered three topics. First, I want to say that we use log transformation to stabilize variances when we analyze data. As I showed you on the last slide, we can always take that back to the untransformed numbers, the absolute RNA copies/mL when we want to look and see what the numbers actually mean.

What we found from this in kit comparisons is that estimates from the Chiron bDNA were lower than those from the Roche Monitor OT NASBA assays. Stay tuned when we start looking at the new version of the OT assay and we will do this all over again in about six months and see how things compare. The differences between the bDNA assay and the others actually vary with concentration. They are actually larger at lower concentrations.

Biological variation--I will put it this way, 80 percent of total variation in a patient, excluding treatment effects and trends, assuming that the patient is not on average changing. I said that it is up to 80 percent because that is assuming some minimal level of assay variation. The percentage of variation represented by the patient will go down as assay variability goes up. The total variability will go up as assay variability goes up and the percentage represented by the patient will go down.

What the confidence limits are saying, what this really translates into is that unconfirmed changes of 0.5 to 0.7 logs may have very little clinical meaning because those are just within the realm of possibility for fluctuations, including assay variability.

Lastly, I think it is important to monitor laboratory performance over the long term, perhaps using

external standards, however one chooses to do it, in order to ensure that that component is kept as small as possible so that what we are looking at is primarily the biological component. Thank you.

DR. HAMMER: Thank you, Don. Before we move on I might give the Committee a chance to ask questions of Dr. Iacono-Connors and Dr. Don Brambilla. This provides a lot of important data for us to discuss issues of application of these assays over the next couple of days. Are there any questions? Mark?

MR. HARRINGTON: It looks like there is a lot of variation even among the very best labs. My question is, we are talking about a standard of care that uses these measurements on a routine basis, how much greater variation do you think there is out there in the labs that were not certified but, yet, from which clinical data come from trials, and then how much even more variation is there in commercial labs that are actually giving people measurements that are used to make treatment decisions?

DR. BRAMBILLA: Well, I don't have any hard information to use to answer that question. You know, it is entirely possible that a clinical lab maintains its own very strict standards and, therefore, would meet these performance criteria. It is also possible that one doesn't

know unless one has some method of performance. I think that is the real problem. The answer to your question is that one needs to have a measure of performance and for a lot of laboratories I don't know what it would be.

MR. HARRINGTON: But what happens in an ACTG trial when the lab fails the certification but has been gathering data for the trial?

DR. BRAMBILLA: If I remember correctly, a laboratory has to be--typically in an ACTG trial the samples are saved in batches in the freezer and assayed at the end of the trial or at some point in the trial and the laboratory has to have maintained current certification in order to analyze the samples.

DR. HAMMER: Other questions?

DR. LIPSKY: When you look at variation are there particular trends, for instance, over replicates? Do the means go down? Do they stay the same? Do you get any hint of causes when you are doing an assay of why there is variation?

DR. BRAMBILLA: Oh, you mean from the proficiency testing program?

DR. LIPSKY: Yes.

DR. BRAMBILLA: Occasionally you do. A lot of times what you will see is that there are two different

kinds of variability that creep in here. Occasionally you will see a panel where everything looks great except one or two specimens in which something happened. Occasionally we can explain it easily. I have seen NASBA assays where it looked like the specimen never made it into the assay. The signal is basically zero. I have seen other assays where it looked like two specimens were switched. So occasionally you get a panel that looks fine except for a couple of outliers. Sometimes you do not. Sometimes you see panels, and I think this is the more common explanation, where at multiple nominal concentrations there is an increased spread over and above what you like to see. Why that happens I do not know. That is the point at which the VQA laboratory starts working with the personnel at the ACTG lab to try and figure out the problem.

DR. FEINBERG: Don, I think this is probably directed at the FDA and not you, but are there any comparable analogous data for biologic variance, etc. for the CD4 data that have been the basis for approval up to this time? I was not on the Committee when those things were laid out real carefully. I am just curious to know that.

DR. FEIGAL: We would probably have to get back to you with specifics, and you may have some comments too, but

I think it is of the same order of magnitude.

DR. BRAMBILLA: I have never looked at serial samples before from this point of view.

DR. EL-SADR: Does the biological variability differ by the actual RNA level in the patients? Did you look at people with higher values versus those with lower values?

DR. BRAMBILLA: Yes, I started to do that because I know that the transformation data I showed you was based on assay variability. No, I don't see a lot. If you go to the log scale, that tends to suppress a lot of the difference. If you don't go to the log scale, then, yes, absolutely the patients with a higher value show greater fluctuations. But on the log scale it is largely suppressed.

DR. VALENTINE: Just to respond to his question, the ACTG had a certification program for CD4s and the data was all analyzed by SDAC, and they did similar types of analyses.

DR. MATHEWS: Could you comment on what is the impact of these parameters if you were to pick clinical trial endpoints based on achieving, say, a certain threshold value, say, undetectable or whatever it is, a certain drop in baseline at a fixed point in time, whenever that might

occur? So time achieving a certain value versus trying to estimate a slope for an individual patient? In other words, it seems to me from what you have said that if there could be a 10-fold difference in a single measurement and 95 percent confidence interval, it makes it much more difficult to pick a single threshold value as an endpoint versus trying to estimate changes by slopes over time, which might have less variability.

DR. BRAMBILLA: if you are talking about picking a single threshold value for making endpoint decisions for individual patients, I think the solution is to not believe anything until you confirm it with a second sample, something along those lines. One can define all sorts of regiments, whether it is two samples or three samples or a cutoff and then increase the confidence that the patient actually is below that cutoff by doing repeated measurements. I think that is the main thing you get here from those rather broad confidence limits.

The other problem that creeps in when you start talking about clinical trials, and it is one of the things I left out of this deliberately, is variation in treatment effect among patients or variance in treatment response. I leave that to my colleagues at SDAC to work with. That is more their job.

DR. HAMMER: Don, do you want to say something about the assay performance at the lower limits of quantification, where there is so much interest individually in patient management and certainly for clinical trial endpoints?

DR. BRAMBILLA: There is some tendency, and you saw this in the plot for the NASBA assay, for the standard deviations in log transformed data, at least on the NASBA assay, to increase at the low end. I don't know if this is true for the new NucliSens assay. I haven't seen any data. I am speculating in talking about the other assays because I really don't have as much data.

In fact, that is something that the VQA is working on, on characterizing. We have, for example, a set of panels in the field right now that are designed to characterize the Roche Monitor assay from 1,000 copies on down to 35,000 copies, looking at reproducibility and rates of values below detection limits. The data are not due in to us for analysis until next week.

DR. HAMMER: One quick last question, when you have your patient estimated numbers, those are taken from the mean of certified laboratories? What is the gold standard for identifying the patient numbers in your panels?

DR. BRAMBILLA: Oh, we just took the median across

all laboratories. That is a good point, but I think the thing to keep in mind is that the medians reproduce the 5-fold dilution series pretty well.

DR. HAMMER: Thank you.

DR. MATHEWS: I want to follow up on your comment to my question. You said that in individual patient management it is important to replicate. But in clinical trials you have to decide on an individual patient basis when they have achieved the endpoints. So you have the same problem as you do in individual patient management.

DR. BRAMBILLA: Absolutely, yes.

DR. MATHEWS: How many replicates do you need, based on data you have presented, to state that a person has confidently reached an endpoint, whatever that value may be?

DR. BRAMBILLA: Let's see, how do I answer that? I have to give you sort of a nebulous answer. It is hard to nail that down. The point is that if you have a value--for example, suppose you have a value that is so low that the upper 95 percent confidence limit on that value is less than the endpoint that you are talking about, then you might begin to think that one is enough. If you have two that are somewhat higher but the upper 95 percent confidence limit on the pair is less than that endpoint, then you might be happy with that. It all depends on how you want to define it.

That is the problem, the more replicates you have that are near or below that endpoint, the tighter the confidence limits that one can put around it.

DR. ELASHOFF: We are going to address that in more detail tomorrow so there will be a lot more data. I think maybe comments for Dr. Brambilla could may be put off.

DR. HAMMER: Thank you.

DR. VALENTINE: Just to pursue the point that Mark was making, clearly, we would presume that in clinical trials you would have quality assured laboratories, and so forth. Does the Agency know if there is some way of regulating or quality assuring commercial laboratories which are going to be predominantly used for patient management, which I suppose is what Mark was getting at I think? Is there a plan in place or how are those laboratories regulated? I know the College of Pathologists have sent out samples but I don't know with what stringency those are analyzed.

DR. FEIGAL: Laboratories are regulated by the accrediting bodies of those laboratories, which is not an FDA function. We would review the performance of the test and then it would be the College of American Pathologists review or sometimes the Joint Commission on Hospital Accreditation. Sometimes state licensing works. In fact,

it may vary state by state. Most commercial laboratories and hospital laboratories will have quality assurance programs and should be able to make it available how they do that.

DR. HAMMER: Thank you. Thanks, Don. I think we should move on. The next speaker is Dr. Winston Cavert, from the University of Minnesota, who will speak about comparative tissue compartment activities.

Comparative Tissue Compartment Activity, W. Cavert, M.D.

DR. CAVERT: I want to apologize to the Committee for using overheads. I found out on Thursday that all three of the locations within my neck of the woods that turn copy into slides were out of commission or on vacation.

(Slide)

We have clearly come a long way since Salk first coined the term viral load with regard to HIV in 1987, as this proceeding attests. But I think one of the implicit assumptions when we talk about plasma viral load is that the plasma reflects what is going on in the total body.

As I understand my charge from the Advisory Committee today, it is to see is that, in fact, the case? Does plasma viral load in some way reflect the viral burden of the entire body, in particular the lymphoid tissue, without therapy and during therapy?

(Slide)

None of the studies that I am going to cite today have the sort of statistical power that we just saw. All of the relevant studies really have probably less than 16 patients in each. So I think we are still a long ways from having the kind of information that we would like about this question but there are some things that can be said.

I am going to review what was first called the dichotomy between the lymphoid tissue compartment and plasma or blood compartment, and a little bit of the viral kinetics models that has arisen out of that. I am going to talk about the correlation between plasma RNA and lymphoid tissue RNA in individuals that are untreated or what I am going to flippantly perhaps call lightly treated subjects, subjects with a single nucleoside perhaps; and then the comparative response of these two compartments during treatment; then some suspicion of what possible exceptions to the correlation between the loads between the two compartments might be; and then, finally, a brief comment about DNA load as well as RNA load, which is going to be the substance of this talk.

(Slide)

In 1991 Pantaleo and colleagues first identified a significantly higher burden of HIV in the lymph nodes or

adenoids as compared of the blood.

(Slide)

Since then the work from several laboratories and several sites I think has relatively incontrovertibly given us the impression that the lymphoid tissue contains the vast majority of the HIV viral burden within the body. Generally about three orders of magnitude perhaps more HIV is found in the lymphoid tissues than in the plasma.

In the lymphoid tissues we have identified major compartments or two major pools of virus. The first is follicular dendritic cells with germinal centers of lymph tissue. There, intact virions are complexed to antibody stored on the foot processes of follicular dendritic cells. Secondly, productive mononuclear cells within the lymphoid tissue. Most of these are CD4 cells but some as well are macrophages and other mononuclear cells perhaps.

(Slide)

When we are talking about lymphoid tissues we are primarily referring to secondary lymphoid tissues of lymph nodes, adenoids and tonsils, the spleen and the gut-associated lymphoid tissue within the intestines, and less so perhaps the thymus and the bone marrow.

(Slide)

To emblematically, if you will, put the two

different compartments within lymphoid tissue into some perspective, this is a computerized 3-D image of a series of sections from a single germinal center of a small lymphoid biopsy. What you see here, in the yellow is the follicular dendritic cell holding intact virions, and the red colored dots are individually productive mononuclear cells, producing virions with full-length HIV RNA.

(Slide)

There are at least as many, and actually probably a number of additional methodological issues that go along with looking at lymphoid tissue, even more so than there is in sampling blood.

For starters, there have been several different lymphoid tissue sites used for monitoring viral load: lymph nodes, tonsil and adenoids, otherwise sometimes referred to as mucosa-associated lymphoid tissue; gut-associated lymphoid tissue, most commonly from the colon; and spleen, taken from individuals who, for example, have thrombocytopenia and need, for clinical reasons, their spleens extracted.

There is also a number of different sampling techniques that have been used in these studies that we will be looking at. Autopsy specimens have been used. Lymph nodes have been biopsied with fine needle. Larger biopsies

have been done visually, either directly, for example of the tonsils, or direct larger needle biopsies of lymph nodes and also via endoscopy for the gut-associated lymphoid tissue, and finally tissue excision as in, for example, the spleen and whole lymph node excisions.

(Slide)

There are also different assays for dealing with this tissue, essentially variations on the same things that you have already heard about. There is a tissue bDNA assay, a number of versions of the RT PCR, including a tissue adaptation of the Amplicor assay, and at least two variations of the NASBA assay that I am aware of. There are also single-cell technology techniques, all of them essentially based on in situ hybridization, and some of those using more formal methods of quantitation including computerized image analysis or simply visual microscopic quantification.

(Slide)

All of this raises a number of potentially troublesome issues in analyzing HIV-RNA lymphoid tissue, and I have highlighted just some of the more prominent ones that I think are, at least some of them, unique to lymphoid tissue.

The first is the question of whether HIV is

uniformly distributed within lymphoid tissue. This is both between different types of lymphoid tissue and also whether it is uniformly or relatively uniformly distributed within particular lymphoid tissues, for example within one lymph node. When one obtains a sample of lymphoid tissue there is an important question and that is does the sample actually have lymphoid tissue in it. In a number of the biopsies that have been looked at, for example from gut or other sites, other types of tissue are included in the biopsy. So it is important to know what percentage of the biopsy is actually lymphoid tissue and what percentage is stroma, fat, epithelium etc.

The question what the standard of reference by which we quantitate the viral load in the lymphoid tissue is I think still the major question. Several of the ways that it has been expressed so far is to express samples as per gram of lymphoid tissue or per the number of human cells, presumably T-lymphocyte but not necessarily always, cells that are counted, or perhaps the total amount of RNA that is extracted. Obviously, comparing between studies that use these different methods is going to be difficult.

There is also the question of total HIV-1 RNA in the lymphoid tissue versus looking at the amount of virus that is associated with two major cellular compartments,

that is, the FDCs versus the mononuclear cells.

Then there is the issue of assay sensitivity and range. The sensitivity calculations and considerations are analogous to those for plasma but quite unique or varied for lymphoid tissue. Finally, there is obviously the issue of quantitative reliability or validation.

(Slide)

At the risk of being, unfortunately, somewhat superficial due to constraints of time, I have lumped a number of the studies together in general categories as to essentially how read them.

I thought I would start off with a quick reference to six pioneer studies, very important studies but earlier studies that were cross-sectional studies, if you will, in individuals that were either untreated or treated perhaps with a single nucleoside or a single additional nucleoside. One of the earliest of these by Dr. Pantaleo in 1993--all of these studies that are on this page, I should mention, to my reading or to the reading of the authors do not show correlations, or at least do not show particularly strong correlations between the viral load in the lymphoid tissue and viral load in the blood. All of these studies are small, for the most part a dozen or fewer patients.

In the black, underneath each assay, I have put

some of the issues that I think make the data somewhat difficult to interpret. In the first study there are issues of assay sensitivity, and I think the assays are rather more sensitive at this point.

The second study, by Sei in 1994, used tissues that were from autopsy and some of them had been 24 hours or more before they were harvested and, therefore, I think it would be expected that the lymphoid tissue RNA would be rather degraded and so correlation wasn't seen in that study.

There are two papers published from DATRI 003, the first in a group of individuals, between 250 and 500, I believe it was, CD4 cells. There were 8 patients in this 1995 Cohen study that were--actually, I should say that all of these patients were on AZT at the baseline point, and it is the baseline that I am focusing on from a cross-section of perspective. In general, as I look at this paper, I don't think that they found a correlation between the lymphoid tissue and the plasma. However, there are significant differences in how they expressed the denominators of those two compartments so I think it is a little difficult to read. I am not sure, there may be some correlation in these studies. I don't have access to the primary data from that but looking at their papers I am not

sure.

In a group of pediatric patients who were failing double nucleoside therapy in 1996, Mueller did not find a correlation between the two compartments.

Finally, Pesca Meylan published a paper last year using fine needle biopsy in 8 patients. Both the issue of fine needle biopsy, which is a sampling problem that needs to be looked at more carefully, as well as differences in how the denominators in the two compartments were expressed. They did not find a correlation either.

(Slide)

There are three papers published in the literature that, in my estimation do show a correlation between the relative amount of virus in the lymphoid compartment and the relative amount of virus in the plasma. All of these were in small numbers of patients. Mary Ann Harris and her colleagues in Vancouver in conjunction with Dr. Pete Daley and his colleagues working on tissue bDNA at Chiron provided an abstract at Vancouver last summer, a study comparing plasma bDNA, I think using the second generation assay, with the tissue bDNA assay. As you can see in these 14 patients or so, they found a reasonably decent correlation with a p value of 0.02.

So I guess what I am saying when I talk about a

correlation is when the plasma viral load higher, is the lymphoid viral load also higher? Do the two relatively reflect each other? And I think this study shows, at least for the small number of patients, that they do. The reasons why I think this is a better study than the others are two-fold. First, although this is a lymph node biopsy, they had histological touch confirmation from the fact that they were dealing with lymph node tissue, and almost entirely with lymph node tissue.

Secondly, I think the tissue bDNA assay, of all the sort of homogenation assays available, is probably the one that has gone the furthest so far in terms of validation in looking at the biologic variability and the intra-assay variability.

(Slide)

Two other studies for which I will not show the log-log plots in the interest of time but, essentially the plots look very similar. Faust published in Otolaryngology Surgery in 1996 a comparison of plasma DNA versus tonsil biopsy, again with histological confirmation that it was, in fact, lymphoid tissue and, again, using bDNA assay performed by Pete Daley at Chiron. Plasma bDNA assays in this case were performed by the ACTG lab at the University of Minnesota.

Then finally a study out of our lab, senior authored by Dr. Ashley Haase, published late last year in Science, correlating plasma bDNA second generation assay with tonsil using a new technique of quantitative computer analysis of in situ hibernization. Again, both of these studies show reasonable correlation with fairly good p values in the Faust study in 10 patients and the Haase study in 8 patients with 13 biopsies.

(Slide)

To look at the issue of what happens to the lymphoid tissue load during treatment, there are several studies of lightly treated patients where when we see a drop in the plasma viral load the lymphoid tissue may or may not drop. One of these is the previously cited Cohen paper from DATRI 003 of 6 individuals who had already been on AZT previously were treated in addition with ddI. As you can see, the log changes in the plasma RNA, in the closed black symbols, are relatively small but, nevertheless, in the open symbols for each of these 6 patients there was a somewhat correlating drop in 5 of them in the lymphoid tissue load. One of these patients was clearly an outlier and I think it is an important point to recognize that in statements that the two compartments seem to correlate in their loads we have seen, in addition to this study, a couple of other

circumstances, patients that are definitely outliers, where they have in particular low plasma virus load and quite high lymphoid burdens.

This is a study published last year by Dr. LeFalay, from Toulon in Marseilles, in southern France. There are 4 patients, previously naive, treated with AZT, ddI and 3TC in a pilot study of that combination. The squares are lymphoid biopsies and the circles are plasma. As you can see, each of these patients, although the times of biopsy varied as well as the times of plasma sampling varied in these patients, we see some corresponding drop in both compartments. The drop, at least in this early study on the lymphoid tissue looks like it was smaller than it was in the plasma, but I think that probably has to do with the sampling method.

(Slide)

That same group published a second study in ME just a couple of months ago in 10 patients treated with the same 3 nucleosides plus saquinavir, previously antiretroviral naive. In the top panel are the 10 patients' plasma RNAs. As you can see, they have roughly a mean 3-log drop between them. In the bottom panel are the plots of their lymphoid tissue viral load changes over 8 weeks, but a 2.5-log drop in the lymphoid tissue.

(Slide)

Two final studies on the comparison between plasma and lymphoid burden in treated patients, this one from Switzerland, again Dr. Meylan's group. There were 9 patients again with fine needle. This is interesting. This paper was just published, in February I believe it was. This is an interesting comparison that hasn't been done so far in any of the other studies that I am aware of, that is, the log change in the RNA level on the Y axis versus the log change with treatment in the plasma HIV-RNA on the X axis. As you can see, there is a correlation to some extent between these.

(Slide)

A study that we performed in conjunction with colleagues from University of Amsterdam, using triple therapy of AZT, 3TC and ritonavir, 34 adults were split between 2 arms, half of them received all the drugs from the start and the other half received just ritonavir from the start and added the nucleosides 3 weeks later. Onto the study a tonsil biopsy substudy was appended. Ten of these patients ultimately were included into the tonsil biopsy substudy.

(Slide)

This is from Dr. Don Notterman, one of the

principal investigators in Amsterdam. This is what happened with the plasma viral load with this triple therapy over the course of 24 weeks. As you can see, this curve of plasma viral load looks similar to that we have seen in other potent combination therapy trials. Over the 24 weeks they had about a 3-log drop in their plasma in both arms.

(Slide)

In the lymphoid tissue over those same 24 weeks, the analysis being done in our lab by in situ hibernization and computer image quantification, we saw a 3.4-log drop in the FDC pool and greater than a 2.3-log drop in the mononuclear cell, actively producing cell pool, in the frequency of those cells. If one were to homogenize the lymphoid tissue, the FDCs would be the predominating component in that. You can essentially say that the total lymphoid tissue viral burden drop was greater than 3.4 logs.

(Slide)

Unfortunately, I don't have direct comparison curves of this, but in the curves for individual patients there is an obvious similarity. Part of the difficulty in comparing curves too closely is because the plasma was sampled much more frequently than the lymphoid tissue. The lymphoid tissue was sampled at baseline, then at day 2, at day 22 and at 24 weeks.

Here is what the curve looks like from the first 3 time points. The FDCs are the top line, in the red and the mononuclear cells, are the blue line underneath. As you can see, they both fell off rapidly within the first 2 days and then a slower phase, if you will, although it is a little hard to talk about a phase when you have only two time points and you draw a line between them but a slower phase of viral decline from the lymphoid tissues between day 2 and day 22.

One of the reasons I wanted to show this slide was because to us this expresses a model of what is going on in the lymphoid tissue and in relation to the plasma. It had been proposed in the past that because FDCs bind antigen antibody complexes, they would be a long-term storehouse. In fact, what we found is that the drop in the FDC lymphoid pool paralleled the drop in the productive CD4 cell pool. Therefore, we think that there is a quite fluid equilibrium between the FDC "storehouse" and the mononuclear cells in the lymphoid tissue. By comparison with the plasma curve, there also seems to be fluid equilibrium between both of these compartments in the plasma. So we think that what is driving the plasma, viremia and the FDC loading is the production of virus by the mononuclear cells, and when you shut that off with potent therapy, or nearly shut that off,

all three pools drop in a roughly parallel fashion.

(Slide)

Just a couple of brief words about the relationship between lymphoid tissue proviral DNA and plasma RNA. This is an important issue because if we can drop lymphoid tissue RNA down to levels that are getting close to detectable, then the issue starts to emerge what about cells that are latently infected that may have simply proviral DNA within them?

I think earlier studies suggested that there may be a correspondence between the amount of proviral DNA in lymphoid tissues and the plasma RNA viral load. This is part of a study published by Sei in 1994. This, again, was a study that had autopsy tissues. I did not show a relation probably because of that between the RNA components but the lymphoid tissue DNA, in the open triangles up here, and the plasma RNA, in the closed circles on the bottom panel, show a rough correlation or, actually, a pretty decent correlation. So that was one of the first studies suggesting that there may be a correlation between lymphoid DNA and plasma RNA.

(Slide)

Here is another one that seemed to confirm that. It was 8 patients, in a letter to Nature Medicine in the

middle of last year by Diazoni. Again, I think you can see the correlation pretty much by just looking at the number of DNA infected lymphoid cells in this column versus the RNA copy numbers.

(Slide)

However, a more recent paper, published just about a month and a half ago by a group at Johns Hopkins, Ocono was the main author, using really I think quite sophisticated PCR assay techniques, suggested that there may, in fact, not be a correlation between lymphoid tissue DNA and what is going on in the plasma RNA. So I think that is really an open question at this point.

(Slide)

One other point to mention is that when I am talking about lymphoid tissue we are ignoring a lot of other potential sites. Lymphoid tissue is obviously where most of the viral burden in the body is located but there may be small amounts of virus in other locations. One study showed 16 percent of individuals with HIV infection have active replication going on in their bone marrow. CNS obviously is a potentially worrisome site of sequestration, as is genital tissue, and a number of others have been suggested as a possibility, including lung, myocardium etc.

One slide that somehow I lost out of the deck here

was to point out likely situations where the plasma may or may not reflect what is going on in the lymphoid tissues. Several of these possible situations include very early after seroconversion. What we thinks happens early on, after someone seroconverts is that mononuclear cells producing virus in the FDCs are slowly loaded over a period of days to weeks, possibly months. So in that early phase we would expect the lymphoid tissue burden to be relatively much lower compared to the plasma viral burden.

Second, that may also be the case in very late AIDS. It has been reported that lymph node architecture is quite destroyed in end-stage AIDS patients and, therefore, the follicular dendritic cell network that holds onto virus is going to be incapable of holding a significant viral burden theoretically. I think this is still open to question. End-stage AIDS may be another problem in trying to correlate these two.

There are also the issues of penetration of drug into lymphoid tissue and other tissues. So when I am showing these treatment trials, I think one cannot necessarily generalize across all potential antiretrovirals. It undoubtedly depends on drug penetration and also perhaps on individual metabolic features in different compartments.

Joe Wong and Doug Richman, at the January

retrovirus meeting, reported a couple of cases of individuals who had been on potent triple therapy for a period of time, then had a lapse in their therapy of several days and then had their lymphoid tissues and their plasma sampled two or three weeks after that. They found a disjunction between the plasma load and the lymphoid load, and the plasma load was uncharacteristically low compared to the lymphoid burden. They proposed at that point a short lapse in therapy perhaps resets the clock or, if you will, reloads the follicular dendritic cells.

Finally, there are reports in the literature of variable resistance patterns in body compartments, and I think that needs to be worried about as well.

(Slide)

In conclusion, as I said, the lymphoid tissue HIV-1 RNA is much greater than the plasma viral burden if you use comparable denominators.

The better cross-sectional series of untreated or lightly treated patients show in general a correlation between these two compartments. That is to say, when the lymphoid tissue burden is high the plasma burden is higher and vice versa.

There are occasional outliers and, as I mentioned, there are exceptional situations that are likely to exist.

The plasma RNA, as we have seen in the few studies that have been published to date in small numbers of individuals, in suboptimal treatment doesn't always seem to correlate, or at least in one of the studies there was some apparent correlation.

(Slide)

In individuals that are given potent combination therapy there seems to be parallel declines in the viral burden in the lymphoid tissue and the plasma. This is probably because there seems to be equilibrium between the lymphoid viral pool, both the FDCs and the mononuclear cells, and the plasma.

Finally, the lymphoid tissue HIV burden probably correlates with peripheral mononuclear cells in the blood, but the plasma RNA and the lymphoid tissue proviral DNA burden may not correlate. Then there is the question of other tissues besides lymphoid tissue. Thank you very much.

DR. HAMMER: Thank you. We are running a little behind but before we move on to the pediatric data, are there any pressing questions for Dr. Cavert from the Committee? If not, thank you very much.

I think Dr. Cvetkovich, from the FDA, is going to introduce the pediatric section of the program.

Introductory Comments on Pediatric Data

DR. CVETKOVICH: Except for the next three studies in this presentation at this meeting, we have been drawing exclusively from studies conducted in adults. However, the picture of the actual history and the response to treatment derived from viral load studies in HIV-infected infants and children is now beginning to emerge. While it may be premature to draw firm conclusions from these data, only now may we begin to evaluate what may be important similarities and differences between the adult experience and that found in the pediatric population.

We are very pleased to include in our meeting the next three speakers who will address the topic of viral load in HIV-infected infants and children. Dr. Lynne Mofenson, from the National Institute of Child Health and Development, will present current natural history data. Then Dr. Palumbo, from the University of Medicine and Dentistry of New Jersey, will present virology from ACTG 152. Finally, Dr. George Johnson, from Medical University of South Carolina, will present very recent virology results from ACTG 300. Dr. Mofenson?

Natural History, Lynne M. Mofenson, M.D.

DR. MOFENSON: I have been asked to summarize the results of the pediatric literature in 20 minutes. So what you are going to have is me speaking very rapidly.

(Slide)

Just to review what we see in infection in adults, with primary infection in adults there is an initial burst in viral burden with a peak of about a million, but after several weeks to months it declines to under 10,000 and most levels remain low or undetectable for many years, and RNA changes as well as CD4 changes together have been shown to be independently predictive of prognosis.

(Slide)

What is different about children that could affect the natural course of HIV viremia? First of all, HIV infection in most children perinatally infected is primary infection. Most children are negative in culture and virologic tests at birth and become positive only after about a week of age. So perinatal infection is really primary infection.

Uninfected newborns are immature both in cellular and humoral immunity. So we now have primary infection occurring in immunologically naive infants. Additionally, in normal newborns there is an increase in activation markers on CD4 and CD8 cells because they are rapidly expanding and differentiating, making them potentially more infectable by HIV. Infants also have a higher absolute lymphocyte number and higher CD4 levels than seen in adults

which, again, gives you more of a population to infect, and the CD4 cells are primarily naive.

So based on this, one might anticipate that there might be some differences in the natural history of RNA in perinatally infected infants.

(Slide)

I am going to review the published studies. The first study is by Paul, sitting over there, and this study was published in 1995 and looked at serial plasma samples during first year of life from 14 infected infants who were involved in a prospective natural history study. This was based on whole blood in heparin. RT PCR was used to measure. The mean RNA copy number in these infants was over 500,000. The decrease over the first year of life was only 2- to 10-fold less than 1 log, very much in contrast to what you see with primary infection in adults. Additionally, despite the high RNA levels, most of these infants had age appropriate CD4 counts and there was both rapid as well as slow progression.

(Slide)

This slide basically shows the data. As you see, levels remain high for a year and well near a million copies.

(Slide)

Moving along to the recently published data from the Women and Infants Transmission Study, this is a prospective natural history study of infected pregnant women and their infants from 6 different sites in the states that are listed on the slide. Blood is collected serially in infants, and the population that was reported in the recent New England Journal was perinatally infected singleton infants, born between 1990 and 1993 who had more than 12 months of follow up.

(Slide)

There were 673 plasma specimens on 106 children, giving over 6 samples per child. Again, this was heparinized blood specimens with RT PCR, with RNA extraction via the Boom technique. The lower limit of detection for this assay was 400.

(Slide)

This just goes to show you the plasma HIV-1 RNA copy number by age. What you see is that there are very high copy numbers with the first few weeks of life. Although these copy numbers declined, they declined only very slowly, and most children have levels over 100,000 during not only the first year of life but also only slowly declined to under 100,000 by three years of life. So early peak, slow decline.

(Slide)

This looks at rapid versus slow progression of disease in these children. Basically, what you can see is that children who had rapid progression, which meant that they developed AIDS or died before 12 months of age, had higher HIV-RNA copy number at almost data points with exception of the birth specimens.

(Slide)

This looks at a Kaplan-Meier curve based on whether or not the infant's median RNA level during the first 2 months of life was above or below the median. The median was nearly 300,000. You can see in adults RNA copy numbers significantly correlated with risk of disease of progression. These are kids who were above the median, a significant number progressed, versus those who were below the median. The difference one sees in adults is that the levels associated with progression are significantly different. We are not talking about levels of 10,000, we are talking about levels over 300,000.

(Slide)

This slide shows you the first 4 months of life in these children. Again, RNA copy numbers in progressors, which is here, were higher than in non-progressors. However, the important thing to look at is that there is

significant overlap for individual children between progression and non-rapid progression.

(Slide)

So basically in the WITS study RNA levels increased rapidly until about 1 to 2 months of age and then slowly declined through 2 years of age, and reached a median of about 35,000 copies at age 24 months, in distinct contrast to what we see in adults. This pattern was observed in both rapid progressors and non-rapid progressors, and was also observed in kids who had cultures positive at birth versus those who had cultures positive later.

(Slide)

Infants with rapid progression had higher peak RNA levels during the first 2 months of life than those with non-rapid progression. Infants with rapid progression also had higher mean geometric RNA levels during the first year of life. However, because of overlap between rapid and non-rapid progressors, there was no single threshold level that could be identified that would be very predictive of progression. However, it should be noted that infants who had RNA levels under 70,000 to 80,000 at 1, 2 and 4 months of age had rapid progression.

(Slide)

Moving right along to the next study, Penn McIntosh and colleagues from Boston Children's Hospital recently published two articles. They looked at children with perinatal infection, followed in their site for clinical care between 1986 and 1993. The children were seen every 3 to 6 months, and they took stored sera from excess samples that were processed rapidly and stored at minus 70 degrees, and again used RT PCR.

(Slide)

They looked at 48 children who had a mean age of 29 months or a little over 2 years at first visit, and looked at them over the subsequent 2 years. There were about 6 samples per patient. Most of these children had received therapy at some point during the study.

(Slide)

This is very similar to what one saw in the Shearer paper. Basically, this is looking at a cross-sectional look. As you can see, RNA copy numbers were very high, over 100,000, in the first year of life and showed a consistent climb over time, persisting for long periods of time, many years, and this was seen regardless of treatment or non-treatment. The untreated kids are in the little triangles and the treated kids are in the little circles.

(Slide)

This slide shows you basically the same thing except looking at slope. This is a negative slope, here, a positive slope here and zero slope. So you can see that there is negative slope for about the first 4 to 5 years of life and then a slow increase in RNA, more consistent with what one sees in adults.

(Slide)

So to summarize the McIntosh study, there was a slow consistent decrease in RNA from about 1 to 5 years of age that was about 1 log. RNA levels were highest in the youngest children. The slope was about -0.1 to -0.2 log per year. The set point was not reached until infants were nearly 5 years old, and this decline was seen regardless of degree of immunodeficiency or receipt of antiretroviral therapy.

(Slide)

This just shows you their next study, which was published just a month or two ago, that correlated RNA copy number with weight for age Z score and CD4 for age Z score, and this is just to show you that there was a significant correlation as RNA copy number increased, and these are different quartiles for the children, so did the weight for age Z score. So children had an increased level of weight loss as RNA copy number went up.

(Slide)

The same type of thing is seen for CD4, although it is not as well correlated with weight.

(Slide)

I would like to move on to the study I know the best, which is our recent publication from the Journal of Infectious Diseases, where we looked at stored specimens from the NICHD IVIG clinic file. This was a clinical trial of IVIG versus albumin placebo that occurred between 1988 and 1991. We had 376 children enrolled in this trial, most of them perinatally infected. We obtained vital status updates on these children through September, 1996. So we had approximately 5-plus years of follow up on the children.

(Slide)

Blood was central storage at entry and every 3 months during the study. It was stored at minus 20 to minus 70 and then shipped to a central repository. Specimens from children who had more than one sample were retrieved and tested, and here HIV-RNA was measured by the NASBA assay by one ACTG certified lab.

(Slide)

This shows you what the patients were like. We had 254 patients that were in this study that were a mean age of 3.4 years, mean entry CD4 count a little over 1,000,

mean entry CD4 percent 20 percent. This is pretty much normal for children that are 3 years.

(Slide)

This slide shows the distribution at baseline. As you can see, there was a wide range of HIV-RNA copy numbers in these children. The median RNA copy number in these 3-year old kids was 105,000. The geometric mean was about the same. Only 8 percent of these children had undetectable levels; only 16 percent were under 10,000; a full 50 percent were over 100,000 and 2 percent were over 500,000, very different than one would see in a relatively asymptomatic adult population.

(Slide)

This shows you again the longitudinal geometric mean by age. You should be pretty used to this picture, very high levels early in life, near a million; slow decline that continues for a considerable period of time. The most rapid decline is within the first 2 years of life.

(Slide)

This shows you the same graph by age but in children who died versus those who did not die. As you can see, those children who died had significantly higher RNA levels at almost all ages than those children who did not die during follow up, although both of them showed some

decline.

(Slide)

This shows you the gradient of mortality with increasing RNA. Again, this is 5-year mortality. With children with RNA copy number under 10,000 the mortality rate at 5 years was 23 percent. For those between 10,000 and 100,000 the mortality rate was similar, 24 percent. Between 100,000 and a million increased to 40 percent, and over a million it was over 70 percent.

(Slide)

This slide breaks it down into smaller categories. You can see that as RNA increases there is an increase in the risk of mortality. But really the sharp increase occurred at about 100,000. Just to note, for those children who were undetectable, 24 percent died by 5 years. Between 4,000 and 50,000 the mortality rate was 28 percent, and 50,000 to 100,000 was 15 percent. Then when you jump up to over 100,000 it goes up to 40 percent. So there didn't appear to be much differentiation between those children who had undetectable levels and those children who were about 100,000.

(Slide)

This is a Kaplan-Meier plot broken into the same categories of under 10,000, 10,000 to 100,000, 100,000 to a

million and over a million. What you can see is that there is a difference in terms of the Kaplan-Meier plot. However, the under 10,000 children and the children between 10,000 and 100,000 are not statistically significantly different from each other. This one, 100,000 to a million is different than these two and this one is, as you can see, clearly different.

(Slide)

This is an ROC curve. I just want to point out the difference in terms of sensitivity in false-positive rate depending on what threshold one picks to look at. For example, if one picks a threshold of 10,000 to define risk for mortality, you will be very sensitive and most children who are above 10,000 will have a death. However, you also have a significant false-positive rate, nearly 80 percent, and that means that 80 percent of children with levels over this much do not die. If you compare to the highest level, a million, the false-positive rate is much less but the sensitivity is also less. At about 100,000 sensitivity was about 80, false-positive rate was about 40.

(Slide)

This is to show you that again there was a difference by age. This breaks kids down to under 2 years and over 2 years. As you can see, in the under 2 children

one does not see a significant increase in 5-year risk of mortality until you hit a million copies, whereas in children who are over 2 you begin to see this rise at about 100,000. So there is probably a point in between birth and 6 years or so where the RNA predictive value changes from very high levels to lower levels. I think Paul will show this data as well.

(Slide)

This is just to show you CD4 was also significant. It correlated particularly in the lower CD4 range. For each 5 percent drop you saw a rapid increase in mortality.

(Slide)

This shows you the correlation between HIV-1 RNA and CD4 percent. There was a borderline statistically significant relationship between the two. The r value was minus 0.4. This is very similar to the data published by Mellors.

(Slide)

This is a multivariate analysis looking at baseline RNA and baseline CD4 percent and their correlation with long-term risk of mortality. As you can see, they were both independently predictive with about a 2.5-fold increased risk of mortality for every log increase in RNA and about a 1- to 3-fold increase in risk of mortality for

every 5 percent decline in CD4 percent.

(Slide)

This just looks at change over time. Basically the data is the same for every log increase in RNA during the course of the study. Mortality risk increased about 2- to 8-fold and the CD4 percent increase was about the same as before.

(Slide)

This looks at the predictive value by baseline RNA quartile. This is RNA quartile 1, the lowest quartile, and this is the highest quartile. And now we have looked at mortality risk when you break it down by CD4 count. My main point here is that no matter what RNA level, CD4 count was still predictive of increased disease progression. It did increase mortality.

(Slide)

However, when you break it down and you look at it by CD4 percent, less than 15, 15 to 25 and over 25, in the children who were most immune suppressed the mortality risk appeared similar regardless of whether you had nearly undetectable RNA or very high copy number, and it was only in the children who had a better immune status where RNA levels showed the step-wise increase in terms of predicting disease progression.

(Slide)

This broke down HIV-RNA into different combinations. The most favorable combination was RNA under 100,000, CD4 over 15 percent and the 5-year mortality risk in these children was 15 percent. As you can see, there is a step-wise increase, and when you get to bad parameters for both variables, over 100,000, less than 15 percent, the mortality rate at 5 years was a whopping 81 percent.

(Slide)

This shows you the Kaplan-Meier for those categories. The main thing I want to point out here is that when you use RNA and CD4 together you get better differentiation in terms of prognosis than when one uses RNA alone.

(Slide)

In conclusion, what can we say about the natural history? Well, high peak levels, over a million, are reached by 1 to 2 months of age in perinatally infected infants. In contrast to the several log fall over about 6 months that one sees in adults, the RNA decline in perinatally infected infants is less than 1 log during the first year of life. RNA also continues to decline over the first several years of life regardless of therapy, and not appearing to reach a zero set-point slope until about 5

years of age, and when it does, it appears to be higher than that reported in adults.

(Slide)

As in adults, higher levels of RNA were associated with increased risk of mortality and also disease progression. You will see that in the next two presentations. However, similar to the CD4 threshold for PCP prophylaxis, RNA levels indicative of increased risk of mortality or disease progression appeared to be higher in infected children than has been reported in infected adults.

Just a comment, however, infected children progress more rapidly than adults. One of the reasons for their more rapid progression may be due to the higher RNA levels that they have.

Finally, RNA and CD4 when used together were better predictors of progression risk in terms of mortality than either used alone. Thank you.

Response to Treatment: ACTG 152, Paul Palumbo, M.D.

DR. PALUMBO: I want to thank the Committee for giving me the opportunity to present ACTG 152 and for recognizing that pediatrics actually is an important consideration at this meeting, with some very unique characteristics which Lynne very nicely outlined for us.

(Slide)

This is a relatively old study, between 1991 and 1993. The entry criteria were quite broad and liberal, such that the majority of infants in our clinics were eligible for this particular protocol. The organizers of this protocol recognized that young infants were going to be considerably different than older children.

There were approximately 830 infants who were enrolled into this protocol over a 2-year period. They were stratified by age, realizing that young infants were going to have significantly different progression than older infants. The stratification was below 30 months and above 30 months. This was a double-blind, randomized, placebo-controlled trial of monotherapy with 2 nucleosides, ZDV and ddI, compared with combination therapy.

(Slide)

It is important to consider the endpoints in this trial. They were entirely clinical and were composed of either time to death or time to first HIV disease progression. Far and away the majority of clinical endpoints were either weight-growth failure or CNS abnormalities, which is very typical of pediatrics. In some respects we are fortunate to have these endpoints that are very definitive and well defined in this population.

(Slide)

This is a Kaplan-Meier plot of the entire cohort for data through November, 1994. At this point a DSMB review recognized that the ZDV arm, which is in green, was performing less well than the other two arms, the ddI monotherapy and the combination arm, in yellow and black. The ZDV arm was stopped at that point but the other two arms were continued in a blinded fashion.

(Slide)

This is the data for the less than 30 months age group, which shows a more dramatic separation of the ZDV arm compared to the other two arms. There was no clinical difference between the ddI monotherapy arm and the combination arm.

(Slide)

That is a finding that was very similar to ACTG 175 that Scott Hammer led for the adults, which is a very similar trial to this pediatric trial.

(Slide)

I mentioned that there were about 830 children enrolled in ACTG 152, and 579 are included in this virologic analysis. There were actually 566 who had both baseline RNA and CD4 results. So about 3/4 of the children in the study are represented in this virologic analysis that I will present today, and they were very representative of the

entire cohort. There were no significant differences between this subset that I will show you and the overall cohort. There were 311 under 30 months of age. We had about 1,500 specimens or 5 specimens per child in that group, and 268 in the children greater than 30 months of age.

The bottom two boxes here represent the baseline plasma RNA results. These are in untreated infants and children. So this was basically a natural history study, following up on what Lynne Mofenson just presented to us. In the top box I have eliminated the 19 individuals who had negative or undetectable RNA at baseline. Those individuals are represented in the lower box, if anyone wants to look at that particular one.

Again you see very high median baseline RNA results for infants under 30 months of age, 620,000, a mean of 5.7 log. For comparison purposes, in ACTG 175, the adult study, was similarly designed. The mean baseline RNA results were 4.2 log, so about 1.5 log lower than our infants who were less than 30 months of age. We see about a 10-fold decrease in baseline RNA results for children over 30 months of age, where the median was 61,000, the mean was 4.8 log.

(Slide)

Graphically depicted through 18 years of pediatric untreated experience at baseline, you see the data presented here. Very high levels, over a million, in the early phases of the study, bottoming out or plateau-ing at around or a little below 100,000 in older children.

(Slide)

Depicted a little differently here by age groups, I do this for two purposes. One, to show the number of infants in each age group, and we had 164 infants who were less than or equal to 12 months of age in this study. That is a very high number to be studying. Again, you can see very high baseline RNA levels at entry in the youngest infants, greater 10^6, and staying at or above 10^5 generally throughout the age groups that were entered into the study.

(Slide)

This graph depicts the mean changes from baseline in log RNA copies/mL at defined time points. In ACTG 152 the first plasma specimen we had available to study after starting therapy was at 24 weeks, or about half a year. So we were missing a lot of potentially informative data before 24 weeks. Nonetheless, it is quite interesting. This is the entire data set of all age groups combined and we do see differences between the 3 arms, significant differences, at all these time points between 24 weeks and 96 weeks.

The mean changes from baseline are very similar to adults, using similar antiretrovirals. Even though in pediatrics we are starting at a much higher level, we are seeing quantitatively similar decreases in pediatric patients.

(Slide)

This is a very similar plot for the subgroup who are less than 30 months of age. We see the most dramatic decreases in this population. At 24 weeks we are seeing 0.3 log reduction in the ZDV monotherapy, about 0.6 logs in the ddI monotherapy arm, and 0.9 log in the combination arm. These are modest increases but they are associated with significant clinical benefit.

(Slide)

In the group over 30 months of age we are not seeing quite as much difference between the treatment arms. We do see significant differences at 48 and 72 weeks when the poorest performing ZDV arm is compared with the other two arms. Otherwise we see much more of a general trend between the three arms.

(Slide)

There was more extensive data for the ddI and combination arms. Those were continued in a blinded fashion for a longer period of time. We have data out to 144 weeks.

As you recall, we saw no clinical differences in the performance of these two arms but we did see significant differences between week 24 and week 72 in the reduction in RNA, with the combination arm performing better than the ddI monotherapy arm. This also was seen in ACTG 175. No clinical differences between these two arms in adults, but differences in RNA. So this is the group less than 30 months of age.

(Slide)

Now, some Kaplan-Meier plots for tracking progression-free survival based on the baseline RNA. We divided the baseline RNA into 4 quartiles, 4 fairly even quartiles, undetectable to 150,000 being the lowest quartile, very different than what we see in any of the adult studies obviously; 150,000 to half million in the second, half million to 1.7 million in the third quartile, and greater than 1.7 million. And there is a very nice correlation between baseline RNA and ultimate course.

The other point here is that this is a very linear increase in risk with increase in RNA, with no evidence of any threshold above or below which there are dramatic differences in disease-free progression.

(Slide)

These are the children over 30 months of age. The

same can be said for relative linear effect. You can see that the quartiles are 10-fold less now than they were in the infants under 30 months of age. The quartiles are more like adult quartiles now as we get into the older age group of children. There were no endpoints seen in the 66 children who were in the lowest quartile. So there may be some argument made for some threshold effect in the older children, but in general a linear effect of RNA with risk.

(Slide)

Another way of evaluated the issue of threshold in the infants under 30 months of age was to divide the RNAs into octiles if there was any information lost in the quartile information. But when one plotted baseline RNA with the hazards ratio we saw a very linear effect across the broad range of values that were present at baseline in infants under 30 months of age.

(Slide)

We also looked at hazards ratios or the Kaplan-Meier plot looking at the week 24 RNA, the absolute value obtained after 24 weeks of therapy. Again we divided it into quartiles. This is children over 30 months of age. What we start to see here is a threshold effect. That is, if you fell into the lowest quartiles, that is, less than or equal to 50,000, the risk was extremely small, 10 percent

less relatively speaking, for 24 month progression-free survival after that 24-week time point. If one got above 50,000 the risk increased relatively dramatically for that particular population.

I don't have a similar plot for the infants under 30 months of age but, again, the quartiles are higher and there is a threshold effect at about 100,000, with relatively low risk for progression or death in infants who fell under 100,000.

(Slide)

One of t he questions that we wanted to address is, is pathogenesis or risk in older pediatric populations with lower viral load going to be different than younger children, that is infants, with higher viral load?

What I show you here are hazards ratios, and these were taken from the Kaplan-Meier plot. So this is the percent failure at 2 years plotted against RNA. The infants under 30 months of age are in the solid line and the quartiles are presented by the children. The children over 30 months of age are in the dotted line and the quartiles are represented by the dots. What you can see is overlap here. Obviously there is 10-fold RNA in the younger infants than in the older children, but at a given RNA level the risk is essentially the same. So there is overlap.

The other point here is that in both cohorts we do see a linear effect of risk versus RNA, again without evidence of a threshold.

(Slide)

If one looks at the week 24 plasma levels and then follows these children for another 2 years after that week 24 time point and calculates the percent failure at that 2-year time point, one again sees similar risk in the 2 age cohorts at similar RNA levels.

One also starts to see what I was alluding to before, that there may be threshold effects and that they are fairly similar for the 2 age groups. There may be somewhat higher thresholds for the younger infants, again around 100,000, and they may be lower for the older children, possibly around 50,000 or potentially a little bit lower.

(Slide)

As Lynne alluded to, RNA is not the only variable we should be considering and combining RNA with CD4 has potential for adding more power to our ability to calculate where a given individual is. All of the ACTG 152 data I am presenting today are, obviously, group data. I am not dealing with individuals. But a number of Cox proportional hazards models were generated with this data base for the

infants under 30 months of age, and this is what I am presenting here.

When RNA and age were included in a Cox proportional hazards model, RNA alone was independently predictive of outcome and of risk. The age fell out. Baseline in model 2, both CD4 and age were independently predictive of outcome. In model 3, when one compared both baseline RNA, CD4 and age. Both baseline RNA and CD4 were independently predictive of outcome in a fairly significant fashion.

The final model I will describe to you is model 6 where all of the variables, baseline RNA, RNA attained at 24, baseline CD4 and CD4 obtained at 24, are included in a complicated model. But, by and large, all 4 variables are independently predictive of where the child or where the group is going to go. We see risk reductions in the range of 40 to 50 percent for each log decrease in RNA at baseline or each log RNA decrease from baseline to week 24. These are fairly similar risk reductions to what has been described for adults in adult trials.

(Slide)

Just to show what some of our challenges are in the future, this is basically a graphic demonstration of where we are or post-ACTG 152. The top line here represents

baseline RNA data for infants less than 30 months of age. With nucleoside therapy as used in ACTG 152, I show a maximum of about a one log reduction in RNA, a modest reduction but clinically significant.

As you can see, we still have 3 to 4 logs to go to reach undetectable RNA. Whether this is going to be achievable in young pediatric populations is yet to be proven. I think it is very reasonable to expect in children over 2 or 3 years of age that they will probably perform similarly to adults on very aggressive therapy. However, it is still unsure how many of these younger infants we are going to get to undetectable RNA with aggressive therapy, and those trials are just beginning or ongoing.

(Slide)

I will just take one minute to acknowledge some of the people who played important roles in these analyses, Claire Raskino and Tim Ramacietti, at SDAC at Harvard School of Health, did most of the analysis and management of the data base. Jane England, Carol Baker and Steve Spector were the principals on protocol ACTG 152, and the four laboratories that performed the assay, and the assay that we used was the NASBA assay, are here. These are all ACTG certified laboratories. Thank you.

DR. HAMMER: Thank you very much. I think what we

will do before move to the third pediatric talk, is to take a 15-minute break here since we are running a few minutes behind. So we will return at 11:18 and start promptly.

[Brief recess]

DR. HAMMER: We are going to complete the pediatric data section. The next speaker is Dr. George Johnson of the Medical University of South Carolina, who is going to present the virologic results from ACTG 300.

HIV-RNA Results: ACTG 300, George Johnson, M.D.

DR. JOHNSON: I would like to thank the Committee for inviting me to speak. Actually, Dr. McKinney should be here as the protocol chair and driving force behind this protocol, but he is having a lot more fun refining his blues guitar skills in West Virginia right now.

(Slide)

This slide is to reflect the number of people that are involved in these protocols and give everyone credit.

(Slide)

ACTG 300 was substantially similar in design to ACTG 152 but we are at a much more preliminary stage of analysis of the data so our virology is going to be extremely limited and I am just going to present the data that we have, with a summary of the trial.

The objective was to compare AZT-3TC combination

therapy with the better of ddI or ddI plus AZT therapy with respect to progression, with a decision being made as to which of the two ddI-containing arms should be determined based on the clinical results of 152. Actually, the AZT-ddI dual therapy arm was discontinued in May of 1996 on the basis of the clinical comparison of the 152 results. That was prior to our virology data being completely available.

The patients were very similar. They were symptomatic HIV-infected children, and they had to have less than 8 weeks of prior antiretroviral therapy. This was also slightly different from 152 initially in that this allowed for perinatal AZT therapy, and they were between 7 weeks and 15 years of age.

(Slide)

The endpoints were very similar but there were some minor differences, primarily related to neurologic endpoints. But the principal endpoints were development of a new category C disease or death, inadequate weight growth velocity of deterioration neurologically.

(Slide)

As of April 4, 1997 when the data set was frozen for DSMB analysis, there were 615 children who had been enrolled in the study; 596 were able to be evaluated, and the data was on the comparison of ZDV-3TC or ddI with 471

evaluable children. The data was current within 2 months for 98 percent of the patients on treatment, and the median follow-up at that point was 9.4 months.

(Slide)

This is just to show the similarity of the population in the 2 treatment arms that were analyzed for the DSMB and to give you baseline data relative to CD4 counts, which were very comparable, and median RNA log titers, which were also comparable.

(Slide)

The primary analysis was on the time to clinical progression or death, and there was a statistically significant difference between the arms, the top solid arm being the ZDV-3TC, the bottom being ddI.

(Slide)

This just gives you a breakdown of the number of endpoints. There was a total of 53 endpoints which contributed to that analysis, 38 in the ddI arm, 15 in the ZDV-3TC arm. As you can see, there was also a large difference in death as a first endpoint, but a large number of the endpoints, similar to 152, were CNS deterioration and weight-growth failure.

(Slide)

This is the breakdown and stratum looking at

children under age 3 relative to time to clinical progression or death. Again, this is more impressive than the total set and, actually, 44 of the clinical progression or death endpoints were in this subgroup or stratum of the population. That was 83 percent of the endpoints.

(Slide)

This is looking at the same data relative to those in the over 3 age stratum, and there was not a significant difference. Looking just at survival for the whole population, you can see there was a significant difference in survival in favor of the ZDV-3TC arm.

(Slide)

I am going to through this very fast because this is just weight growth presented as Z scores for these, with zero being normal growth. You can see that even with the ddI arm, which is the lower arm, there is some benefit in that it increases above zero, and the ZDV-3TC arm did better as far as weight growth than the ddI monotherapy arm.

(Slide)

Similarly for height Z scores, although there was much more variability in measuring height, particularly in young children--and the data presented is only for the under 3 stratum--there is a similar trend in the over 3 stratum but it is not nearly as impressive.

(Slide)

Probably data that people want to look at more is on CD4 counts from baseline. This is a change in absolute CD4 count from baseline, with the bottom line being ddI monotherapy, zero being no change from baseline; and the top line, solid, and being ZDV plus 3TC, and there was a substantial increase, particularly early on, but it was sustained out through 48 weeks of therapy. This is under 3 years of age.

(Slide)

This is the same data presented as percentage, which is more stable in the young infants. There is a gradual decrease over the first several years of life in CD4 counts, as has been presented. So this is CD4 percent, younger stratum, and similar in the older stratum, both favoring the ZDV-3TC arm.

(Slide)

This is the mean change in viral RNA on a log 10 scale comparing the ZDV and 3TC. There was a sustained difference out to 36 weeks, again favoring the ZDV-3TC arm.

(Slide)

This is including all of the patients that were simultaneously enrolled to the first 3 arms prior to discontinuation of the ZDV-ddI arm, and comparing them for

time to clinical progression. You can see that the ZDV-3TC and ZDV-ddI arms, at the top, overlap and were indistinguishable, both being substantially better than the ddI monotherapy arm.

(Slide)

Survival was similar in the 2 dual therapy arms compared to the ddI monotherapy arm, which is different than what had been found in 152 and we don't have a answer at this point yet as to what the difference was.

(Slide)

This is looking at t he viral load data including the ZDV-ddI dual therapy arm, which resulted in levels between the ZDV-3TC and ddI monotherapy effect. This is again limited to the patients who were simultaneously enrolled in all 3 arms.

(Slide)

So our conclusions are that ZDV plus 3TC was superior with respect to HIV disease progression, survival, weight and height growth rates, and CD4 and RNA changes.

These effects were sustained through 48 weeks of therapy. There was a mean increase in CD4 cells of 125 to 190 cells, and a mean increase between 0.7 to 0.9 logs of viral RNA.

There were no significant differences in toxicity

between the arms. There was a slightly higher rate of hepatotoxicity in the ddI monotherapy arm.

(Slide)

When patients were randomized to all 3, the 2 dual therapy arms wee superior to the monotherapy arm with respect to HIV disease progression and survival.

The differences between this study and 152 may be due to slight differences in progression criteria, and that will be looked at; some other factors which we really don't have any handle on; and chance.

The combination therapy should be preferred to monotherapy for antiretroviral therapy of infected children, particularly under 3.

And I have lost the last slide. The other thing I think I have already mentioned so we can just end it and see if there are any questions.

DR. HAMMER: Thank you very much. I would just ask the panel if there are any immediate, pressing questions for Drs. Mofenson, Palumbo or Johnson on the pediatric data. Dr. Feinberg?

DR. FEINBERG: I just have one clarification question. For the ACTG 152 group, what was the eligibility criterion around prior antiretroviral therapy?

DR. PALUMBO: The ACTG 152 pediatric population

essentially had to ...[not at microphone; inaudible.]

DR. HAMMER: Any other questions? Chris?

DR. MATHEWS: Also for Dr. Palumbo, on the 152 data, where you presented those Cox models with main effects for CD4 change and RNA change at, I think, 24 weeks, did you look in the analysis at discordant responders, or what was the prevalence of discordant responses between those 2 markers? In other words, CD4 going up, viral load going up or the opposite?

DR. PALUMBO: Those analyses are ongoing. Other than the standard analyses presented, the other analyses looking at trends and counter-trends, are ongoing. So we don't have a look at whether RNA going one direction and CD4 in another are issues that we see commonly or uncommonly in this pediatric population.

DR. EL-SADR: It seems to me that looking at the contrasting pediatric experience and the adult experience is that the clinical events, primarily CNS and growth events, happened pretty rapidly. You can see differentiation in these trials pretty early. So it is different from the adult clinical trials where you have to continue the trial for a very long time until you get the clinical events. How much of an advantage is it to look at the 24-week viral endpoint event versus 36 weeks clinically then? I mean, did

you actually try to look at it this way?

DR. JOHNSON: I can say for 300 that we haven't yet looked at that. This is about what we have. The virology that we have is about 70 percent of what needs to be done.

DR. PALUMBO: Yes, I would concur with George. ACTG 300 ended rather abruptly and quickly from start to finish with clinical endpoints only. The problem I think we are going to see though, as we progress in pediatrics and as we have seen with our adult colleagues, is that as we are able to lower viral load to lower quartiles, shall we say, the risk is going to become much lower for disease progression and I think in pediatrics, hopefully, we will lose the ability to use clinical endpoints routinely in very aggressive protocols. So I think we are really going to need to use CD4, RNA and any other laboratory markers we can find to better define how clinical regimens are working or not working in future clinical trials.

DR. JOHNSON: That is ongoing in the pediatric ACTG, that transition to using virologic and other surrogate markers.

DR. HAMMER: I just have one quick question for Dr. Johnson. One thing that is striking in the ddI arm therapy is the 0.2 to 0.3 reduction at week 12, a lower

reduction than one might expect as seen in other studies with ddI. Is there any drug adherence data? One wouldn't expect that much in the way of resistance.

DR. JOHNSON: That was collected but we don't have that analyzed yet. That was collected by report so there was no monitoring of levels on a compliance type of basis.

DR. HAMMER: Thanks very much. I think we should move on now to the open public hearing. We have a number of speakers signed up.

There are a couple of issues first. Please limit your comments to five minutes or less. Also please make any financial disclosures. If there are no financial disclosures to make, please state so.

The first speaker in the open public hearing session is Dr. Victor DeGretolla, from the Harvard School of Public Health.

Comments, Victor DeGretolla, Ph.D.

DR. DEGRETOLLA: Dr. Victor DeGretolla, from the Harvard School of Public Health. In terms of the financial disclosures, I am one of the principal investigators of a project funded by Glaxo to look at the role of markers across a wide range of studies.

As everyone knows, a number of drugs have been approved in the past few years that have lowered viral

burden and extended AIDS-free survival, and the effects are seen in national surveillance, as well as clinical trials and every doctor's office. So it is natural to want to change standards for regulatory purposes.

I have a few concerns about this. Drugs, of course, have a number of important effects. In addition to reducing viral burden, they can also induce resistance both to the drug the patient has taken and to other related drugs.

So in addition to knowing RNA, you have to know something about whether a drug maintains or reduces future treatment options, and whether it works in patients who may have reduced options because of multi-drug resistance. The problem, of course, is that new drugs will be used in combination and patients take sequences of treatments. So how can we evaluate the contribution of an individual drug?

I think one thing we can do is try to do long-term follow-up regimens that include the drug of interest to find out the duration of suppression and other longer-term consequences, like what treatment options remain after a patient has had an adverse virologic response.

In addition, I think studies of the best strategies for how to use drugs may aid regulators in determining what features a drug should have. For example,

if a strategy study showed that the best way to initiate therapy is to use a protease-containing regimen and if they showed that patients should switch therapy as soon as there was any detectable virus it would be clear that what a drug should do is either work well in a protease-containing regimen to maintain virus below detection as long as possible, or work well in a group of patients who have had and already failed a protease-containing regimen.

On the other hand, if we were to see that the best way to initiate therapy was not to use a protease right away and the best time to switch would be after a patient had a viral load count of 5,000, then there would be different standards that a drug might be held to.

Now, there has been a lot of discussion about strategy trials. I saw Ellen Cooper in the audience, who has spoken about the usefulness of such trials. We haven't yet seen one. Perhaps if regulators and the Advisory Committee were to consider whether such evidence would be useful for regulation, it might help encourage the development of studies.

My only question might be do we need clinical endpoint studies any more or does RNA provide enough information? I think it might depend on patient class. If you have a class of patients for whom you can get nearly

complete and durable suppression, as well as CD4 count increases and no other serious consequences, that information might well be enough. But if you are talking about patient populations that are harder to treat, with drug-resistant virus where you can't attain full antiviral suppression, then partial suppression of RNA may not be enough and clinical endpoints might be useful.

So in conclusion, the recommendations that I would like the Committee to consider are the usefulness of long-term studies of suppression; evaluation of the effect of drugs on future options of treatment for the patient, as well as the initial RNA response; and evaluation of whether a drug might work well in a hard to treat population; and, finally, some recommendations regarding strategy studies that might demonstrate how best to use RNA or resistance information.

DR. HAMMER: Thank you very much, Victor. The next speaker is David Scondras, from Boston.

Comments, David Scondras

MR. SCONDRAS: Thank you. I represent a non-profit organization called Search for Cure that is funded by a great many sources, community groups, Roche Diagnostics, IRC, Agrom Pharmaceuticals, state departments for public health.

I have passed to each of you a more complete rendition of the testimony that I wanted to introduce today. Given the burden of testimony that you are under, I will try to make this brief and limit my comments to a reflection on the nine points that I put forth in the document which was circulated to each of you.

The first three really would be no surprise to you inasmuch as the community reflects the general community feeling, which I think is fairly widespread and which I infer from conversations with dozens, if not hundreds, of people over the last six months in the HIV-positive community and advocates that, essentially, there seems to exist sufficient data to warrant the use of viral load for the approval of therapeutics.

Having said that, I would like to turn my attention to those things that might not come on your table, and to bring to the table a handful of concerns regarding that issue that might otherwise not be discussed and that we would feel a need for some reassurance.

The first is the assumption we make, and we would hope this Committee would communicate that assumption, that the FDA will continue to ensure that companies take responsibility beyond full approval for seeing, in a long-term sense, what the effects of the approved drugs are, both

in combination, side effects and so forth. That is a presumption we make. How do you ensure that there are Phase IV trials that are, in fact, adhered to?

Another assumption we are making is that the FDA-- and this is a bit of an irony, we are sitting here, discussing the use of these tests for FDA to determine which treatments are effective while they are not permitted to be used by clinicians to make treatment decisions; they are approved for prognosis at this point. In fact, the best of them, second generation tests that Chiron has and Roche, have not been approved at all and, yet, we know there is a need for them. I think a mentioning of that to the FDA would be greatly appreciated.

A third concern of ours is that we really hope and we assume that if, indeed, you find that viral load is an applicable surrogate marker to approve with therapies, are we referring to all therapies, not just antivirals? Inasmuch as there is a wealth of data showing independence of the viral load numbers, whether you are talking about reduction in or the actual absolute value of, and clinical progression etc., it would be extremely disturbing if it turns out that there was a therapy that had implications for viral load that was not approved or supported because it was not a drug. That would create a significant concern.

We also assume that duration as well as viral reduction is viewed as important by the FDA. And we applaud them for the work they have already done and the concern they have expressed about the whole issue of duration.

However, the most important concern we have is one that I know you will find hard to focus on and I question the extent to which you can address it. But it goes like this, the FDA decision to use viral load could very well have the unintended consequence of reducing research under development of immune-based therapies, including those to restore immune function.

It is the old story of folks looking for their lost wallet while the light is good. The problem that we have with this is that there are a lot of folks who at this point do not see any reason to believe will necessarily benefit a great deal from the new therapies, people for whom drugs have failed, people in whom the immune system damage is such that it will not be reconstituted with antiretrovirals. We can show you studies, if you would like to see them, on the extent to which FDA decisions in effect directly impact the amount of dollars that go into research in certain areas.

We would, therefore, ask you, given the potential negative consequences of this decision, to tell the FDA to

take a step to ameliorate that. And I think it has an ethical responsibility to do that, in particular, to convene an assemblage of immunological experts to identify and cause data to be assembled on immunologically relevant surrogate markers so that the message that comes out from here is not that this is the gold standard at the end of the rainbow to which every dollar in the world flows but, rather, as one of many, to be followed by others.

I know it is unusual to request that the FDA become proactive instead of reactive, but I know that David Feigal has done an extraordinary job in that area at CDER and will do so at CBER. And I am hoping that you will take the time and be willing to mention to the FDA the extent to which many people are concerned that the decision will actually end up not narrowing the tunnel of researches to those chemotherapies which reduce viral load rather than broadening them to look at restoration and look at other concerns that we all have.

Thank you for taking the time to listen to me, and I really appreciate the work that you are doing.

DR. HAMMER: Thank you very much. The next speaker is Jules Levin, from the National AIDS Treatment Advocacy Project.

MR. LEVIN: I pass today. I would like to speak

tomorrow.

DR. HAMMER: Okay. Make sure you are listed for tomorrow's public session. The next speaker then is Alan Norburn, from the AIDS Treatment Project in London.

Comments, Alan Norburn

MR. NORBURN: I am speaking today on behalf of Rafid Babikanian of the AIDS Treatment Project, Executive Director of the AIDS Treatment Project. ATP is funded by a wide range of drug companies, corporate donations and organizations such as the Elton John AIDS Foundation.

There are several unique difficulties in the currently facing regulatory authorities evaluating anti-HIV drugs. The European AIDS Treatment Group believes that there is a way forward navigating between these obstacles to rational and compassionate system of HIV drug approval.

Twenty-four week surrogate marker trials can show the efficacy of the new drug in combination. Viral load and CD4 are important tools in determining prognosis and treatment efficacy. But 24 weeks is not long enough to determine the duration of the surrogate marker changes and 24-week analyses should be part of longer surrogate marker trials, one year or more, which would continue to look at duration of effect, including longer-term safety and when to switch drugs as an endpoint.

I can illustrate the point with my own personal circumstances. I am now in week 26 of combination therapy which is failing to realize the goal of complete suppression of viral replication. Having already been through the trauma of deciding which therapy to use as a first-line treatment, I am now in something of a hit and miss situation in deciding how to continue.

The reality of the situation is that surrogate marker trials are highly unlikely to fully predict all of the clinical ramifications of a particular combination therapy, and it is probably impossible to conduct traditional clinical endpoint studies with most drugs that will be seeking licensing approval in the next few years.

Ideally, the day a drug is approved I would like to know everything about it, not only how safe and efficient it is, but also guidelines on how and when to use it. This reality urgently necessitates exploration of non-traditional methods of assessing clinical efficacy of drugs, including treatment strategy trials and observational data bases. Such methods may not be as statistically sound or scientifically accurate as traditional blinded clinical endpoint studies, but at the moment they are the best hope we have for obtaining some clinically useful information. They could not say how efficient a drug is, but they would

complement the information we already have, helping us attain maximum methods once efficiency has been shown by surrogate markers. They could provide valuable information with direct clinical relevance, particularly if there were to be uniform data collection from trial to trial, allowing for easier meta-analysis.

In addition to focused safety and interaction studies, wide extended access programs should be encouraged. Targets or guidelines for the number of people who should be exposed to a drug prior to filing for licensing approval and the lengths of exposure time will help to ensure that additional safety and interaction problems are identified early on.

Patent extensions could be offered to drug companies as an incentive to conduct uniform studies and access programs. This would mean a degree of cooperation between the drug companies and frequently convened antiviral expert groups. Such groups could be responsible for advising on evaluation on post-approval commitment.

New drugs are desperately needed by HIV-positive people for whom current drugs are no longer working or whose regimes have become too unrealistic to comply with. Each month that passes by without new drugs more people become ill and die. We hold the fate of these people in our hands.

Thank you for listening.

DR. HAMMER: Thank you very much. The next speaker is Francois Houyez, from the European AIDS Treatment Group in Paris.

Comments by Francois Houyez

MR. HOUYEZ: Good morning.

DR. HAMMER: I don't think your mike is on. For your overheads, just stand aside when you show them so the entire audience can see them. We had some trouble with the earlier overheads.

MR. HOUYEZ: My name is Francois Houyez. I am with the European AIDS Treatment Group. This group is based in Europe. This group is funded by a wide range of grants from the European Union.

DR. HAMMER: I think your microphone is still not on.

(Slide)

MR. HOUYEZ: ACTG presented a month ago a document to the European Medicinal Evaluation Agency which is reviewing the criteria to evaluate new drugs. These new guidelines should be published in September or October. So in a few words, we came to a decision on such indications like these results.

(Slide)

Dale Kemp and others presented in St. Petersburg a few days ago data which clearly shows that maximum suppression of viral load is one of the main predictors of the duration of the effect of an antiretroviral treatment.

(Slide)

Based on this information, we looked at the kind of information that we have and the kind of information which people need to make their decision and to evaluate the treatment.

(Slide)

We came to the point that it is not enough just to know the percentage of people who achieve results of a viral load below such-and-such a threshold. The main information we have to know is what maximum suppression of viral load we can achieve and how long this will last, and how we decide that we have to change our treatments.

So on these curves I show you some indication of first treatment which comes to the nadir, the maximum suppression of viral load, and the different situations afterwards. The question is how long a time clinical trials have to be run to show such maximum suppression. Ideally, if the treatment really works well, you won't be allowed to see such results for a long period of time. So we came to a decision that a 24-week trial should be the best average

time, not too short but not too long. We came to the point that the slope of decay could be an indicator of treatment potency from a single drug if you compare to another regimen which doesn't contain that drug.

(Slide)

So the principles are to grant conditional approval on the basis of 24-week surrogate marker with more preliminary safety data. The 24-week surrogate marker can show initial efficacy of the new drug combination. HIV-RNA and CD4 evolution have to be evaluated or else you wouldn't be able to really know what the treatment benefit is. As I said, there is a strong correlation between HIV-RNA nadir and duration of maximum response, and there is not such a correlation between RNA at baseline and duration of treatment.

(Slide)

But 24 weeks is not long enough. This analysis should be part of a much longer trial which will document the time to switch from one treatment to another one. So if approval should be based on a 24-week trial, then we should go on with such trials to document more data about the best way and when to switch treatment.

(Slide)

In addition to this, we addressed this point with

the European agency, that trials need to look at a wider set of drugs when considering drug interactions, not especially recreational drugs but also a broader range of drugs.

(Slide)

And to compile the theories and profile in vivo as a part of the dossier. We consider resistance to be a safety issue as well because of individual consequences when you fail because of resistance. It also has public health issues from the point of view that resistance is a safety issue. This is why we would like to address the resistance data as soon as any other data in the development of new drugs, which is not the case right now.

So these are the main points that should be raised to the CPMP and I will stop here and thank you for your attention.

DR. HAMMER: Thank you very much. The next speaker is Linda Grinberg, from the Foundation for AIDS and Immune Research Project Reform. If she is not here, the next speaker is Bill Bahlman, from ACTUP, New York.

Comments, Bill Bahlman

MR. BAHLMAN: Thank you. Bill Bahlman, from Act-Up, New York. I am also an officer of the Community Advisory Board at NYU Bellevue. I wear a number of hats. ACT-UP, New York accepts no pharmaceutical grants, mostly

their own fund-raising efforts.

I am going to be speaking more at length tomorrow but I wanted to raise one point, and that was in reference to what Don Brambilla said today bout the variability of response of viral load tests and I just wanted to touch on that before giving my full remarks tomorrow.

I think it is very important that what he said relates very closely and very importantly to one's own personal care, to not make quick judgments on a drug regimen that somebody may be taking with their private physician or in a clinic, and to have confirmation viral-load tests done, as well as those who enroll in clinical trials should not jump out of a clinical trial based on one viral-load result.

But when you relate it to what the FDA is proposing in terms of accepting viral load and percentages of patients that go to undetectable levels, the 6-month period and then 48-week period, it doesn't represent the same kind of problem because you are going to be doing a viral-load tests every 2 weeks initially, then every 4 weeks and then maybe less so often but still doing them regularly for the second 24 weeks of a study, and there outlining test results should not present a problem to analyze percentages of patients that go undetectable. Thank you.

DR. HAMMER: Thank you. The next speaker is Dr.

Iris Long.

Comments, Iris Long, Ph.D.

DR. LONG: My name is Iris Long. I reside in Jackson Heights, Queens, New York. I am a pharmaceutical chemist by background. Since 1987 I have been an advocate of people living with AIDS HIV, focusing on medication development, especially during the experimental phase.

I am a member of AIDS Coalition to Unleash Power, ACT-UP, New York, a volunteer organization, AIDS activist organization, and a member of the Community Advisory Board for the AIDS Clinical Trial Unit at Mt. Sinai Medical Center, both in New York City. I have no financial association with any drug company or device company.

Women AIDS treatment activists would give their full support to new viral-load trial designs and associated drug approval process if women living with AIDS HIV are included in the process of developing and implementing these trials.

Before such trials are designed, issues concerning women's access to these new trials need to be addressed. Women with life-threatening AIDS HIV disease are still being explicitly and implicitly excluded from many drug trials sponsored by the federal government and pharmaceutical manufacturers. Monitoring women's access to these

treatments has not been effectively done by either the FDA or the government-funded ACTG program.

In 1996 Centers for Disease Control reported that the nation's female adult, adolescent population represented 20 percent, with the race ethnicity profile being 21 percent white, 59 percent black and 19 percent Hispanic. According to the ACTG's inflated data, women's participating in AIDS ACTG trials is 15 percent. However, they include in their 15 percent female participation figure a significant number of pregnant women participant enrollees, 1,400 or the 6,000, of the pediatric ACTG program. Pediatric trials are not testing treatments for women. Subtracting pregnant women gives 12 percent women participants, not 15 percent.

What needs to be done is that an ethical national research policy should be developed. Women living with AIDS HIV around the country should have ample opportunities to discuss with the FDA government-funded clinical trial programs and drug companies, patient options and issues involved in new antiretroviral trial design. The FDA must require the inclusion of women in new trial designs so that meaningful analysis of women can be done.

The National Institute of Allergy and Infectious Diseases and the FDA should closely monitor women's access to government and non-government new viral-load based trials

in all phases of development, including local site monitoring. Effective community advisory boards should be a part of this process. Thank you.

DR. HAMMER: Thank you. That concludes the open public hearing section. There is one final thing, Dr. Johnson has asked just to show his last slide to conclude our morning.

(Slide)

DR. JOHNSON: The first thing that I have already talked about was the very suboptimal therapy with ddI. Monotherapy did have some beneficial effect on disease, and that is useful information for pediatricians where compliance may be even a greater problem than in adults, and some families will only be able to comply with a monotherapy approach, at least initially.

The speculative second one there about more than a 0.3 log drop at 12 and 24 weeks may be associated with differences in rates of survival and clinical progression.

But the last, relative to this portion of the day, is the imperative to study combinations, particularly those containing non-nucleoside reverse transcriptase inhibitors and protease inhibitors rapidly in children, to allow access of those therapies.

In addition to this, just as a pediatric advocate,

I would like to push for both development and testing of formulations which are able to be administered to particularly young children, which is lagging behind adult drug development.

DR. HAMMER: Thank you very much. I have just been apprised that there is one additional request to speak at the open public session, Ron Baker, from San Francisco.

Comments, Ron Baker

MR. BAKER: Good morning and thank you for the opportunity to speak here this morning. I am Ron Baker, Editor in Chief of the AIDS Treatment publication from the San Francisco AIDS Foundation. In terms of financial disclosure, in the current fiscal year the San Francisco AIDS Foundation has been the recipient of educational grants from Glaxo Wellcome, Inc. and from Hoffman-La Roche.

I am here today to present the views of the Foundation on three issues related to the use of HIV-RNA testing in clinical studies. First, we feel that the time has come to eliminate suboptimal treatment arms in comparative clinical studies. Second, a number of studies reviewed here, this morning, have shown that viral-load measurement demonstrates the usefulness of experimental drugs more quickly and just reliably as clinical endpoints, in our view. Thirdly, we feel that drug labeling should

reflect the way in which drugs will actually be used clinically.

The recently issued guidelines for the use of antiretroviral agents are clear regarding decisions to initiate or change therapy. Such decisions, according to these recommendations, should be guided by measurement of HIV-RNA and CD4 T-cell counts, as well as by the clinical condition of the patient.

At the San Francisco AIDS Foundation we feel that it is no longer justifiable to extend clinical studies until clinical endpoints, such as the deaths of study participants have been reached. In our view, accelerated approval can be granted, as it is now, based on viral-load suppression in a significant number of study participants with perhaps as few as 18 to 24 weeks of data. Full approval could be granted to a new drug based on about a year's data, provided that the magnitude of viral suppression is great enough and remains durable. Defining "great enough" and "durable enough" remains problematic.

At the same time, it is very important for clinicians to know who does not respond well to drug so that individuals can switch to a potentially more effective regimen.

Concerning drug labeling, the Foundation supports

the use of language in the FDA indication that a drug should be used "to achieve maximal viral suppression" rather than treatment of HIV disease. This more precise and descriptive language in labeling will provide greater clarity to physicians and to patients concerning the real objective of antiretroviral therapy, which is to reduce the viral load to as low a level as possible for as long as possible.

In summary, the Foundation asks the Committee to recommend, first, the elimination of suboptimal study arms. Second, we strongly support the use of viral-load measurement as an endpoint in AIDS drug studies. Finally, we advocate the use of labeling that accurately reflects how a drug affects HIV viral load.

These are urgent concerns and we feel they should be implemented by FDA as quickly as possible. Thank you.

DR. HAMMER: Thank you very much. If there are no other requests to speak at the public hearing, I am going to adjourn the morning's session. We can reconvene in one hour, at 1:20.

[Whereupon, at 12:20 p.m., the proceedings were recessed, to be resumed at 1:20 p.m. this same day.]

AFTERNOON SESSION

DR. HAMMER: We are waiting for the sound engineer, but if you will please take your seats, we will start in a moment. We will start with this one microphone that appears to be working. This afternoon's session is devoted to a series of presentations, first, a review of antiretroviral guidelines, and then clinical confirmation of HIV-RNA changes. I would like to begin with Sherilyn Stanley, from NIAID, who will review the antiretroviral guidelines.

Review of Antiretroviral Guidelines, S. Stanley, M.D.

DR. STANLEY: Thank you very much. I would like to take this opportunity to thank the organizers for inviting us to review these for you all.

(Slide)

I am sure I don't have to remind this group of the rapid evolution that we have experienced in antiretroviral therapy over the past several years. This slide just sort of summarizes that schematically. In '87, with monotherapy we could get nice suppression of viremia but it did not persist. By '94 we could give combination 2-drug therapy, and again we could get nice suppression of viremia but generally it did not persist. Now, in 1997 we have available to us the combination therapies, including the

protease inhibitors, and we are seeing again in a lot of our patients marked suppression of viremia with some long-term suppression in some patients.

(Slide)

The evolution of therapy and the advances, along with the rapid increase in number of drugs that have been approved for HIV therapy, have caused a lot of confusion in the field of the clinical practitioner. This was recognized by the Secretary and she requested that Eric Goosby, head of the Office of HIV AIDS policy, in conjunction with Henry J. Kaiser Foundation, convene a panel that would have a three-year life span at a minimum that would address clinical issues of therapy for HIV-infected patients, not just antiretroviral therapy but also other issues, for instance pediatrics perhaps, perhaps managed care issues, whatever clinical aspects of HIV management this panel wished to address. The first issue that with some urgency the panel felt needed to be addressed was antiretroviral therapy.

(Slide)

This is the panel that was put together, again, co-convened by the Henry J. Kaiser Foundation and OHAP. The co-chairs were picked to be Dr. Fauci, as the federal representative, Dr. John Barton, from Johns Hopkins, as the private community representative, thus, again solidifying

that this was a unique public-private partnership.

The panel members were drawn from a large variety of community activists, active commissions, researchers and, as well, multiple HHS agencies were represented as members of the panel.

Again, this panel has been constituted for three years and they have just released, in the Federal Register, the draft guidelines for antiretroviral therapy. Those guidelines are available through a 1-800 number. We are still in a period of public comment. At the end of that period the comments will be considered by the panel and a final document will be released.

Let me just summarize for you the process that the panel went through. The panel met three different times and discussed various aspects of antiretroviral therapy and the data that was available. They relied on the recently formulated principles of antiretroviral therapy, which is a document that is traveling in companion form with the guideline document.

The principles document was derived by an NIH-convened panel to develop principles of antiretroviral therapy. These are basically the eleven pathogenetic principles that provide the rationale for why to treat and how to treat. So the clinical practices panel used that

information as well as the experience of experts and their own clinical experience to derive the draft guidelines.

Critical in their considerations was one of the principles of HIV disease, which is the fact that HIV replication is constant and active during all stages of HIV disease. This would lead one to think that one should perhaps treat early and treat aggressively.

(Slide)

However, the panel also considered very seriously the data by Mellors et al., showing that depending on the level of RNA copies, time to progression to AIDS is very much faster with the higher viral burdens as opposed to those with less than 9,000 copies who have a prolonged progression to AIDS and a fairly prolonged symptom-free survival.

(Slide)

This is also shown in this graph where, as virus levels increase going in this direction and as CD4 levels decrease going in this direction, you get increasing progression with likelihood of development of AIDS within three years.

(Slide)

Considering this data, the panel felt that there were also other considerations that would weigh either for

or against early therapy in the asymptomatic individual. These were summarized in a table in the document. They were also discussed at length in the text of the document. Again, this is for making the decision for the practicing physician when do you treat the asymptomatic HIV-infected individual. The aggressive clinician, on the basis of active viral replication, might say treat early. The more conservative clinician or the patient that has other considerations might say let me look at my chance of progression to symptomatic AIDS, and let me weigh these other factors in here and decide whether I wish to delay therapy for some time or aggressively treat at this point. The panel tried to reach this sort of center view that allows the aggressive clinician to treat early, but also allows for and recognizes the legitimacy of delaying therapy in the asymptomatic individual, such that the recommendations for therapy are really summarized in Table 5 in the document.

It was absolutely universal that any patient with symptomatic HIV infection or AIDS, no matter what the CD4 or HIV-RNA levels, should be treated. However, in the asymptomatic individual with CD4 T-cells less than 500 or HIV-RNA greater than 10,000 by bDNA or 20,000 by RT PCR, in this category of patients treatment should be offered based

on the scientific principles that we understand with HIV disease. But the strength of the recommendation for treatment should be balanced on the prognosis for disease-free survival and the willingness of the patient to accept therapy and, some of those other considerations that I showed you in the last slide. So this is really the group where the approach of the clinician, whether aggressive or conservative, and the considerations of the patient will make the most difference.

Then in the asymptomatic patients where the CD4 counts are greater than 500 and the HIV-RNA is very low to undetectable, some experts would delay therapy and observe but still there are some experts who would treat these patients based on the principle that there is always ongoing viral replication and this is detrimental to the immune system. Therefore, this option was allowed by the panel.

(Slide)

In order to help make this document as user friendly as possible and to help the clinician make these decisions, there is a table in the document that gives the Mellors data: plasma viral load, CD4 count and percent developing AIDS in three years, six years and nine years. This is the number of patients that fell into each category. These are the less than 350; these are 351 to 500. Again,

allowing clinicians to use either bDNA or RT PCR, with an explanation of how these convert from the MAX data which, of course, was generated on frozen, stored specimens. Then for the patients with greater than 500 CD4s this data is given. So the document really provides the clinician with the survival best data that we have to date, as well as the other rationale, for or against, for the early treatment of the asymptomatic patient.

(Slide)

Once the decision to treat has been made, the panel was unanimous in stating that three drugs are the first choice of therapy including a protease inhibitor, so that two nucleosides or reverse transcriptase for instance and a protease inhibitor, or perhaps a non-nucleoside, reverse transcriptase inhibitor. But it was unanimous that once you decide to treat, you need to be aggressive so that you achieve maximal viral suppression with all the benefits that that obviously gives, which is decreased development of resistance and other positive factors.

(Slide)

That decision was really based on the multitude of clinical trials we have showing benefits of triple-drug therapy, either virologic or clinical benefits.

(Slide)

Table 6 in the document shows what the panel considered to be the optimal choices and the alternative choices. You will notice that in this table the preferred choice is a highly active protease inhibitor with two NRTIs, and the potential combinations are shown.

An alternative regimen, again, for initiating therapy in the asymptomatic patient was considered to be nevirapin and two NRTIs, as above, saquinavir in its current hard-gel capsule formulation.

(Slide)

Less desirable and, in fact, not recommended unless there was some sort of special clinical situation, is the use of only two NRTIs without a protease inhibitor or other third drug. Again, absolutely not recommended and probably contraindicated are all monotherapies and these particular combinations of two NRTIs.

(Slide)

The panel addressed the issues of when do you change therapy. In relationship to this meeting here, based on the data that was reviewed by the clinical guidance panel as well as the principles panel, the scientific principles enumerate in the principles document that a ten-fold reduction in viremia at four weeks is a sign of successful therapy, and that virus that is not under ten-fold by four

to six months after therapy probably reflects suboptimal treatment.

This is a table from the document. The first bullet is discussed at more length in the text of the document. I will recommend that that be read. The guidelines for changing therapy would include suboptimal reduction of plasma viremia after initiation which, as I said, the panel took to be ten-fold reduction at four weeks; reappearance of viremia after suppression or detectable virus at four to six months after therapy; significant increases in viremia from the nadir of suppression; decline in CD4 numbers and clinical deterioration. Again, there were several caveats of things to consider when a clinician is considering changing therapy. I won't go through all of those. They are included in the document.

(Slide)

The panel went so far as to try, with the minimal data is available, to help the practicing clinician by making some recommendations for what a potential change would consist of. Again, this table, which is in the document, outlines what the prior regimen might be; what you might consider switching to, given again that this is mostly expert opinion because of the paucity of data that is available.

The panel did consider acute retroviral syndrome. It was felt it was important to reeducate the clinician on what that is so that we can perhaps achieve better recognition of this. So in the document is included a list of the symptoms. The panel basically said that most experts would probably treat a recognized case of acute retroviral syndrome but was unable to come to a firm conclusion about the length of such treatment, and ended up stating that probably treatment should go on indefinitely once it has been initiated, again because of a lack of data.

(Slide)

Again, relative to this meeting, the panel obviously put great emphasis on HIV-RNA testing and felt that although it is still very important to get CD4 counts to understand the immunologic condition of the patient, viral burden data is what people are really using out there in the academic settings and in the clinical settings to guide their decisions about antiretroviral therapy.

So there is a table that instructs the clinician on what the clinical indications might be for RNA testing; what information you would hope to get; and how you would use that in your decision making.

(Slide)

This is simply the second half of that table.

(Slide)

Just to give a flavor of some of the tables that are in the document, I am not going to go through them but there is a table describing characteristics of the NRTIs, the NNRTIs--

(Slide)

--the protease inhibitors--

(Slide)

--and this one goes on for some time.

(Slide)

Drugs that should not be used, so that this will be available, hopefully, for the clinician's reference. We ran these past David Feigal and his staff at the FDA, and we hope that they are accurate and we will continue to update them as more information becomes available. We appreciate David's staff time on this.

(Slide)

Again, drugs which should not be used in combination by category.

(Slide)

Various clinical drug interactions that are of significance.

(Slide)

This continues.

(Slide)

Also interactions between the PIs and the NNRTIs listed to date.

(Slide)

And this is the second half.

(Slide)

Finally, the panel also addressed the use of antiretroviral therapy in pregnancy, and the recommendation is basically that the woman's clinical status should be the primary determinant factor in treating or continuing therapy with, of course, special considerations that one must think of the unborn fetus and possible teratogenicity and carcinogenicity effects. There is a table in the document that gives what information we have on that.

So where are we at? We have a week left in the public comment period. I urge you, if you haven't obtained the document to get it and to please give us feedback. After the comments are then put together, the panel will review these, will make appropriate changes in the document and the document will then be published, hopefully, within the next several months.

There is a mechanism being put in place for this panel not only to go on and address other issues, but to continually update this document so that as we get more data

on various agents or new agents appear they can be added to the document.

I will be glad to answer any specific questions.

DR. HAMMER: Are there any questions? I think no. Thank you very much. We will now move on to the session on clinical confirmation of HIV-RNA changes. The first speaker will be Dr. Jeffrey Murray from the Division of Antiviral Drug Products, FDA.

Introduction, Jeffrey Murray, M.D.

(Slide)

DR. MURRAY: I am Jeff Murray, one of the FDA reviewers who helped in planning this Advisory Committee session.

I would like to spend just a moment to acknowledge the other individuals and groups who helped put this meeting together because it was quite a bit of an effort. We have members from FDA Antiviral Drugs, but also we had a lot of help from the Surrogate Marker Collaborative Group, other IND and NDA holders, and also this meeting was put together with some feedback from community groups at a meeting that we hosted a month or so ago.

(Slide)

We are now beginning the part of the meeting at which analyses of the clinical trial data will be presented

and explored. This afternoon there will be presentations on the clinical correlation of treatment-induced HIV-RNA changes. Tomorrow we will examine behavior or pattern of HIV-RNA in response to various antiretroviral treatments.

(Slide)

In preparation of the following presentations and preparation for the Committee's discussion tomorrow, I want to comment briefly on our rationale and our objectives for this meeting and for the presentations that will immediately follow.

There are several reasons this is an opportune time to reevaluate clinical studies supporting traditional approval. First, there are new therapeutic goals and guidelines, as you have just heard from Dr. Sherilyn Stanley. Specifically, we are in the midst of a shift in the way HIV-infected patients are managed clinically. In contrast to the setting in which the past clinical endpoint trials were conducted, we are now working in the setting of HIV monitoring, potent drug combinations and new goals of therapy, that is, to maximally suppress virus.

Second, we realize that it was an opportune time because there was a sizeable accumulation of clinical trial data describing relationships between virologic changes and clinical disease progression. There is also a large amount

of clinical trial data describing the behavior of HIV-RNA response to some potent drug therapies. We felt that the state of this is sufficiently rich to start making some recommendations for improving trial design and we thought it would be a missed opportunity not to explore this wealth of information.

(Slide)

So, in short, our overall goals for this meeting were to explore using HIV-RNA as an endpoint, and also as another option for additional approval, and also as a label indication. We anticipated that using HIV-RNA as a primary endpoint could have potential advantages in clinical trial design for participants, for investigators, for sponsors and also for FDA.

(Slide)

The reasons for considering the use of RNA as a clinical endpoint for efficacy trials are really some potential advantages. That is, it is a less complex endpoint than the current clinical endpoint which includes approximately twenty different infectious diseases, malignancies and other conditions.

Second, using RNA as an endpoint coincides with medical practice, and you can't ignore this. Physicians use this test to make recommendations.

Third, we felt that an endpoint that we have actually proposed, an endpoint such as time to loss of virologic response could be appealing for trial participants and also for those analyzing the data. This type of endpoint could easily allow treatment switches before clinical failure and, at the same time, treatment switches would not necessarily disturb the study analysis because an endpoint would have been achieved before a treatment switch was actually required.

Last, with the advent of more potent regimens, powering studies for the relatively infrequent clinical endpoints was and is becoming more difficult. So powering studies with respect to detecting differences in RNA endpoint could offer us economy with respect to sample sizes.

(Slide)

Although we want to consider RNA endpoint studies as another option for traditional approval for antiretroviral drugs, we were confident that certain conditions have to be preserved. First, clinical endpoint studies need to remain an option and still should be encouraged for answering certain clinical questions as necessary. That is, RNA studies can support an indication for lowering RNA. Clinical studies can support an

indication for delaying disease progression, for AIDS dementia and for other type of clinical questions.

Second, we are fully committed to keeping all early access mechanisms intact, including accelerated approval based on earlier, such as 16 to 24-week changes in both CD4 and RNA. For the RNA endpoint short-term changes would just simply be confirmed by longer-term RNA studies demonstrating durability.

Third, we are also committed to an accurate evaluation of safety and tolerability. Although RNA studies would be powered for an RNA endpoint, clinical disease progression data would still be collected. Information on safety, tolerability, CD4 data, we all realize this is still important to be looking at.

(Slide)

Before RNA could be used as a primary endpoint for traditional approval, we believe the following questions need to be addressed: First, we need to be confident that there is evidence that RNA reduction in itself is associated with decrease in clinical progression rate.

Second, we need to know best how to measure and analyze these changes, what was most clinically relevant. Third, we want to explore if there are any considerations for using this endpoint in special subpopulations, and that

is why we devoted a considerable amount of time this morning to looking at the pediatric data that is available. The clinical correlates presentations that follow will address numbers one and three.

(Slide)

For this set of presentations that will immediately follow we developed a number of analyses questions pertinent to the relationship between RNA reduction and clinical benefit. These are, to evaluate whether reduction of plasma RNA confers benefit; to describe the relationship in magnitude and duration of HIV-RNA reduction and clinical disease progression; to describe the relationship between virologic nadir and clinical disease progression.

(Slide)

To explore whether the prognostic significance of HIV-RNA is dependent on baseline factors; to describe appearance HIV-RNA changes around the time of the clinical event, before and after; and also to explore the proportion of antiretroviral treatment effect that is mediated by changes in both virologic and immunologic endpoints.

(Slide)

So we are asking the Committee to evaluate a collection of clinical studies. This involves more than

just pooling of data to produce a single mathematical estimate of treatment effect explained. In fact, this has not been done and you will not hear that.

Instead, what will follow are five separate presentations. Some use data from one trial and others have combined data when appropriate. There is a heterogeneity in the studies. Not all studies use the same assays; not all studies measure the same time points or study the same populations.

Although this heterogeneity can sometimes prevent pooling of data, it is also very informative. For example, you will see that all of these analyses show that reduction of plasma HIV-RNA is associated with a decreased risk of disease progression, and they all show that the more HIV-RNA is lowered the greater the reduction for the risks for disease progression. It is the consistency of results across different studies and different patient populations that constitutes very strong evidence.

However, making a decision about the relevance of RNA requires beyond the presentations today. There are other pieces of this puzzle. For example, we must apply what we know biologically, particularly with respect to the development of resistance. We can't ignore the fact that concerns about drug resistance and the continued usefulness

of the regimens that we are approving is directly linked to the maximal suppression of virus.

We will now begin these very interesting presentations. The first speaker will be Dr. Ian Marchner, from the Harvard School of Public Health, and he will be presenting some data from various ACTG trials.

Presentation by Ian Marchner, Ph.D.

DR. HAMMER: Thank you very much.

DR. MARCHNER: My name is Ian Marchner. I am a statistician with the AIDS Clinical Trials Groups, statistical data analysis center at Harvard.

(Slide)

I am going to be talking about a pooled cross-protocol analysis of a number of ACTG studies that we conducted to assist the utility of treatment-mediated changes in plasma HIV-RNA for predicting clinical progression rates.

(Slide)

Just to give you some background with regard to the studies we are analyzing, this was a cross-protocol analysis of seven ACTG studies involving a variety of different treatment regimens. Most of the treatment regimens we are dealing with involved nucleoside reverse transcriptase inhibitors, particularly ZDV, ddI and ddC.

There was one study that included nevirapin and one study that included saquinavir.

We made a decision in this analysis to include even monotherapy arms in the analysis, the reason being that we are interested in looking at the association between RNA responses with clinical progression, rather than the effect of therapies on response.

The sample that we are dealing with is going to be all individuals in these studies that had baseline RNA measurements and CD4 measurements. Some of the analyses that were performed included all such individuals, and they numbered 1,330 altogether. Most of the analyses, in fact probably all the analyses that I am going to present today are concerned with changes in RNA up to week 24, and their association with clinical progression. So we could only include individuals who were at risk of clinical progression at week 24 and who also who had CD4 and RNA data at week 24. These individuals numbered 1,000.

Everything that I am going to be talking about today, with the exception of one brief comment about the Chiron assay, is concerned with the Roche PCR assay, the original assay.

(Slide)

This is just a very small summary of the data that

I have included in the packet that the Committee got by way of background. The crucial elements in this table are, firstly, the sample size which is 1,000, primarily for the individuals who had the data that I talked about on the original slide. There were 120 clinical progressions, defined as a new AIDS-defining clinical event or death among those 1,000 individuals.

The CD4 at baseline was approximately 200, the median RNA was 50,000 and, in terms of the follow up, the median follow up overall was approximately 1 year, with follow up out to about 3 years in some individuals.

Here also is a summary of the different therapy arms that we have on the studies. You can see that there are monotherapy arms together with triple therapy arms, and triple therapy here as well.

(Slide)

So what I am going to do is basically just present you with three or four key questions, for the purpose of brevity, and associate with each question just a single key graphic to give a picture of what I think the answer to the question is.

The first question I am going to deal with is how is the magnitude of response related to the reduction in clinical disease progression rate? What I am going to use

as response is the change in HIV-RNA over the first 24 weeks of each of these trials.

(Slide)

This is really what I would see as the key graphical information relating the magnitude of response to the reduction in clinical risk. What I have on the X axis is the change from baseline to week 24 in HIV-1 RNA. what we have done is to split the change up into 8 groups, octiles, and then to estimate essentially the hazards ratio for clinical progression in each of these octiles. Then the hazards ratio is plotted against the median change in each of the 8 octiles.

The first thing that obviously stands out is that individuals with no reduction or even a slight increase are at the greatest risk of progression, whereas individuals who had greater reductions are at less risk for progression. You will notice that for this data set we are dealing with reductions at about 1.5 logs from baseline.

But probably the more striking feature is the fairly strong linearity between the adjusted hazards ratio-- I should say that this is adjusted for the baseline level. So this is the effect of changes adjusting for the baseline level. There is a very strong linear relationship, indicating that the decrease in clinical progression risk is

proportional to the reduction in HIV-RNA out to 24 weeks. Implicitly what that tells us that larger reductions are more beneficial than smaller reductions, but still smaller reductions have some clinical benefit.

This plot that I just showed you in some sense gives us information about the next question, although I prefer another plot for answering it in more detail. The question is what descriptors of the magnitude of response are most clinically relevant. I have sort of taken this to mean are we interested in absolute response or absolute level of RNA achieved, or is there some threshold beyond which there is no real change in clinical progression risk?

(Slide)

The analysis that we have done is analogous to the one on the previous plot, but now on the X axis what we are dealing with is the absolute week 24 RNA level as opposed to the response or reduction over the first 24 weeks. We can see a very striking linear relationship, indicating a proportional relationship between the risk of clinical progression and the absolute value achieved after 24 weeks of therapy on these studies.

What we see here is no evidence of a threshold relationship in the sense that the lower your week 24 RNA is, the better your clinical progression risk is, and this

doesn't seem to level out as we get to the lower week-24 RNA values.

The next thing I want to talk about is the effect of baseline RNA level on the interpretation of treatment-mediated reductions and, in particular, this question really has two aspects to it. The more transparent aspect is, is baseline RNA level an independent predictor of clinical progression risk over and above the reduction in RNA? And we would perhaps expect that it would be. But a more subtle question is whether or not the baseline level modifies the interpretation of the HIV reduction over the first 24 weeks, and that is what statisticians like to refer to as interaction between baseline and treatment-mediated response.

(Slide)

In terms of the independent predictive ability of baseline and response, this is just a crude summary giving a feel for the fact that baseline level and reduction are both independent predictors if you do more sophisticated analyses, and that certainly comes out.

What we have here is individuals divided into 4 categories depending on whether or not they had any decrease in HIV-RNA out to week 24. So there are two groups, one who had a decrease; one who didn't have a decrease. Then

whether or not they were above or below the median of 55,000 at baseline. What we see is, for example, the top 2 curves refer to groups below the median at baseline. We have separation here, statistically significant separation dependent on whether or not they had a reduction in RNA after week 24 and then, likewise, for the individuals who were below baseline we again have separation dependent on their response, suggesting to us that baseline and response are independently predictive.

(Slide)

Now getting at the more subtle question of whether or not there is interaction between the two or whether the baseline level modifies the interpretation of an HIV response, what we have here are hazards ratios again. On the X axis what we have is individuals divided into whether or not they had 1 of 3 week-24 responses: no decrease, or an increase, a moderate response, from 0 to 0.5 log reduction and a better response of greater than 0.5 log reduction. In the 3 lines are just the baseline level categorized approximately into 1/3 percentiles, so less than 20,000, 20,000 to 100,000 and above 100,000.

In each of the 3 baseline categories we see that the risk increases as the week-24 response gets worse. What we see between each of the 3 curves is that the higher the

baseline level, the higher the risk of progression.

The important point though is that the increase in risk or the decrease in risk, however you want to look at it, is very similar for the 3 different baseline categories, suggesting that the interpretation of a given week-24 response is similar for the different baseline categories. So, for example, a 0.5 log reduction decreases your risk of clinical event by about the same ratio regardless of what your initial baseline level was.

(Slide)

We had a fairly substantial talk today about issues of variability in interpreting these responses. I won't spend very much time at all on this. But I just want to say that this is basically an intersection of some of the data that Don Brambilla collected. This is looking at peak baseline measurements of HIV-RNA. In terms of the Roche assay, we found basically about 90 to 95 percent of successive measurements, in other words, repeated baseline measurements in the absence of a treatment effect, were within a range of about 0.5 log peps up to about 6 logs.

So the Chiron assay, and this is the only comment I will make about the Chiron assay, I don't have any clinical progression data related to the Chiron assay, but what we did find for another study, ACTG 306 in fact, was

that we had a similar level of variability for individuals who started out at greater than 10,000 copies but the variability with the Chiron assay was somewhat greater than with the Roche assay for individuals less than 10,000 copies. This is a result that we have picked up for descriptive analyses and it requires further investigation. It was hinted at in some of the data, I think, that Don put up today.

(Slide)

What I will do before summarizing the results is just make some comments about the relationship between CD4 response and RNA response. What we have here is a plot of the week-24 RNA response against the week-24 CD4 response. As you would probably expect, there is a negative relationship in the sense that individuals with reductions in RNA tend to have increases in CD4.

Importantly, this response is not particularly strong. It is statistically significant but the correlation coefficient is only 0.3. In fact, if you look at any given range of RNA responses, say, zero to 1 log reductions, the range of CD4 responses is very wide-ranging, from almost 200 cell decreases to 200 cell increases.

(Slide)

So what this suggests is that given the potential

for discordance between the RNA response and the CD4 response, the two markers could potentially both be useful in assessing treatment response and predicting clinical progression.

What I have here is just one summary of the various joint analyses that we did for CD4 and RNA. What we have again on the X axis is the RNA response divided into 3 groups, no response, moderate response and better response, and 3 curves corresponding to the CD4 response, no response, moderate response, greater response.

In all cases we see, as we go from the better RNA response to the moderate response, we get an increase in clinical progression risk. This is, in fact, statistically significant. But then as we move from the moderate response to no response, the increase in clinical progression risk is really very dependent on the CD4. If you combine a poor RNA response with a poor CD4 response, that is clearly much worse than if you combine a poor RNA response with perhaps a moderate CD4 response.

So that is really suggesting to us that both CD4 and RNA contain prognostic information and perhaps both should be used in assisting prognosis.

(Slide)

By way of summary of the key points, the reduction

in clinical progression is proportional to the reduction in HIV-RNA out to 24 weeks. That basically implies the second point here, the larger the reduction, the better the decrease in clinical progression risk.

I should point out the caveat that we didn't have a high percentage of individuals reaching below the limits of delectability by the Roche assay in this data set, and we were dealing with responses out to about 1.5 logs so whether this proportional relationship persists for larger reductions is not clear from these data.

The implication then from this proportionality relationship is that any response beyond what could be deemed to be just assay variability or biologic variability could be seen as clinically beneficial. So one figure that has been banded around is about a 0.5 log drop as being indicative of biologic variability. Don had a slightly larger figure this morning, maybe 0.5 to 0.7 .

(Slide)

The last two points are to summarize for you that the baseline level of HIV-RNA doesn't seem to modify the clinical interpretation of the week-24 RNA response in the sense that a given response can be interpreted singly for individuals of different baseline levels.

Although the baseline level is an independent

predictor of clinical progression, it just doesn't modify the interpretation of the RNA response.

The last point that I would just reiterate again is the importance of CD4 as an additional indicator of prognosis, over and above RNA. Thank you.

DR. HAMMER: Thank you. I would like to ask the Committee members to hold their questions until the end of this section and we will have a combined and answer period later.

DR. LIPSKY: I would just like to clarify one thing, please.

DR. HAMMER: One clarification, sure.

DR. LIPSKY: Could you put up the only Kaplan-Meier plot that you have, please?

(Slide)

Does that indicate that a group where the viral counts actually went up greater than baseline did better than some groups where they actually went down?

DR. MARCHNER: Yes, in a sense because you might have a group that started out lower and had perhaps no decrease or slight increase, and then comparing that with individuals that started out very high, the former group may have a better prognosis because they started, say, from 10,000 and went to 15,000 or 20,000 compared with starting

at 100,000 and maybe going down to 80,000 or going up to 120,000.

DR. LIPSKY: Clearly, something like that is happening?

DR. MARCHNER: Yes, that is what you would expect because the former group had a lower level to begin with.

DR. LIPSKY: Even though they are virologically getting worse?

DR. MARCHNER: Right. You have to distinguish between improvement in progression rate and absolute progression rate. The first group that you talked about didn't have any improvement but they were starting from a better beginning point than someone who had a very high absolute progression rate and had a slight improvement to improve that high progression rate but still didn't get down.

DR. LIPSKY: So your data would say that an increase in viral RNA is not necessarily a bad thing?

DR. MARCHNER: Well, it is a bad thing in the sense that you haven't got the improvement of the therapy. You haven't gotten any benefit from the therapy. But if you have an increase in RNA, it doesn't necessarily mean that your absolute risk is going to be worse than someone who had a moderate decrease. That person with the moderate decrease

started out from an extremely high level.

DR. HAMMER: If you went up from 2,000 to 10,000, you went up but your risk is going to be lower than if you went from 500,000 to 100,000.

DR. LIPSKY: It seems though, you know, we are looking at a cutoff at 55,000 on baseline--that is up there. I mean, there is something that seems a little bit disturbing about that Kaplan-Meier plot because we are not talking about 10,000 and 2,000. You are up at 55,000.

DR. MARCHNER: Yes, but we are dealing with very large categories here. If you break things down into smaller categories or deal with things on continuous levels the same sorts of results are going to apply. I don't think it is particularly surprising that a person's absolute level should be primarily indicative of where their absolute risk is at any given point of time. Therefore, someone with a lower value could well be better off even if they didn't have a good response.

DR. LIPSKY: Well, I think that certainly has strong implications for changing therapy etc., because you clearly have a group that appears to have done better than another group --

DR. MARCHNER: The point is you have to take into account two things, not just the response but where the

person is.

DR. LIPSKY: You showed some graphs that were called linear. Could they not be sigmoidal, particularly the first one, actually log linear and intercepts below zero. Does that not have implications about threshold?

DR. MARCHNER: The evidence for a sigmoidal relationship in any of those plots wouldn't have been very strong. I certainly want to draw the caveat that we don't have a lot of individuals going down to extremely low levels, and whether or not that trend persists is not 100 percent clear from these data. But I wouldn't be prepared to argue that those curves were supporting any sort of threshold relationship.

DR. LIPSKY: But it is certainly log linear if it is a linear relationship.

DR. MARCHNER: Yes. That is the sort of natural mathematical scale to present risks on.

DR. HAMMER: Thank you. There will be time for more questions later. We are going to start now having a group of four presentations by pharmaceutical manufacturers. The first presentation will be by Ralph DeMasi and Lynn Smiley, from Glaxo Wellcome.

Presentation by Lynn Smiley, M.D.

DR. SMILEY: On behalf of Glaxo Wellcome, we

appreciate the opportunity to present our data today. I would also like to echo the acknowledgements given by Dr. Jeff Murray at the reopening of the session earlier this afternoon. The work presented today and tomorrow really represents a culmination of collaboration over the past year and a half among many groups.

(Slide)

Following some introductory overviews of the trials, we have analyzed, and that will be presented today, I am going to turn this over to Dr. Ralph DeMasi, a statistician within our group, who was project leader internally within Glaxo Wellcome's initiative.

(Slide)

What you are going to see today is a cross-study analysis done retrospectively of six controlled trials, sponsored by Glaxo Wellcome, that were completed within the past two, two and a half years. About two-thirds of the data are on patients who received zidovudine plus 3TC. The remaining one-third were randomized to the control treatments.

The cross-study analysis included the CAESAR trial. The CAESAR study which was our adult clinical endpoint study for 3TC, 85 percent of those patients were treatment experienced.

The NUCA and NUCB3001 studies, which were surrogate marker trials in less advanced patients, and these patients were naive. The NUCA and NUCB3002 studies were conducted similarly but included treatment a experienced population.

The AVANTI01 trial was a trial of ZDV-3TC versus ZDV-3TC plus and investigational nucleoside reverse transcriptase inhibitor.

Our data are from 1,581 patients who had RNA done at baseline and at least one follow-up visit. Our endpoints are from 197 patients and included 268 clinical disease progression events, either new AIDS-defining events or death.

(Slide)

All clinical trails were randomized, double-blinded, controlled studies, with a mean duration follow-up of 1 year. All AIDS events in the CAESAR trial were independently reviewed by an endpoint committee. The plasma samples were tested using the Roche Amplicor assay in all studies. As I mentioned, RNA was measured at baseline and approximately every 4-8 weeks on study.

(Slide)

The metrics of RNA response included looking at nadir or the lowest level achieved, as well as the peak

response, which was the maximum change from baseline. Ralph will show some data that includes the 8-28 week mean change from baseline, as well as an 8-52 week mean change from baseline. Cox multiple regression model was used and the intent-to-treat population was analyzed.

(Slide)

The demographics and baseline characteristics showed that approximately half of the population in this cross-study analysis was treatment naive, as defined by having received less than 6 months of nucleoside therapy. Predominantly male population. The disease stage is shown, partition across CD4 A, B and C. The mean CD4 count was around 200 and the mean RNA level at baseline was about 4.8 logs, or about 60,000 copies/mL.

At this point I am going to turn it over to Dr. Ralph DeMasi to present the results and conclusions.

Presentation by Ralph DeMasi, Ph.D.

(Slide)

DR. DEMASI: I would like to start off this part of my presentation with some descriptive analyses, looking at the correlation between the magnitude and duration of RNA reduction and the reduced incidence of risk of clinical progression. I realize that some of this may be hard for you to read so I am going to walk the axes, the Y axis and

the X axis and then the particular points that we are looking at on the plot.

This particular plot looks at the progression incidence by duration of RNA below 5,000 copies/mL. What we have here on the Y axis is progression incidence, defined as the number of events per patient year of exposure. The X axis represents the duration of RNA reduction of 5,000 copies/mL. The particular points along the X axis, the first point here is zero weeks. This means the patients who have actually shown no reduction in RNA to below 5,000. The next point is 0-12 weeks or up to 12 weeks reduction; 12-24 week reduction and then greater than 24-week reduction to the far right. Again, the Y axis is the number of events per year of exposure, and the higher it is, the higher the risk for incidence of clinical progression.

What this first analysis indicates is the relationship here between the more durable RNA response and the decreased risk of clinical progression, and that the effect of longer duration of suppression is roughly proportional to the decreased risk in clinical incidence. The numbers here indicate the actual number of events, or 204 events and the 722 at the lowest point, and then at the highest 7/529 patient years.

(Slide)

Now I would like to turn to looking at different metrics of the magnitude of response. This first metric is the maximum change in RNA. It is a peak response over the treatment period, and now I have stratified this analysis by baseline RNA values.

The Y axis here is actually the progression incidence over 100 patient years of exposure. The X axis is the magnitude of the peak response categorized into certain distinct categories. A 2 log reduction, 1.5-2 log reduction, 1-1.5, 0.5-1, 0-0.5 and an increase in RNA over the treatment period.

We have two lines here, the top or pink line is for patients who started out with RNA values above the median. So these were patients who had higher baseline RNAs. The green line here is for the patients who had lower baseline RNAs.

What we can see here again is proportionality of effect between a better RNA response or a 2 log reduction on the far left, translated to a very rare incidence of clinical progression, and on up to patients with a very modest or no reduction having the highest risk of clinical progression. Then in between you can see roughly the proportionality of the effect.

Another point to recognize is that the patients

who start out with higher baseline values have a higher risk of clinical progression independent of the actual RNA response. I would just like to note that these bars here are the 95 percent confidence intervals for the estimates.

(Slide)

This is an analogous presentation showing again on the Y axis the progression incidence over 100 patient years of exposure, and on the X axis the magnitude of the reductions with the same categories as I just discussed.

The metric we are looking at now is an 8-28 week change from baseline and, once again, that is stratified by baseline RNA. So the pink line is for patients having a higher baseline; the green line is for patients having a lower baseline.

We can also see using this metric the proportionality of effect. Patients having a better RNA response measured by this metric have a lower incidence of progression. Patients having a worse response have a higher incidence of progression.

I would just note that the uncertainty here, the stratum having a 2 log reduction, is reflected in the width of this confidence band.

(Slide)

The next metric we are going to be looking at is

the actual nadirs. So this is the lowest value achieved on treatment. Once again, that is stratified by baseline RNA. The Y axis again is number of events per 100 patient years. On the X axis I have now used the categories of actual response. To the far left you see less than 400 copies/mL. The next group is 400 to 5,000 and 5,000 to 20,000 and greater than 20,000. Once again, we have two lines here, the pink for patients with higher baseline values and the green for patients with lower baseline RNA values.

Once again here we can see the proportionality of the effect between the actual level achieved and the incidence of clinical progression. I would also like to say that in this plot it appears that the baseline value is not significant looking at it this way, but when we do this type of analysis in a multiple regression setting, using the actual baseline values and the actual nadirs, the baseline value is still significant.

(Slide)

This is the same plot as the previous plot but we pooled data over the patients who have lower and higher baseline RNAs. There are two differences here. The Y axis now is the clinical progression rate and the X axis is 10 categories of nadir achieved instead of the 4 that we had in the previous plot.

What you can see here from this presentation is this proportionality of effect, linear effect relationship between the actual nadir achieved and the risk of clinical progression, with patients achieving the lowest values at lowest risk and patients achieving the highest values at the highest risk, and in between this proportionality of effect.

Before I move on I would like to say that these analyses have been done looking at Kaplan-Meier estimate of progression rate and also Cox model, showing that these results are fairly consistent.

(Slide)

Now I would like to turn to looking at the joint effect of CD4 and RNA, looking at the relationship between these two variables and the risk for incidence of clinical progression. This plot shows the correlation between 8-52 week CD4 count and RNA and clinical disease progression.

The X axis is RNA response in terms of the log scale and the Y axis is the CD4. The reference lines on this plot represent a value of about 3.7 log-10, which is about 5,000 copies, and 200 CD4 counts on the Y axis. The blue dots are patients who did not progress on the study and the red squares are patients who progressed on the study.

What we can see looking at it this way are a couple of points. We know that the highest clinical

progression incidence occurs when patients have RNA values greater than 5,000 and CD4 counts less than 200, in particular, the rate is about 28 percent. When the CD4 count is above 200 and the RNA value is less than 5,000 clinical progression is relatively greater. We can also see a slight correlation here between CD4 and RNA but, nevertheless, a wide variability. For patients with a given RNA value, they have a range of CD4 values.

(Slide)

This is a similar presentation to the previous one. What I have looked at here is the 8-28 week mean changes from baseline and now I am looking at subsequent clinical disease progression. So this is clinical disease progression after 28 weeks and we are trying to look at the temporal relationship between the changes in CD4 and RNA and the subsequent risk of clinical progression.

The X axis ranges from minus 3 to 1.5 and the reference line is at minus 1. So this is patients who have had a 1 log reduction. The CD4 axis runs from minus 200 to 300, with the reference line drawn at 50 CD4 cell rise.

What we can see here is that patients who did not achieve 1 log reduction or 50 cell CD4 rise in the 8-28 week period were at the highest risk of clinical progression and that patients who achieved such a CD4 and RNA response were

at the lowest risk of clinical progression. Once again here we see a slight correlation, about minus 0.5, between CD4 response and RNA response.

(Slide)

Now I would like to look at the effects of baseline CD4 and baseline RNA and CD4 response and RNA response. What we are doing here is using a Cox multiple regression model to predict subsequent clinical disease progression based on an 8-28 week metric. So this is a mean change from baseline at 8-28 weeks. I have fitted this Cox model, calculated the hazards ratios and then from those obtained the percent reduction in risk of clinical progression. These estimates are for 50 cell CD4 increase or 1 log reduction. In other words, the hazards ratio for baseline CD4 count was about 0.5 so that corresponds to about a 50 percent reduction in risk. The confidence intervals are noted here. You can see that the fact that they don't overlap indicates that all these variables are highly statistically significant, particularly the effects of baseline CD4 and RNA of about 50 and 60 percent reduction in risk, or 50 cell increase in CD4 or 1 log reduction in RNA. Then for the CD4 and RNA response we have a 60 to 70 percent reduction in risk of subsequent clinical disease progression. It is important to note that we looked at the

interaction between CD4 and RNA and this was not significant.

I would now like to spend a couple of minutes looking at RNA and CD4 as surrogate markers for clinical disease progression to new a new AIDS event or death in the CAESAR trail, the clinical endpoint trial. What I am going to be showing you is the concordancy between the treatment effects on the CD4 response and the RNA response and the clinical response as measured by progression to new AIDS event or death. Then what I would like to show you is what happens to the treatment effect in the clinical progression if you remove the treatment effects on the CD4 and RNA responses.

(Slide)

The objectives of such a surrogacy analysis, the main objective of this type of analysis is to answer the question of whether or not the effect of antiretrovirals on delayed clinical disease progression is actually mediated by the antiretroviral therapy on immunologic and virologic endpoints as measured by CD4 and RNA.

I just want to note that there are two other methods in looking at surrogacy for CD4 and RNA. One of these is looking to see whether or not treatment regimens which confer immunologic and virologic benefit compared to

control regimens also confer clinical benefit compared to these regimens. Then, whether or not the converse is true. In other words, do treatment regimens which do not confer immunologic and virologic benefit compared to control regimens also do not confer clinical benefit compared to these control regimens?

I just want to note that we are conducting some collaborative work with other sponsors and ACTG to look at this and that work is still in progress.

(Slide)

I am going to show you now the RNA responses for the two treatment arms in the CAESAR trial, the placebo arm here in the pink, and the 3TC arm in the green. What this plot shows is a median change from baseline in log RNA for each treatment arm, and this is the time on study.

So we can see that the current therapy plus placebo arm remains essentially unchanged, flat throughout the treatment period for about one year, as opposed to the 3TC arm which shows a sharp reduction, a 1.5 log drop, and then a gradual return to baseline but, nevertheless, a sustained 0.5 log reduction out to about a year of study.

The treatment comparisons of the 3TC arm with the placebo arm here, they are all highly significant. We looked at different metrics of response, shorter term as

well as more median-term responses, using simple metrics such as the mean.

(Slide)

The next thing to look at then is the treatment effect on the CD4 count, and here we see the concordancy between the treatment effect on RNA and the treatment effect on the CD4 count. What we see here in the placebo arm, and this is time on study. This is CD4 count median change from baseline. There is an actual reduction in the placebo arm, current therapy plus placebo, versus about a 35 cell rise in the 3TC arm. Then this is followed by a return to baseline but, nevertheless, we see about a 35 cell difference throughout the treatment period.

(Slide)

This is the Kaplan-Meier of estimates of AIDS-free survival. What we have here again is the 3TC arm in the green. I know it is probably difficult to see. In the red line is the placebo arm. We can see here a very highly statistically significant relationship in that adding 3TC was beneficial to placebo with respect to clinical disease progression.

I would like to note that the time scale for the X axis is from 0-32 weeks because on the next overhead we are going to be looking at a metric of 12-20 weeks for RAN

response and seeing how well that explains this observed treatment difference during the 20-32 week time frame.

(Slide)

When we do that, what we do is we fit a Cox proportional hazards model, and we cover the baseline survival function and then use that to obtain the treatment specific survival curves. These are like Kaplan-Meier curves. What we want to demonstrate is that after we account for the fact that the treatment arms are different with respect to CD4 and RNA. If, in fact, they were the same, what would the difference in the clinical progression rates be?

When we do that, we see that the treatment effect is clearly non-significant and that the adjusted Kaplan-Meier curves are essentially superimposing.

(Slide)

In conclusion, I would like to reiterate that treatment-induced reductions in RNA reduce the risk of clinical disease progression. That CD4 adds independent prognostic information to RNA on the risk of clinical disease progression.

Furthermore, the prognostic significance of RNA does not depend on baseline CD4 or CD4 response, and that disease progression is rare for patients with very low RNA

and is most common for patients with very high RNA.

Finally, delayed progression to new AIDS or death caused by antiretroviral therapy in the CAESAR trial is actually mediated by the antiretroviral effect on CD4 and RNA.

DR. HAMMER: Thank you very much. Unless there are pressing clarification questions, we will hold them until the general questions. The next speaker is Christy Chuang-Stein, from Pharmacia and Upjohn.

Presentation, Christy Chuang-Stein, Ph.D.

DR. CHUANG-STEIN: Well, I hope everybody is still awake. Good afternoon, everyone.

(Slide)

During the next twenty minutes I will share with you results from Pharmacia's and Upjohn's effort with the Food and Drug Administration to understand the role of viral- load reduction in evaluating AIDS treatment.

(Slide)

The data that formed the basis of my presentation came from two large trials conducted by Pharmacia and Upjohn. Over 2,300 patients were enrolled into these two trials, and the data of the two studies were combined for this exercise.

The baseline CD4 count for the combined analysis

population was around 230 and the geometric mean of the baseline HIV-1 RNA was around 75,000 copies. The median follow-up duration was over 1 year, and the longest follow-up duration exceeded 2 years.

Even though the treatment regimens used in these two trials have become obsolete by today's treatment standards, the data collected in these two trials, nevertheless, has offered a wealth of information on the relationship, or the lack of it, between the viral-load reduction and the risk of clinical progression.

This morning we heard two presentations on the RNA PCR assay as well as the assay's characteristics. Because of the information shared with us by Dr. Iacona-Connors and Dr. Brambilla, I will only mention very briefly here the results from Pharmacia's and Upjohn's own efforts to examine the combined HIV-RNA assay variability and the short-term biologic variation in HIV-1 RNA.

For the two trials here, Pharmacia and Upjohn Company used and RNA PCR assay developed by its own clinical research laboratories. This RNA PCR has been thoroughly validated against the Roche Amplicor assay, and was found to produce results highly correlated with those produced by the Roche assay, with a correlation coefficient around 0.93. All the RNA PCR values included in my presentation have been

converted to their Roche equivalent.

These two trials offered a great opportunity to examine the combined HIV-RNA assay variability and the short-term biologic variation in HIV-RNA because two pretreatment RNA measurements were taken 14 days of each other after each patient was adequately washed out of their current antiretroviral therapy.

(Slide)

Applying the tolerance limits technique to the difference between the two pretreatment RNA measurements, we concluded that a viral-load reduction of 0.5 logs --

(Slide)

--or more was beyond the combined HIV-RNA PCR assay, as well as the short-term biologic variation in HIV-RNA.

(Slide)

We will next concentrate on the prognostic value of baseline CD4 count, baseline HIV-RNA, as well as the change in RNA for the subsequent risk of clinical progression. For convenience I will use the term clinical progression to include death.

There are several metrics that one can use to characterize the change in RNA or the RNA response. The metric we chose was the peak reduction in RNA during the

first 12 weeks of treatment, with the peak reduction expressed on a log 10 scale. To evaluate the merit of using this peak reduction during the first 12 weeks of treatment, we restricted the analysis to those individuals who were in the studies for at least 12 weeks and did not experience any clinical progression during these 12 weeks.

Two measures were used to quantify the risk of clinical progression. The first was simply the proportion of individuals who experienced clinical progression. The second one was based on the incidence rate of clinical progression, defined as the number of individuals experiencing clinical progression with 10,000 days of follow up.

(Slide)

We used the FDA's suggestion to classify the baseline HIV-1 RNA into 5 categories ranging from less than 20,000 copies to greater than half million copies. As for the peak reduction, we first classified the peak reduction into 3 categories, greater than 1 log, between 0.5 and 1 log and less than 0.5 log. This classification was based on an earlier observation that a 0.5 log or greater reduction represented a real RNA response in our trials. Our second classification was based on the quintile of the peak response of the analysis population. The second

classification was used to decide if the risk of clinical progression was a monotonal function of the RNA reduction in the first 12 weeks.

These two graphs show the risk of clinical progression for the 3 reduction categories. The X axis here represents the baseline HIV-1 RNA on the original scale. This first graph, here, shows a greater risk of clinical progression for smaller reduction in the HIV-RNA during the first 12 weeks. This relationship was especially clear for higher baseline HIV-1 RNA values.

(Slide)

An almost identical pattern was observed when we expressed the risk in terms of the incidence rate of clinical progression for 10,000 days of follow up.

(Slide)

The next two transparencies show the risk for the 5 reduction categories. Except for a few instances where things get a little bit switched, the inverse relationship between the RNA reduction and the risk of clinical progression was apparent in this graph. In addition, the risk of clinical progression continued to decline with higher baseline HIV-1 RNA values. Furthermore, these findings were independent of the measures used to quantify the risk of clinical progression.

(Slide)

We have also looked at the role of baseline CD4 count and the risk of clinical progression. In order to do that, we fit a proportional hazards model using the 5 reduction categories as well as the 2 baseline marker values as predictors. Recall that the 5 reduction categories were determined using quintiles. Therefore, there were about the same number of patients in each of these 5 categories here. In addition, the baseline CD4 count, which was treated as a continuous variable, was expressed in a unit of 25 cells and the baseline HIV-1 RNA was expressed as a unit of 0.5 log.

Our analysis showed that both baseline CD4 as well as baseline HIV-1 RNA were highly correlated with clinical progression. Under the fitted model, the model suggested that a higher baseline CD4 count in the amount of 25 cells was associated with a 15 percent less risk of clinical progression. On the other hand, a higher baseline HIV-1 RNA value in the amount of 0.5 log was associated with a 65 percent increase in the risk of clinical progression. The estimated hazards ratio of the various reduction categories relative to the first one decreased monotonally as the amount of reduction increase.

I would like to point out here that in including these 5 categories in the model we treat this one, the

category with the least amount of reduction, as the reference group. Therefore, the risk was measured against the reference group. That is where the hazards ratio came from. Therefore, the hazards ratio for the reference group was set equal to 1.

(Slide)

So far we have shown you results from classifying the baseline HIV-1 RNA according to the suggestion of the FDA. We have also conducted an analysis using the baseline RNA classification suggested by the Surrogate Marker Collaborative Group.

The baseline RNA categories suggested by the SMCG ranged from less than 5,000 to greater than 200,000 copies. Here I have shown you the calculated risk of clinical progression for each of the 5 baseline HIV-1 RNA categories. You will notice that the distribution of patients into these 5 categories is not as smooth or not as even as that among the 5 categories recommended by the FDA.

The striking finding here is that none of the 123 patients in our analysis population who started the trials with a baseline HIV-1 RNA less than 5,000 experienced any clinical progression during the trial's period.

(Slide)

In addition to this observation, our earlier

estimate of the hazards ratio also suggested that an 0.5 log reduction in the HIV-1 RNA during the first 12 weeks was associated with a 38 percent less risk in the risk of clinical progression. The corresponding figure of 1 log reduction during the first 12 weeks was 57 percent. Also, the clinical benefit of a viral-load reduction beyond 1.3 logs appears to flatten out in our trials.

(Slide)

In addition to looking at the peak reductions during the first 12 weeks, we also looked at the duration of virologic response to see whether this duration has any prognostic value for subsequent risk of clinical progression.

For this analysis we included only those patients who were in the studies for at least 24 weeks and did not experience any clinical progression during the first 24 weeks. Also, in order to conduct this analysis we needed a definition for virologic response. Based on our earlier observation of what constituted a true RNA response, we defined for our analysis for the data from our trials a virologic response by at least 0.5 log reduction in HIV-1 RNA from the baseline value. For those individuals who did achieve a response during the first 24 weeks, we defined virologic failure for our analysis a viral-load rebound to

be within 0.5 log from the corresponding baseline.

For this analysis, when we looked at the duration we looked at the 24-week period, and the response duration during this 24-week period is then calculated as to the period between the time when the virologic response was first observed and the time when the virologic failure was first declared.

For those individuals who did not fail by week 24, we truncated their response at week 24 and calculated the response duration accordingly. This convention, indeed, did not differentiate between virologic failure from the continued response at week 24. However, since more than 95 percent of the responders responded by week 12, a short response duration, such as less than a 8 weeks, implied a true virologic failure at week 24. Therefore, the results from this analysis can be best interpreted who never responded and those who had very short response duration and those who had a long response duration.

(Slide)

We divided the response duration into 5 categories, again by the quintile of the response duration distribution. Because the plots based on the true risk for the two measures for risk of clinical progression are extremely similar, I will only show you here the one based

on the incidence rate of clinical progression.

The curve for the non-responders is above all the remaining curves. This might be a little hard to see. In order to better differentiate among the four middle response duration categories, I added green to the four corresponding curves. In addition, I added blue to the curve pertaining to the longest response duration group.

As can be seen from this graph, there is a trend for a decrease in the risk of clinical progression or an increase in the responders' duration.

(Slide)

The estimated hazards ratio for the various response categories relative to the non-responders using the proportionate hazards model are given on this transparency. The pattern among the estimated hazards ratio confirms the trend observed earlier, and there is a trend of decreasing risk with a longer response duration, consistent with our earlier observation on the relationship between the baseline CD$ and the baseline HIV-1 RNA with clinical progression. This analysis only confirms that highly significant association between the baseline marker values with the clinical progression.

(Slide)

What are the implications of our findings for

study designs of trials to evaluate the HIV treatment? Obviously, it is important to consider stratifying randomization by the baseline CD4 count and baseline HIV-1 RNA values because of their strong association with clinical progression.

Our second point here is really a question for the members of the Advisory Committee as well as the experts in the AIDS arena. Remember, in our trials we did not observe any clinical progression in those individuals in our analysis population who started the trial with an HIV-RNA less than 5,000 copies.

In view of that observation, how can we conduct the benefit against risk of the highly active antiretroviral therapies in patients with very low viral load while the clinical benefit of HAART might not be realized for a long time? The risk of the HAART can be felt acutely through treatment side effects and drug toxicity. We don't have an answer to this question. We would simply like to bring the issue up for the community to consider.

Finally, considering the relationship between response duration and the risk of clinical progression identified in our analysis, we feel it is important that the trials be long enough to capture information on the response duration. In our opinion, trials to evaluate a regimen's

ability to suppress viral load should be at least 24 weeks long.

This concludes the Pharmacia and Upjohn presentation this afternoon. We would like to thank the Agency for the opportunity to participate in this important project and the chance to share the important scientific knowledge learned from Pharmacia's and Upjohn's effort to combat HIV infection. Thank you.

DR. HAMMER: Thank you very much. It is not scheduled but I am going to take the prerogative to have a 10- or 15-minute break now. Then we will return promptly and finish up the afternoon's presentations.

[Brief recess]

DR. HAMMER: Please take your seats. We are going to continue the pharmaceutical sponsor presentations. The next speaker is Lesley Struthers, Hoffman La Roche.

Presentation, Lesley Struthers

(Slide)

MS. STRUTHERS: My name is Lesley Struthers. On behalf of Hoffman La Roche, I will present the data analysis we performed exploring the relationship between RNA regimens and the time to first AIDS-defining event or death. I will present data on study design and baseline information.

We looked at various definitions in defining a

virologic responder and how long this definition identifies those patients who later go on to have AIDS-defining events or die. We explored the impact of duration of RNA effect, and I will also show you the importance of CD4 as a surrogate marker.

(Slide)

All analyses presented here are taken from one large study, ND-12456, U.S. study. This was a double-blind, randomized study involving 940 patients in the intent-to-treat analysis. The patients who had experienced AZT but were naive to protease inhibitors, ddC and ddI. There were three treatment arms, ddC, saquinavir and the combination of these two treatments.

The primary endpoint was time to first AIDS-defining event or death. There were 223 endpoints, not 243 as on this slide. These were patients who were followed for a median time of 17 months. Both RNA, using the Roche Amplicor kit, and CD4 were measured every 4 weeks up to 16 weeks, and then every 8 weeks up to 80 weeks. So you can see this is a large study with a considerable amount of information.

(Slide)

The distribution of baseline CD4 and RNA are represented here in the pie graphs. A large proportion of

patients, 70 percent, had a baseline CD4 between 100-300. The median baseline was approximately 170 across all 3 treatment arms.

For RNA, the patients were split to 26 percent having RNA values below 50,000, 36 percent of patients with a value between 50,000 and 200,000 and 39 percent of patients with a value greater than 200,000. The median baseline across the 3 treatment arms was approximately 5.1 logs.

(Slide)

This graph shows the absolute levels of CD4 and RNA together, looking at the area under the curve over the 48 weeks, with RNA along the X axis and CD4 along the Y axis. The red circles show all patients who were alive and did not have an AIDS-defining event. The squares indicate the patients who either died or had an adverse AIDS-defining event.

It is clear from this graph that the majority of patients with an AIDS-defining event who had died, the yellow squares, had CD4 values of less than 200 and RNA values greater than 10,000.

Using these cutoff values, we can also see that a large number of red circles are in this lower right-hand quadrant, meaning we are also falsely identifying these

patients.

(Slide)

One of our first objectives was to see whether there is a virologic response cutoff level identifying patients who would later have an AIDS-defining event or die. Virologic response is defined as an absolute level of RNA which the patient had to reach in the first 24 weeks before they could be classified as a responder.

In this graph we looked at two cutoff values, 10,000 and 100,000. The bottom line is the time to an AIDS-defining event or death for those patients whose absolute RNA level remains above 100,000 during the first 24 weeks. The patients in the top line have at least one value of less than 10,000 in the first 24 weeks, and the middle line containing the remaining patients.

This clearly shows that an absolute RNA cutoff level in the first 24 weeks has an effect on time to first AIDS-defining event or death.

(Slide)

In addition to looking at the 10,000 and 100,000, we also looked at different RNA cutoff levels. In fact, we looked at 15,000, 20,000, 30,000, 40,000 and 50,000. In the graph we tried to summarize our information.

(Slide)

These cutoff levels are displayed along the X axis. We looked at the effect of having 1 response value in the first 24 weeks. This is the blue line. We looked at 2 consecutive response values, the pink line, and then 3 consecutive values, the yellow line. The Cox model was used here to calculate the relative risk and includes baseline CD4 and RNA and treatment, where the RNA cutoff value defined the patient as a responder or a non-responder.

The relative risk, displayed along the Y axis, is the ratio of the hazards of progression to AIDS-defining event or death for responders and patients who did not respond. The greater the difference from a relative risk of 1, the greater the difference between responders and non-responders. All levels of RNA on this graph are important with the relative risk between 0.5 and 0.3, and all of them clearly very different from 1. A cutoff value of 10,000 or above is equally as effective in predicting AIDS-defining event or death as the higher cutoff values are.

(Slide)

We also looked at the effect of change from baseline over the first 24 weeks. Here, the patients are split into 3 groups according to whether they had greater than 1 log decrease in RNA. This is the top line. Between a 0.5 and 1 log decrease, the middle line, and less than a

0.5 log decrease, the bottom line.

The clear difference in lines indicates the smaller the change from baseline, the greater the probability of suffering an AIDS-defining event or death. Interestingly, this metric did not split the patients as widely as when we looked at the virologic response definition as absolute difference in RNA of 100,000.

(Slide)

Next we looked at the additional effect of baseline CD4 and RNA on time to first AIDS-defining event or death after the RNA responses were taken into account.

(Slide)

Here, when looking at patients who had a value of RNA less than 100,000 during the first 24 weeks, and then switching these patients by their CD4 median baseline value, there is a clear effect of CD4 baseline on the first AIDS-defining event or death.

(Slide)

Now we look at the effect of baseline RNA on the percentage of patients who progressed to the first AIDS-defining event or death. The baseline is split in quartiles so each of these 4 arms has the same number of patients in it. The change from baseline in the first 24 weeks is also split into quartiles.

This 3D plot demonstrates that the smaller the change in RNA over 24 weeks, the higher the block. So more patients have progressed. This trend is demonstrated over the different baseline levels. It is also clear that the baseline RNA has an effect and the higher the baseline, the higher the block, and so the greater the likelihood of progression.

(Slide)

One of the objectives was also to examine the duration. The area under the curve, AUC, takes into account the level of RNA as well as the duration. The patient who has a low RNA and maintains this will have a low AUC, whereas, a patient whose RNA drops initially and then increases will have a higher AUC. This graph demonstrates that the metric AUC for the first 24 weeks, when split by quartiles, is clearly a nice, strong prognostic indicator.

In the next graph we are going to show you the importance of following RNA throughout the study and using a virologic failure definition.

(Slide)

We used virologic failure cutoff levels of 5,000, 10,000, 15,000, 20,000, 30,000, 40,000, 50,000 and 100,000. These are displayed along the X axis in this graph. As soon as the patient has 1 RNA level above our cutoff used in this

virologic failure definition, they are classified as a failure, the blue line, at the time point that the failure occurred. We have also examined the effect if patients have 2 consecutive RNA values as failures, the pink line, and 3 consecutive failures, the yellow line.

The Cox model used here, in this analysis, includes baseline RNA, CD4 and treatment, with the RNA value defined as failure or not a failure being used in the model as a time-dependent variate. So this means that we don't just look at the RNA values during the first 24 weeks; we look at them through the life of the study.

The relative risk on the Y axis is the ratio of Cox hazards regression based upon an AIDS-defining event or death. A failure is compared to patients who do not fail. Again, the further the relative risk from 1, the greater the difference between failures and non-failures.

Here, the relative risk demonstrates that with cutoff levels of 10,000 or above the patients defined as failures are more likely to progress than patients who have not failed. The risk also increases as we raise the cutoff level along the X axis.

If we just look at 2 examples on this graph, and look at the pink line, so those patients who have 2 consecutive measures that count as a virologic failure,

using a cutoff of 10,000 the relative risk is 1.5, which means patients classified as failures have a 50 percent increase in having an AIDS-defining event or dying compared to those whose RNA stays below 10,000. With a cutoff level of 100,000, the relative risk is 2. So patients classified as failures have a 100 percent increase in having an AIDS-defining event or dying compared to those who remained below 100,000.

(Slide)

This slide looks at the effect of CD4 during the study, as well as the effect of RNA. Here the data is split into 3 groups based on likely change of response in CD4 and RNA, using the area under the curve minus the baseline over the 24 weeks compared to their respective means. The plus indicates patients who were above the mean. This means they have a high CD4 or high RNA. The minus indicates patients below the mean.

The top line shows that patients who have a high CD4 and a low RNA do considerably better compared to patients who have a low CD4 and a high RNA.

We have seen from the data so far that both CD4 and RNA are important prognostic factors, and that baseline CD4 and RNA provide additional prognostic information.

I will now go through surrogate marker analysis

that we used to investigate how effective our surrogate markers are using the Prentice criteria.

(Slide)

For this analysis 2 treatment arms are used, ddC and saquinavir plus ddC.

(Slide)

The Prentice criteria states that first we need to have a clear treatment effect in our time to AIDS-defining event or death, and we clearly demonstrate this between the 2 arms, with a significant difference, a p value of less than 0.0001.

(Slide)

The Prentice criteria also states that we need to have a clear treatment effect in our surrogate markers. Here, looking at RNA, we have a clear difference between the 2 treatment arms, looking at the area under the curve minus the baseline over 24 weeks.

(Slide)

With CD4 we also see a clear difference over the first 24 weeks in the area under the curve minus the baseline.

(Slide)

Given that we have a clinical difference and surrogate marker differences, we used a Cox model initiated

by the baseline CD4 and RNA in the model, which are both highly significant, with p values of less than 0.01. We can see that this is significant.

(Slide)

Next we add both RNA and CD4, area under the curve minus the baseline. We can see that both baselines and the area under the curve minus the baseline relative to CD4 and RNA are significant. We can also see that the treatment effect is no longer significant. This means that the surrogate markers are explaining some of the treatment effect.

(Slide)

Pooling data together from several models, the AUCMB RNA by itself explains 35 percent of the treatment effect; 49 percent of the effect is explained by CD4 AUCMB in the 24 weeks. Together, as in this model, they both explain 61 percent of the treatment effect.

(Slide)

Here I show this graphically. Here we have the model which is baseline CD4 and RNA values and we can see that there is a treatment effect there.

(Slide)

Once we put AUCMB RNA and CD4 in the model, we can see that the tap is closed, the yellow lines compared to the

blue lines.

(Slide)

In summary, our conclusion from this analysis is that RNA is a very strong prognostic indicator. It appears that the level of RNA may be more important than change from baseline. The duration of RNA when we look at the area under the curve indicates that duration of effect is important. From this data, based on experienced patients, we also clearly show that CD4 is equally as important as RNA. Thank you.

DR. HAMMER: Thank you very much. The next speaker is Margo Heath-Chizzoi, from Abbott Laboratories.

Presentation, Margo Heath-Chizzoi, M.D.

(Slide)

DR. HEATH-CHIZZOI: I am Margo Heath-Chizzoi, and on behalf of Abbott Laboratories I would like to thank the FDA for inviting us to present the correlations of HIV changes with clinical benefit demonstrated in the Abbott study M94-247.

(Slide)

The M94-247 study evaluated clinical benefit in patients with advanced HIV illness. At baseline the patients had to have CD4 cell count less than 100 cells/mL^3, at least 9 months of prior approved antiretroviral therapy.

The patients were allowed to continue up to 2 but not more than 2 concurrent, approved agents during the time the study was conducted. It included only nucleoside agents. They had to have a Karnofsky score of greater than 70, and they needed to not have active opportunistic infections requiring induction of therapy. Maintenance therapy and secondary prophylaxis agents were included in the study.

(Slide)

The study enrolled 1,090 patients who were randomized to either ritonavir 600 mg b.i.d. to their chronic regimen, and that arm included 543 patients, or placebo added to their current regimen in 547 patients.

The primary objective of the study was to evaluate clinical endpoints, which were defined as new AIDS-defining illness, with the exception that recurrent PCP, Candida esophagitis or prolonged mucocutaneous herpes were allowed as recurrent events, and death was included as a primary endpoint if patients didn't have another new disease diagnosed prior to death.

A surrogate marker sub-study was nested in this study for the first 16 weeks of evaluation. During these 16 weeks patients were asked to continue the current antiretroviral regimen that they studied the study on, with the exception that they could stop an agent if they had

toxicity during the first 16 weeks. There were 80 ritonavir patients and 79 placebo patients, and those were the first patients who were involved in the study with a baseline RNA better than 15,000 copies/mL.

During the time that it took us to identify these virology subset patients, patients who enrolled in the study were also analyzed for CD4 and CD8 changes. The enrollment for the entire study occurred during May and June of 1995, and the 191 events that were assumed to be required to have an 8 percent power to detect a one-third reduction in events between the 2 arms were approved and analyzed by the middle of December of 1995, which showed a highly statistical significant benefit both in disease progression and survival between the two arms. That allowed us to offer open-label ritonavir to all patients in early January, 1996.

Since these were the first patients enrolled in the study, they overall had about 7 months of evaluation during the placebo-controlled period. The remainder of this presentation will focus only on the 80 ritonavir patients in the virology subset, looking at changes at their HIV-RNA and how it correlated with clinical benefit over the 7-month observation.

(Slide)

The baseline demographics of the virology subset

are similar to the overall patient group. They were primarily men, 94 percent men; 89 percent Caucasian, with an average age around 39 years. They were more than 5 years from their diagnosis of HIV infection, the majority of the patients and, unfortunately, 88 and 15 belong over here where the majority of patients acquired HIV by sexual transmission.

(Slide)

These, indeed, were advanced patients. Here are their mean and median baseline RNA numbers. The baseline RNA median was 5.4 logs, with a median CD4 cell count of 20.8 cells and a median CD8 of 411 cells. The concurrent antiretroviral agents taken by this group reflected sort of the standard care in the mid-1995's range. There was a large group of patients who actually had no concurrent regimens; a fair number of patients who were taking either long-term therapy AZT or D14; a small number on ddC. There was one patient on monotherapy, ddI, and the other patients were on ddI combinations, and a fair mixture of AZT dual combination regimens, and different regimens were also included.

(Slide)

Nineteen patients had a clinical endpoint at 7 months of evaluation, and similar to the larger study

population, the diagnoses spanned the spectrum of AIDS-defining illnesses, including 4 patients who died prior to the diagnosis or a new event.

(Slide)

The first correlation we looked at looking at RNA changes with the clinical endpoints, we looked at early changes in the magnitude of HIV-RNA increase at 12 weeks or 16 weeks, stratified by change into either 0.5 log or a log at those time points.

As you can see, the 0.5 log cuts at both 12 and 16 weeks had a more balanced distribution of patients than the greater or less than of 1 log cuts at both time points, and using the Fisher's exact test and proportional hazards model we were unable to identify any statistically significant difference in the events rates in these 2 groups.

(Slide)

Here the Kaplan-Meier shows the trend towards a difference when you took changes of less than a log, in yellow, compared to greater than a log at the 16-week time point. Again, there is a fairly small number of patients in this greater than a log decrease which may be damping our ability to detect a statistically significant difference between the groups.

(Slide)

The next analysis looked at maximal suppression at nadir, and here you see more splaying of the groups when you take a maximal suppression of less than a log at any time point, here in green, compared to 1-2 logs, in yellow, or further than a 2-log decline nadir, here in blue. You can see that you do begin to get a separation of the curves even in this early small sample size.

(Slide)

To explore that further, we looked at thresholds for absolute RNA values at the nadir, looking at the level of detection of the assay for the Roche Amplicor, using 200 copies as our level of detection, a threshold of above and below 1,000 copies or a threshold of above or below 5,000 copies.

Here, you can see that using the level of the 5,000 copy threshold we got a more even distribution of patients than at either of the 2 lower thresholds. This above and below 5,000 copies did lead to statistically significant changes in the comparisons between the 2 groups. There was a trend toward clinical benefit in the less than a 1,000 compared to greater than 1,000 analysis, but the statistical significance between the groups is sort of dampened by the unequal distribution of patients. When you got down to using a threshold of the level of detection of

the assay, we got even more imbalance in the patients and we lost our ability to detect a difference between the groups.

(Slide)

To assess the influence of this a little bit more closely we plotted a Kaplan-Meier analysis. As you can see, the group that had a nadir of greater than 5,000 appeared to develop clinical events early compared to the group whose nadir was less than 5,000. That appeared to be sustained through the 7 months of evaluation.

(Slide)

To assess the influence of baseline characteristics on that observation, we looked at grouping the patients by a baseline RNA of less than 300,000 copies compared to greater than 300,000 copies, this being a round number fairly close to a median number for the overall patient group. In the patients that had less than 300,000 copies there are actually more patients with fewer overall clinical events and we couldn't detect statistical significance in that comparison.

However, when you go to patients with a baseline greater than 300,000 copies, despite having a slightly smaller overall number of patients, you do have more clinical events that separated nicely or had the characterization of a nadir of less than 500,000 compared to

a nadir of greater than 500,000, shown here in green. This really isn't surprising from the clinical standpoint since these patients who had very high RNA levels at baseline and didn't really see an appreciable change in their RNA would be the ones you would expect to have the highest rate of disease progression in the short observation period.

(Slide)

In a similar analysis we looked at the impact of baseline CD4 cells above and below the median of 20 on a similar split between nadirs of less than 5,000 and greater than 5,000. Again, it was the group with the most advanced disease, baseline CD4s less than 20, that had the most apparent difference in the nadir of 5,000 compared to a nadir of greater than 5,000. And patients with greater than 20 CD4 cells counts at baseline really didn't show that much difference between this nadir cut of above and below 5,000.

(Slide)

An additional analysis was conducted to look at the impact of duration of suppression on clinical benefit. Here patients were stratified by time to rebound from their nadir of less than 85 days compared to greater than 85 days, which was very close to the median in rebound from nadir for the overall patient group.

This was using a fairly stringent definition of

nadir and rebound where we required the patients to have at least 0.6 logs decline in order to be called a nadir, and they have to have 0.6 logs rebound. As you can see, using those stringent definitions, we have a very small sample size and couldn't detect a difference in the groups.

(Slide)

So within the limits of this data set on the 80 patients with 19 clinical endpoints observed for 7 months, we feel comfortable making the following conclusions: That a nadir decrease of greater than 2 logs during 16 weeks is associated with a greater clinical benefit than having a nadir decrease of less than 1 log, and that having an absolute nadir value of less than 5,000 copies is associated with clinical benefit compared to having a value greater than 5,000 copies.

This ability to replicate it in the subset of patients only with greater than 300,000 copies may well be a function of the advanced patient population of the study rather than something of virologic significance. We think the patients who have baselines less than 100,000 should also have clinical benefit with longer observation and a minimal cut of less than 5,000.

(Slide)

Basically, we don't feel that duration of

observation in this limited sample is enough to be able to make any statements about long-term benefit in the group. Taken altogether, we feel that these data do support the general conclusion that HIV-RNA should be decreased as much as possible for as long as possible to maximize clinical benefit. Thank you.

DR. HAMMER: Thank you very much. The final speaker is Michael Elashoff, who will provide some summary comments and then we will open this up for questions.

Summary, Michael Elashoff, Ph.D.

(Slide)

DR. ELASHOFF: I will be summarizing the company presentations and presenting some of our conclusions. In planning this meeting we tried to achieve some consistency in the format of presentations so that it would be easier to draw conclusions. At the same time, the trials had varying populations, regimens and sizes so no one analysis could be dictated for all presentations. The result is that each company examined their data in a somewhat different way. This allows for us to examine the relationship in RNA changes and clinical events in several different ways and judge how consistent this relationship may be.

(Slide)

The basic question the presentations addressed was

how an initial RNA response predicted the subsequent clinical events. The RNA response was measured in several different ways, change from baseline, nadir and durability of the effect. The area under the RNA curve was also looked at for durability and effect. Clinical events were also examined in several different ways, Kaplan-Meier curves and proportional hazards models, to address the association between RNA and clinical events.

(Slide)

Overall, all five presentations found evidence for an association between RNA changes and subsequent clinical events. This association was seen in all the characterizations of RNA changes and clinical event analyses shown in the previous slide. Together, the data provide compelling evidence for this relationship.

(Slide)

The presentations covered many clinical trials and drug regimens. Here I have summarized the studies and regimens analyzed.

(Slide)

As I have mentioned, the study populations varied across the spectrum of HIV disease. Subjects in the Abbott study RNA subset provide by far the most advanced, as evidenced by the median RNA and CD4. This allowed for the

small number of subjects to yield valuable information. The other larger trials primarily represented less advanced subjects, with a mixture of naive and experienced patients. Overall, the number of subjects analyzed was almost 5,500.

(Slide)

I will now go over some of the specific analyses from the presentations. One of the standard ways in which RNA responses are quantified is by the change from baseline, usually measured in log units.

(Slide)

The ACTG analysis found a strong relationship between change from baseline over the first 24 weeks. Shown here on the X axis are units of 0.5 log drop, and the clinical event relative risk is on the Y axis. The primary story of this graph is in the slope and shape of these lines, rather than the particular lines here which just represent different ways of analyzing the data. This relationship was seen after adjustments for baseline RNA, CD4, treatment and study.

(Slide)

The Glaxo analysis also identified a strong relationship. Here are two curves, one for high baseline RNA and one for low baseline RNA. Change from baseline RNA, again in units of 0.5 log decrease, is shown on the X axis

and clinical incidence rate is on the Y axis. The curves seem to flatten with large decreases, especially for those who started lower to begin with. This effect is in part due to the assay lower limit of detection. More follow up will be necessary to detect differences down in this range.

(Slide)

Pharmacia's results also showed association between change from baseline and clinical progression rate. Here we see that subjects who started with a low RNA experienced few events regardless of their change. But the relationship is more dramatic for higher baseline RNA values.

(Slide)

They also presented these results in a proportional hazards model, which indicated a dose-response type relationship between change from baseline and clinical event rates after adjustment for baseline CD4 and RNA values.

Notice here that above about 0.5 logs, no reduction was seen in the hazard rate. In the region above 0.5 logs, there was a significant reduction. The two categories here above 0.86 logs showed the most decrease in clinical event rate.

[Slide.]

Roche presented Kaplan-Meier curves stratified by the magnitude of change from baseline. The curves showed separation between the three categories indicating that greater decreases are associated with a lower risk of disease progression.

[Slide.]

Overall, the results indicated that even the smallest decrease studied across the trials, about 0.5 log, was associated with clinical benefit. Further, greater decreases resulted in lower clinical-event rates. The relationship for large decreases remains to be clarified where the limit of detection and small event rates make characterizations difficult.

[Slide.]

Another way of measuring RNA response is via the RNA nadir, or lowest value achieved. Since this is an absolute number, cross-study comparisons could be made more easily.

[Slide.]

Glaxo found, as might be expected, that lower RNA nadirs were associated with lower clinical progression rates. This was true for both high and low baseline RNA levels. In this graph, you can also see a flattening of the curve below about 5,000 copies. However, the small number

of events indicate that longer follow up will be necessary to clarify this range.

[Slide.]

Glaxo also showed the relationship between RNA nadir and clinical progression rate. They found an approximate linear relationship. The ACTG also found a similar result.

[Slide.]

Here are Kaplan-Meier curves stratified by the RNA nadir level achieved while on treatment and eventual clinical progression in Kaplan-Meier curve format. Again, lower RNA nadirs were associated with longer times to clinical events.

[Slide.]

Roche also presented their data in terms of hazard rates. This graph found lower hazards for smaller RNA values.

[Slide.]

Overall, the data showed a clear association between the lowest RNA value achieved and subsequent clinical-event rates. The RNA values achieved to date do not appear to have reached the so-called threshold effect beyond which further reduction would convey no further advantage. Since these studies were started some years ago,

the regimens were not, in general, optimal by today's standards.

This meant proportionately few subjects had responses down to the limit of these assays so that characterizing the response curve in that range remains imprecise. Longer follow up and better treatments will be needed to address this area.

[Slide.]

The final characterization or RNA response was in the durability of that response.

[Slide.]

Pharmacia showed that durability was associated with clinical-event rates. In this graph, individual lines represent differing durabilities of effect. For this analysis, effect was defined as a half-log reduction. For the higher baseline RNA values, the lines are seen to separate indicating that less durable responses were associated with higher clinical-event rates.

[Slide.]

This effect is easier to see in the proportional-hazards model. In this model, responses that persisted past 16 weeks or about 114 days resulted in the fewest events while less durable responses meant a higher rate of clinical events.

[Slide.]

Roche addressed the issue of durability using a area-under-the-curve, or AUC. The AUC is analagous to the product of duration times effect. When they broke down subjects into four groups by their AUC, a clear difference was seen in the Kaplan-Meier curves. But since AUC is a function of duration and amount of change, this analysis is an indirect measure of the association between durability and clinical-event rates.

[Slide.]

Durability of response and subsequent clinical-event rates were found to be associated in the Pharmacia analysis. This result was supported by Roche's analysis of the area-under-the-curve. Although we had hoped that the trials considered would have dealt more completely with durability, we should not be surprised at the limited information available. This is due to the age of the trials, the treatment regimens and the population studied.

Long-term durability--that is, longer than 24 weeks--was infrequent and remains to be addressed.

[Slide.]

Both Roche and Glaxo also provided surrogate-marker analyses. These approached the question in a somewhat different way from the other analyses by

incorporating treatment comparisons into the analysis. Initially, Roche found a significant treatment effect.

[Slide.]

After incorporating RNA and CD4 responses while on study, presumably induced by the treatment, they found that the treatment effect was no longer significant. In other words, the RNA and CD4 responses seemed to mediate the treatment's effect on clinical-event rates.

[Slide.]

Glaxo showed this visually for their data as the treatment difference in AIDS-free survival between treatment and control seen here--

[Slide.]

--was virtually gone when RNA and CD4 responses in the first part of the study were incorporated into the analysis.

[Slide.]

While these analyses were suggestive that much of the treatment effect is mediated by changes in RNA and CD4, it is important to point out that the goal of these talks was not to formally validate RNA as a surrogate. When drugs are approved for lowering RNA, the claim of clinical benefit will not follow automatically.

To claim clinical benefit, sponsors will have to

show exactly that. Clinical practice and treatment strategies are largely based on RNA. As Dr. Feigal stated in the introduction, we may be read to move past the question of surrogacy. Tomorrow, we will discuss in more detail, the design of clinical trials making use of long-term RNA changes. The role of clinical endpoints and CD4 in such trials will be considered further.

[Slide.]

In summary, both change from baseline and RNA nadir were found to be associated with clinical-event rates. The analyses presented, on concert with several other considerations, suggest that the RNA nadir may be preferable to change from baseline. This avoids the problem of where you start pointed out by Dr. Marchner.
It makes the inevitable comparisons across studies more interpretable and may avoid some of the problems associated with the quantitative analysis of mean changes.

Further, the assays seem more suited to qualitative interpretations such as RNA decreasing versus RNA increasing or RNA below the limit versus above the limit. The limit of detection is simply the lowest nadir possible. Overall, clinical progression seems to be more related to the absolute RNA level rather than the change from baseline.

[Slide.]

The primary focus of the talks was on the relationship between RNA changes and subsequent clinical events. While these trials are not the final story, particularly for long-term responses at very low viral levels, they provide a compelling evidence that the lower the level of RNA achieved during the study, the lower the risk of clinical progression.

More limited data suggested that the more durable response, the better the outcome. We expect that further information will emerge as more data become available. Finally, these conclusions were based on multiple studies, treatments and methods of analyses and covered over 5,000 subjects.

DR. HAMMER: Thank you very much.

Questions

We now have approximately a half an hour for questions from the committee. I would suggest that questions be directed to any of the afternoon speakers. The warmth in the room has quelled some of the enthusiasm.

DR. MATHEWS: I could address this to Dr. Elashoff, but some of the data that raises this question in my mind were from the Roche presentation dealing with nadir RNA values as an endpoint. The question is while RNA nadirs

seem to perform reasonably well prognostically, is there come confounding with baseline value and interpretation of the nadir?

In other words, the interpretation of an endpoint as a nadir really mixes the baseline value with the treatment effect since, if a drug is known to produce, say, on average, a 0.75 log drop from wherever you start at a certain point in time, that is going to determine the nadir value that is achievable in a given patient population.

DR. ELASHOFF: I don't think that baseline affects the analysis of the data more than change from baseline. I think just the opposite that baseline--I think that this confusion was also noted somewhat after the ACTG talk where, when you look at change from baseline, really you can't think about change from baseline numbers without knowing where you started whereas the RNA nadirs are a more absolute comparison.

Certainly, baseline is important. Nadir and baseline would be important. But if you say that a drug is a 0.7 log drug, that is only true for the population in which it was originally studied. That doesn't necessarily imply that a different population with a different baseline is going to get that same reduction.

DR. MATHEWS: I agree that that is true but it is

problem true of all the metrics that have been presented. Some of the analyses that were presented gave adjusted values in the use of the nadir metric, adjusting for baseline. But many of them did not. They were just unadjusted or crude effects, whether it was using hazard ratios or incidence rates.

I am just not convinced. In my own mind, I guess haven't processed all this data yet that the nadir value is as attractive as your conclusion suggests compared to all metrics. I guess I would argue, just as we have looked at in most of the other drugs I have seen come before this committee, for consistency of effect across whatever metric is used, and not relying on a single metric.

DR. ELASHOFF: I guess I would agree with that although when you have a more effective regimen where, say, you have two treatment arms both of which are getting you to very low levels, differences in change from baseline will only reflect differences from where they started and won't reflect a true treatment difference.

I would agree with you that if they are telling you different things, then that would warrant further exploration.

DR. VALENTINE: I think many of us have been functioning under the paradigm that if we can really

diminish viral replication then we will prevent the selection of resistance. Yet all of us have seen patients who became below "levels of detection" for a period of time of a number of months only to rebound back again.

All the data we have seen today, obviously, is using a lower limit of detection of the 400, 500 or 1000 range, even. I hope that tomorrow we see some data going down lower than 500. I would urge all the sponsors to go back to their frozen samples and repeat these assays. Data is emerging that there is still replication going on less than 400 and it is conceivable that the nadir measurement would be much more meaningful if it were down around 20 so that there would be a long durability of effect and less chance of selection that may be occurring between 20 and 400.

DR. ELASHOFF: Actually, two of the presentations tomorrow are going to address exactly that comparing nadirs in the 400 to 500 range with values down as low as 20 and, actually, a dramatic difference was seen.

DR. VERTER: I must confess that I am a bit apprehensive about making the following comments.

DR. HAMMER: That's okay.

DR. VERTER: I feel a little bit like the little kid with his finger in the dike. But I am going to raise

some cautionary here. First, I would like to thank all five presenters for the analysis. I must confess that it was almost overwhelming. I was writing notes on the first one when the third one, I think, was starting to do their presentation. So I am sure I missed some subtle points.

Also, I confess to no expertise in this area of retroviral analysis or AIDs trials, just some years of background in other clinical trials. However, I must say that I found that a number of the analyses led me to raise the following cautions.

One, I think that if we had days to go over each analysis, we could probably answer some of these. But my concerns start with the potential for selection biases. I believe if I interpreted all the analyses presented correctly that none of them were actually the randomized cohorts but they represented selected cohorts based on either survival to eight weeks or 16 weeks or 24 weeks or whatever period of time the investigators could get a reduction measure.

So, immediately, we have some concern, I would think, that not everyone is in the study. In fact, you had to survive to that point in time, or at least you had to have a measure to that point in time before you could get into the analysis. This, I think, should raise a cautionary

note, indeed, as to what implication that has for any drug use.

The worst case would be, of course, that the drug was actually harmful in those who didn't get to that point. But if you made it to that point, you were okay. In fact, there was some suggestion in other trials, not in this area, that responders might do well. For example, in antiarrhythmic disease, a number of antiarrhythmic agents have been licensed by the FDA because antiarrhythmia is known to increase the risk of sudden death.

Antiarrhythmic agents reduce arrhythmias; therefore, the giving of antiarrhythmics should reduce death. In fact, until about six, seven years ago when a landmark trial was done, this was thought to be true. But then two drugs were tested and found to have a threefold increase in the risk of death even though they were shown, prior to the beginning of the trial, to reduce arrhythmias in that cohort that was eventually randomized.

So I think that is one example. There are others. In addition to that, and I don't want to take too long a time, but there were a few other things that I would like to note. One, a couple of the presenters pooled studies. I can see where, with the short time frame available to the committee to hear all the presentations, that may have been

a necessary evil.

But I hope, at some point in time, that the FDA or the companies present the individual studies. Pooling studies can actually mask or enhance an effect and I think the individual studies, given that there are different treatments, different risk groups, the individual results should be presented.

There were also multiple subgroup analyses that were done. Here, again, the cautionary note that there may have been a differential effect amongst the subgroups, but it may not have been a qualitative difference. It may have been quantitative. In other words, there was one analysis put up of under and over 300,000 copies per ml.

If you noticed quickly, as it went by, there was an effect in both groups. It was just better in one than the other. So one shouldn't preclude the use of the drug just because one is better than the other one.

Maybe I will stop here for now and let someone else get a chance.

DR. HAMMER: I was just wondering whether any of the speakers want to comment on the points raised.

DR. FEIGAL: I think many of the comparisons actually were across randomized groups. Your point about you are going to using a laboratory marker is that you have

to return to clinic to have the laboratory to be in some of these analyses is well taken.

Most of these trials are studies which have been previously reviewed because they have shown clinical benefit across randomized groups. I think that there are some aspects of the technology of the virology and the way the specimens were collected that sometimes left us with subsets where specimens were only collected in a sample of patients.

Part of the rationale and why there were so many similar presentations is I guess we were looking to see what kind of case could be made for the consistency of the evidence.

The issue of how to examine a relationship and how to describe it when you have often a categorical outcome such as progression or survival and you are looking for a dose response across some other variable is the reason, I think, that you saw that the data was often split looking for the kind of dose response.

I recognize that there is the problem of multiplicity but I think the kinds of evidence we are looking for was an ordered response across the direction and some consistency in that from study to study.

DR. HAMMER: Thank you.

DR. LIPSKY: I have a question for the FDA who saw

all the data; we realize historically we are looking at therapy which would, by 1997 standards, be considered outmoded. Although we are trying to glean some good information from it, it is possible we may be misled. One of the things, nadirs, et cetera, could be misleading.

I am just curious for today, for what we were presented today, how many patients were able to go down to levels that were undetectable in viral copies?

DR. ELASHOFF: I think it is a small percentage. I am not sure what the overall one would be; maybe 10 percent or less. Actually, it would depend on the particular study. The studies with more advance populations had actually much lower rates. Some of the NUCA studies I think were in the 60, 70 percent range, maybe a little higher initially and then they bounced back up again.

So it covered a spectrum, but there really weren't long-term response which, as Mike had pointed out earlier, by 24 weeks, it was a relatively small proportion across the board.

DR. LIPSKY: So one potential is I think we would have to be very careful to make conclusions, at least at that level on the data today. Maybe tomorrow we will see some things more clearly.

DR. EL-SADR: I think most of the data were in

patients with median CD4 about 200 or so. How comfortable should we be with extrapolating to people with higher CD4, especially regarding the new guidelines. It seems none of the data presented today were relevant to those with higher CD4.

DR. ELASHOFF: I think the median represents just that, the median, and not the entire range. So the actual CD4 counts across these thousands of subjects range from low to high.

DR. FEIGAL: Typical cutoffs were 350, sometimes 500, but typically 350. There was, I think, adequate amount of data on patients with very low counts. I think one of the things that the data brings into question, though, and particularly looking at some of the quartile plots, is the adequacy of assessing risk from CD4 count alone and whether or not you need to also look at baseline viral load at the same time.

DR. MATHEWS: It seemed from the background documentation and the presentations that the paradigm we are being asked to examine as a primary endpoint is is viral load sufficient versus the preceding paradigm which relied on clinical endpoints. But I think the discussion has implied that we are lumping clinical endpoints with CD4 responses.

DR. FEIGAL: None of these progression endpoints were the mixed CD4 count progression endpoints that were seen in some trials. These were all data with clinical progression. Part of it, I think, Chris, it is a little bit more subtle shift than just clinical endpoints or not.

One of the things that was emphasized in the previous studies where we tried to get six months of data on a new regimen compared to an old regimen was to try and get complete data for that whole time period because, as people dropped out, we got biased estimates of what the average response was because, quite predictably, the people with the crummy counts dropped out.

What we were asking participants to do was to remain in a trial even though we had evidence that they probably were no longer responsive to the drugs that they were in. So part of the emphasis is trying to break the process down a little bit.

An initial phase of studying the response of a drug by asking how deep is the response and what proportion of subjects get an adequate response and then looking, once you have that data on response, from there on from time to failure. So if you can define a loss of a viral-load response or some other type, you don't ask the participants to simply stay in the study on fixed regimens as though

there is no information, no prognostic information, when the viral counts rise or the CD4 counts begin to fall.

So it is, in a sense, trying to change the structure of the way that we study the viral load per se. The issue of whether there are clinical endpoints or not is not an either/or kind of a decision. We are saying that the way in the past that we have asked for information about the viral load with fixed regimens for periods of time without therapeutic monitoring, whether the patient had a response or not, is just not clinically appropriate any longer.

DR. MATHEWS: I agree with that but the point of my inquiry is that the bulk of the data that has been presented, and it is fairly consistent from every one of the trials examined, is that there is independent information from CD4 count and RNA response, that the responses are not always concordant.

For example, in the MAC study that was reported, there was a 20 percent discordance rate. Since both measures are laboratory measured in real time, if you are going to specify a laboratory marker as a primary endpoint, it seems to me it should include both viral load and CD4 response and not just viral load alone.

DR. HAMMER: That probably will come up again tomorrow afternoon, I suspect. I have a question for Drs.

DeMasi and Struthers. It follows up on Chris' point about discordance. Both presentations included a quadrant plot. Obviously, the bulk of patients who clinically progressed were in the low CD4, high RNA, quadrant. But has either investigation looked at the patients in the other quadrants?

I know those are small numbers but what you can tell us about those discordant patients, particularly those--and I know it is very small numbers--but those in the quadrant of low RNA, higher CD4. One of the questions that we will have to debate is this issue of a marker endpoint, a single biologic test, for a traditional approval with that indication.

I think there is some concern about the small proportion of patients who don't exhibit a "classic" response in both CD4 and RNA, at least with the drug classes we are talking about. Is there any comment from either Dr. DeMasi or Dr. Struthers since both showed a quadrant plot?

If there is no comment, there is no comment.

DR. DeMASI: If I understood the question correctly, it was the discordancy in the quadrants, the ones on the lower left and the upper right?

DR. HAMMER: There is about 9 percent of patients I think in your plot that don't fit into the neat right lower quadrant. Of course, we understand these are biologic

tests and the patients don't always fit into these neat categories but it does raise an interesting question about the discordancy potentially. I just was wondering if you have looked at that or have any comments about it.

DR. DeMASI: Yes. Actually, we did look at those events that occurred with, say, a high CD4 count and a low viral load it was relatively rare, as you saw. But what we did see is that some of these events were presumptive diagnoses. Some of them were non-AIDS-related deaths such as a heroine overdose and one accident.

When we restrict the analyses, what we have done further more is look at specific types of AIDS events in the particular bottom right. When we do that, when we limit to certain events such as CMV, disseminated or retinitis, we can even see this more of a concordancy in the CD4 and RNA responses.

DR. HAMMER: Thank you.

DR. STRUTHERS: On our data, we had four patients that had an AIDS-defining event in the high CD4 but low RNA. But we haven't looked at those patients individually to see the reason, what their event was.

DR. HAMMER: One of the interesting things about DDC has been, at least in some either as monotherapy or with ZDB in the 175 trial, you have seen virologic responses

without much in the way of CD4 responses. So that drug, alone or in combination, in some combinations, have shown some of this discordance and, for example, in the 175, experienced subjects, we know, in the virology substudy, there was a virologic response but no clinical benefit that could be determined in that study.

So it is just a note of caution, I think, that the group has to deal with. So thank you.

DR. VALENTINE: Just while we are discussing discordances, another type of discordance that is seen now in several studies is that with a large suppression of RNA copy number and the concomitant rise in CD4 cells, when the RNA copy numbers come back up toward baseline, at least, the CD4 cells do not come down but tend to hang in there at least for 20-plus weeks. That also remains to be explained.

DR. VERTER: I was wondering of any of the persons who presented did a kind of sensitivity analysis; that is, I thought I saw 25 percent of the cohorts that weren't included because they didn't have changes from baseline to 16, 24, whatever weeks, to put in a worse change and rerun the analysis. I wonder if anyone has done that.

DR. HAMMER: That is a general question to all of the presenters this afternoon. No one is leaping to the microphone. Ian has a comment.

DR. MARCHNER: First of all, I think the figure of 25 percent might be inflated a little by the fact that not only are we deleting people if they don't have the marker data, we are also deleting them if they had a clinical progression in the time frame where we are defining response, which we have to do.

That was one of your earlier points. If you are looking at the prognostic indication of a response over a given period, you have to look at the effect of that conditional on not having the clinical event during that period. For example, if you were to include people with clinical events during that period, you run into problems of whether or not having the clinical event will affect the marker value which is probably fairly likely.

So what I am saying is if someone has AIDS, that could bump their RNA up, for example.

DR. VERTER: That is my point, is giving the person the worse possible outcome and including the clinical thing. If you are trying to evaluate a treatment, it seems to me that by excluding the clinical events, you potentially--and I use the word very carefully, could wind up with something that looks beneficial whereas, in reality, you have excluded those in which it was harmful.

DR. MARCHNER: You have to look at the two groups

separately. You have got individuals who have short-term failure and you need to look at that group separately from the individuals who had their response over the period in which you are defining the response.

I think it makes sense to look at both groups. I don't think it makes sense to combine them in a single analysis.

DR. HAMMER: We can return tomorrow.

MS. LEIN: I saved it until last because it is quick. I just wanted Dr. Feigal, for the record, if you could just let folks know that this discussion is really relevant more so to antiviral therapies and, in the context of immune-based therapies, unless the mechanism is supposed to be HIV-specific, that HIV RNA is really looked at as a safety issue more than evaluating efficacy.

DR. FEIGAL: I take your point. I think this is specifically for antiviral drugs.

DR. HAMMER: On that note, I would like to thank the speakers today for providing just an enormous amount of data for us to digest. This day is adjourned. We will reconvene at 8 a.m. tomorrow. Thank you.

[Whereupon, at 4:30 p.m., the proceedings were recessed to be resumed at 8:00 a.m., Tuesday, July 15, 1997.]