My name is Jim Rooney. I'm from Gilead Sciences. I'm here today representing the Intercompany Collaboration for AIDS Drug Development. The ICC, as the group is known, is an organization of pharmaceutical companies open to any pharmaceutical company that is involved in the development of new drugs for the treatment of HIV infection. The current membership is listed on the first slide.

The ICC has been in operation since about 1993. It is relatively unique. I don't think a similar organization exists in any other therapeutic area. And I think it has fostered a great deal of degree of collaboration on both a medical and scientific level and communication between the member companies involved.

Next slide, please. The goal of the ICC is to share information and antiretroviral drugs in an effort to develop improved combination therapies for the treatment of HIV infection.

Next slide, please. In discussions with Dr. Struble in preparation for this meeting, she asked us to comment on three points. The first is the use of multiple experimental agents in registration studies for salvage therapy. The second was the use of placebos in salvage trials, particularly the

sharing of those between companies. The third is to comment generally on issues in the design and suggestions for the design of clinical trials in salvage therapy.

Next slide, please. With respect to combining agents, experimental agents, in salvage therapy, the current approaches, as you are aware, are geared to demonstrating the incremental benefit for each new drug.

However, the long-term durability of the response is more likely to reflect the activity of the entire regimen, rather than the activity of a single agent. Therefore, for salvage therapy, we do believe that it is reasonable to consider combining more than one experimental agent or registrational studies.

And that is, of course, providing that the regulatory environment is such that it is clear how those agents would be approved and that there would be some incentive for performing studies in this particular patient population, as opposed to other patient populations, where the designs are more clearly established and the benefit more clearly expressed.

Next slide, please. The rationale for combining agents is increased potency of the new

regimen and to limit the development of resistance to new agents. And it would be indicated in situations where *in vitro* data demonstrates synergy or additivity.

There would be minimal expected drug interactions or one would be able to compensate for those expected interactions. And there would be minimal overlapping toxicities or toxicities that would be manageable.

Issues are multiple. They have been elucidated this morning. They include isolating the benefit of a single therapy would not be simple given the limitations on study design, a limited number of new agents are available at any given time and they're in different stages of development and there are different amounts of available drug supply for these types of studies.

New data in any trial, new data, safety or efficacy, emerging during the study can affect the conduct, outcome, and acceptability of the study results. And certainly in the case when you're combining more than one experimental agent, this is even more so the case. There are limited data on long-term safety in agents that are not yet approved.

The attribution of safety events can be

complicated with agents whose safety profile is not completely elucidated. Unexpected drug interactions can and do occur and can complicate the interpretation of the study results. And unexpected safety issues with one drug could affect the other drug or the ability to complete the study. And this has certainly occurred in cases even where there are less than two experimental agents in a single trial.

Nonetheless, because of the strong rationale for combining agents, there are a variety of settings where the use of more than one experimental agent has already occurred. These include expanded access registration studies and nonregistration trials.

In expanded access, the rationale is to allow construction of more potent regimens and to help avoid functional mono therapy, to make better treatments available to thousands of patients before commercial availability of the new products. And because of this, most of the most recent expanded access programs have indeed allowed the availability of other experimental agents. And several of them are listed here on this slide.

Next. What about the use of more than one experimental agent in registration studies? This

slide lists some of the trials either currently underway or previously conducted that have allowed the use of other experimental agents as part of the treatment design.

In most instances, these have been included as part of the background regimen in some instances as part of a combined part of the single arm in the study.

Generally speaking, these are examples of studies to date, but I think there are even different types of designs that we could think of where we could use experimental agents in ways that would be beneficial to patients and perhaps with different regulatory guidelines in a way that could allow for either registration of single or both agents. I'll discuss those, actually, or some possibilities when we discuss clinical trials in the salvage setting.

Next. I would like to turn briefly to the issue of the use of placebos. Obviously it is an important component of the design of many trials. This slide outlines the various registration studies that have been conducted where placebos have been obtained from other companies, some of the programs more recent and some more distant.

In most instances, the patient population

has been either antiretroviral-naive or patients with limited treatment experience. There are, of course, multiple examples where placebos have been provided by companies for nonregistration of government-sponsored trials.

Next slide, please. But what about the use of placebos in a salvage setting? In discussions amongst the companies, it was felt that it was unlikely that placebos would be requested frequently in salvage studies, particularly in cases where there would be multiple experimental agents.

Obviously if a single company is conducting a trial with their own agent, they have their own placebo. So this is an issue of really more than one experimental agent and the placebos for those.

There was a concern that it would add pill burden and may decrease compliance. However, in those settings where it did make sense, I know there has been some discussion of factorial design this morning. And if it could be agreed upon in a study like that, then certainly the participating companies who do supply the investigational agents could supply the corresponding placebo.

Turning your attention to issues in

clinical trial design in salvage therapy, the issues are multiple. Most of these have been touched upon this morning. I won't reiterate them.

Next. What are some of the current designs that are currently being used for clinical trial designs by companies in treatment-experienced patients? They include superiority designs. Looking at regimen A, B, C, D and in a salvage setting, this is most commonly an optimized regimen to which you add either drug X or placebo.

One of the issues with this design in the salvage setting or entirely treatment-experienced patients is that the incremental drug benefit for a fifth or a sixth drug is often small and difficult to demonstrate.

Equivalence designs, where you look at fixed regimen A, B, C, D and compare it to, for example, A, B, C, X, where drug X is compared to drug D. In a salvage setting, there really are no fixed standard of care regimens to serve as the control. And the contribution of drug X in the regimen, in addition, the corresponding delta, the statistical parameter used to calculate power, difficult to estimate.

Next. Before turning to proposals for new

study designs, I would like to touch briefly on issues of endpoints. I know they will be discussed this afternoon.

As was mentioned this morning, the endpoints commonly used in other registration studies in less experienced patients, such as percent below detectable, is a categorical endpoint.

Unfortunately, response rates have been low in most studies in treatment-experienced patients. It may not be a sensitive endpoint for patient populations with high levels of HIV RNA at baseline and may miss clinically significant changes in HIV RNA.

Next. Alternative ways of looking at the endpoint of HIV RNA, such as looking at change from baseline or average area under the curve, DAVG, may be a better primary endpoint because it may be more sensitive to changes that could be clinically relevant.

This endpoint is currently allowed by both U.S. and European regulatory guidelines and is included in some antiretroviral labels. And, as you know, meta analyses from studies in patients with advanced disease conducted by the ACTG and FDA have shown clinical benefit associated with about a two and

a half-fold reduction, about a .4 log change in HIV RNA with a degree of benefit increasing proportionately with the degree of reduction in HIV RNA, 72 percent reduction, clinical progression for one-log reduction in HIV RNA at 24 weeks in the Marshener paper.

Next, please. Well, what about some new study designs in this patient population? There are no easy answers, obviously. So what we would like to do is review just a few, some of those that have been discussed in the context of this morning's discussion and a couple of other proposals as well.

With respect to factorial designs -- and this is the simple factorial. The optimized would obviously have arm one deleted and just look at two new drugs, A and B, AB and AB.

Next slide, please. Advantages would be ideally can answer many questions with one study is one way of including more than one experimental agent in a study deign. I think certainly those are favorable aspects. A sample size could be larger or smaller depending upon the numbers of questions asked.

Next. However, there are, as was discussed this morning, some limitations of this design. Included amongst those, if you do believe

that adding single agents to optimized therapy is not indeed an optimal regimen, there would be in one type of factorial design, at least, a single agent added to background therapy.

If three agents were available, then obviously that wouldn't necessarily be the case. But it's possible that patients may be exposed to less than optimal therapy, a significant number of patients. The point of combining experimental agents is because single agents have not been optimally suppressive.

More importantly, however, from a regulatory perspective is that interactions between treatments can undermine study results. Unfortunately, interactions are not uncommon, treatment interactions. They can be based on PK, virologic, or metabolic reasons.

The main effect of the single drugs A or B in this kind of study design could be kind of difficult to estimate in the presence of a significant interaction between A and B.

It was mentioned this morning that if the interaction is positive, it wouldn't negatively affect the regulatory approval of either agents A or B. However, if they're not positive, then it could affect

the likelihood of regulatory approval of those agents.

Unfortunately, there is actually regulatory precedent for this where a recent example of a registration study was invalidated, reviewed by this Committee because of a concern regarding a drug interaction in that trial. And this was Gilead Study 417.

For all of these reasons and in discussions amongst the companies, I think most companies feel that factorial designs are not an optimum way to develop for regulatory approval drugs in a salvage setting. However, they could be considered for nonregistration exploratory trials.

Next. What are some other approaches? As was discussed this morning, allowing use of other experimental agents as part of an optimized background regimen is certainly a possibility, a good design option. It is currently being used in several trials. And it is possible to include agents, not only in expanded access but also earlier in development.

In this case, however, although two experimental agents are involved in the regimen, regulatory consideration is only being given for one of those agents.

An alternative approach, which could be

used for single drugs or for combinations, would be very similar to the proposal, the two-part proposal given by the FDA this morning or the European proposal just presented by Dr. Vittecoq, where short-term activity was assessed during an early period of mono therapy or add-on therapy and longer-term safety was evaluated in combination therapy.

It makes sense for the reasons that have previously been elucidated, demonstrating antiviral activity in the target patient population, limiting time on mono therapy, development of resistance. And on a regulatory basis, this method could provide a more efficient way of identifying and making commercially available new agents that could be used for patients most in need.

Next slide, please. Very similar to the design just mentioned, here drug X and placebo or no therapy are evaluated over an early period of time. The duration of mono therapy would be individualized for each drug intended to limit the development of resistance.

For those agents where development of resistance was especially a concern, it would be possible certainly to add drug X to the failing regimen, instead of being evaluated as mono therapy.

And then all patients would receive drug X plus the optimized background, which could include other experimental agents.

Period of combination therapy, as alliterated by Dr. Vittecoq, would be customized to maximize benefit for each patient based upon resistance testing, past treatment history, tolerability, potential drug interactions, et cetera.

Next. A period of mono therapy would provide evidence of antiviral activity in this patient population and short-term safety and tolerability. And it would be assumed that the short-term antiviral activity should provide evidence of potential long-term benefit in a fully suppressive regimen.

There certainly are many examples of other drugs when given in naive patient populations that when given as mono therapy would be quickly associated with the development of resistance but when combined in an optimally suppressive regimen can provide a durable antiretroviral response for years. So the same principle would be assumed to be applied here and that the durability would be a function of the potency of the regimen, not just the drug.

Period of combination would provide optimal, potential optimal, therapy to all patients,

would examine durability, safety, tolerability. One could also monitor patients for development of resistance to experimental agent. There are various variations on this proposal.

Certainly one is to look at placebo control as part of the period of combination. Most of the companies in discussing this I think favored a proposal closer to the FDA proposal this morning, where all patients would receive optimized therapy.

Next slide. This is a slightly more radical version of the same proposal and in this case attempts to make use of the benefit of using two experimental agents together. And in this case, drug A and B would really be considered as one experimental therapy and, again, would be compared against either placebo or no therapy and then combined as part of an optimized regimen.

Here it would be assumed that the benefit of drug A and B if it were indeed effective, that it would be approved essentially as a combination therapy itself but that the individual activity of these agents and how they contributed to the regimen would be worked out as part of studies ongoing in other patient populations; again, the same issues.

There would be some regulatory issues

obviously involved with these suggestions. Could in the first case of the single agent Drug X based upon this proposal or in the case of the dual combination all experimental agents be approved based upon these types of studies? Would additional studies in the case of combinations be required to clarify the contribution of each agent to the successful regimen prior to approval?

Again, there could be implications also for negative study. These studies are conducted without optimal information about how to use the drugs. So if there was indeed a negative outcome from the study, could it diminish the likelihood of regulatory approval with a second positive study from a different patient population and would it limit the use of the drug in treatment-experienced or heavily treatment-experienced patient populations?

So, in conclusion, general recommendations. ICC member companies support the use of multiple experimental agents in salvage therapy. We do not favor a factorial design for registration studies. Studies of short-term mono therapy combined with longer-term combination for regimen consideration is the basis for regulatory approval.

We do suggest that a positive study of two

experimental drugs could support registration or should support registration of both drugs for an indication limited to that combination in salvage therapy with data from further trials to be used to extend the indication.

Paramount, given the discussion this morning and the complexity issues, obviously flexibility in approach to all of these issues, study design, use of comparators and choice of endpoints, is very important. And, really, each registration package should be customized to the drug in a patient population being explored.

Thank you.

ACTING CHAIRMAN GULICK: Thanks, Dr. Rooney.

That concludes the people who signed up to speak at the open portion of the meeting. Is there anyone else who would like to make a public statement who did not sign up?

(No response.)

ACTING CHAIRMAN GULICK: Okay. So we will close the open portion of the meeting. I return the Committee to our questions.

CONTINUE QUESTIONS TO THE COMMITTEE

ACTING CHAIRMAN GULICK: We have done

Question 1. Question Number 2 is specifically considering some of the study designs. And I would like to start with the three that Dr. Laessig presented in her presentation this morning. Would it be possible to get those slides up from Dr. Laessig? Oh, great. That's a summary.

DR. MURRAY: That's the correct slide.

ACTING CHAIRMAN GULICK: Great. Let's go with that. All right. I think I would like to consider them in the order that you presented them this morning. Let's talk about add-on, the add-on, design first, which is number two on this slide.

So that's optimized background plus drug A versus optimized background plus either a matching placebo to drug A or no treatment. The charge to the Committee is to identify the strengths and weaknesses of this particular design.

DR. JOLSON: I just have one thing to remember. When you look at optimized background, that's with the assumption that access to drugs available and expanded access is equally accessible to both groups.

ACTING CHAIRMAN GULICK: And resistance testing by the same regard.

Dr. Eron?

DR. ERON: One thing we have not talked about with optimized background is actually to have a requirement for people to be sensitive to two drugs, let's say, in addition to the drug being studied. So in a phenotypic assay or in a genotypic algorithm, you're only eligible if, indeed, you are susceptible to two drugs that are available to you.

I think that might be somewhat more acceptable than kind of adding study drug A to anyone who qualifies, whether their regimen can be truly optimized or not. I don't know what other people think of that approach to this design, but --

ACTING CHAIRMAN GULICK: Dr. Mellors?

DR. MELLORS: Yes. I think Joe has hit the nail on the head. It depends on how optimized optimized really is. If your phenotypic or genotypic sensitivity score is one and you're looking at an add-on therapy with a fragile drug, that's an unethical trial. Okay?

If the average phenotypic sensitivity score approaches three for the optimized background, then adding another agent is a nice way cleanly to demonstrate efficacy.

I don't want to go back to the morning. That was the point I was trying to make. It's simple.

All of the comments are well-taken that if you have lousy optimized background, it's a lousy design. And this is one way to individualize designs for given patient populations.

DR. ERON: And having a score of two or three could be an entry criteria.

DR. MELLORS: That's correct.

ACTING CHAIRMAN GULICK: Dr. Cunningham?

DR. CUNNINGHAM: As I have said, the advantage of that is it is clean and easy and it certainly could only be done in people where the optimized background is a reasonable background.

However, we know that even when the optimized background is a reasonable background, the failure rate is very high. So there would have to be an early escape mechanism in that kind of trial.

ACTING CHAIRMAN GULICK: Dr. DeMasi?

DR. DeMASI: Yes. One of the issues that I would like to bring up for discussion is in this type of design, making the distinction between the activity of the individual drug, the study during the first two weeks, versus the efficacy of the regimen within the trial that is being studied in the trial and beyond that two-week phases what is needed in order to demonstrate safety and efficacy of the

regimen, including this new agent.

ACTING CHAIRMAN GULICK: And I think you are reaching towards the next design, actually, which is the two-part hybrid that -- let's stick for a minute --

DR. DeMASI: Just to clarify, in the two-part hybrid, you see that the treatment groups actually come together. But what I am suggesting is the necessity for an additional randomized phase of the study beyond this two-week period to further demonstrate or confirm the activity that you would see during the first two weeks of the study.

ACTING CHAIRMAN GULICK: Ms. Dee?

MS. DEE: I don't know how you would do that. And that would be interesting to see that because, really, when you talk about switching -- I think this is right. Maybe Victor or Dr. Hammerstrom can comment on this.

Once you're over that eight weeks and then you have an early escape switch point, which I think that appears to be more ethical anyway, don't you really have an eight-week trial? I mean, isn't that the result of that? So the data is eight-week data. Is that right or wrong?

DR. HAMMERSTROM: I don't think that's

exactly right because you would say that at eight weeks, those who have failed on the placebo are allowed to add: A) open label. They essentially go into an expanded access program.

But it need not be the case that everybody will have failed on background plus placebo at eight weeks. Perhaps 20 percent, perhaps 40 percent of them have not failed and the rest of them would still be going on. So you would still be collecting data for whatever fraction of them haven't failed out to whenever you make your accelerated approval decision.

In fact, if people continue out to 36 or 40 weeks on background plus placebo and still haven't failed, you would still be collecting data on them on the comparator arm, even relevant to the traditional approval at 48-week data.

So what we would consider an eight-week trial would be one where everybody stops on the placebo arm at eight weeks and switches to accelerated approval or switches to expanded access.

As long as you are proceeding on the background plus placebo, you stay on that, at least out to the scheduled end of the trial, which is 48 weeks, then we wouldn't consider it an 8-week trial.

And you really are getting a comparator

that goes all the way out there. It shows that on optimum background plus placebo, you get a failure rate that goes down like this and maybe 10 percent of them are left at 48 weeks; whereas, on optimum background plus A, you've got a failure rate that goes down hopefully much less rapidly and 40 percent of them are still succeeding out at 48 weeks.

ACTING CHAIRMAN GULICK: Dr. Pettinelli?

DR. PETTINELLI: I was just going to say I agree with the statement that is being made. In all the rest, even if you have like a one or two-year duration, we have early escape for a single patient, not for a trial.

I would also like to comment that the add-on, as was stated by others of my colleagues before, is indeed possible when the patients are defined as being sensitive to drug at least. So, again, we have to look very carefully to the patient population.

I don't know how many of those patients would be in that category. Probably the additional two experimental drugs would be the most common occurrence just because that will increase the sensitivity.

ACTING CHAIRMAN GULICK: Dr. Saag?

DR. SAAG: Yes. That's what I was going to say. Actually, it's probably healthy to have the morning discussion because it seems like we've got a little bit more clarity this afternoon. And that is that looking at the top one compared to the second, the modified versus add-on, I think that's exactly right.

If it's early, which is what our charge is now, to look at people who at an early point failed two HAART regimens, all classes involved, that the add-on may be the wise choice if, again, you have three drugs that are available and you're simply looking for the fourth and, like John said and others, it's clean. That's very nice for registration.

For the little bit more advanced, like, say, you could almost have it in the same study where you do an add-on if there is a score of three and if the score is two or less than you go to the modified factorial. And you could actually either have a separate study that runs concurrently, but at least you can cover the waterfront.

But for those patients who have -- I was worried about the top one if you had three drugs available for optimum background because then you're comparing three versus five-drug regimens. An that's

probably too convoluted for somebody who is earlier in the course of failure. So I think this may be a consensus emerging that's kind of nice to hear.

ACTING CHAIRMAN GULICK: Dr. Mathews?

DR. MATHEWS: You know, the devil is in the details, particularly I think up front, where we are trying to define inclusion criteria and what exactly is meant by optimized background regimen.

The notion of a sensitivity score based on the resistance collaborative group I think is very attractive, but obviously you have to assume that the patients that are being screened for the trial are on some therapy that makes the resistance tests at the time they are screened interpretable.

You can imagine a number of situations where people might have motivations to go off therapy so that they would qualify for a trial.

So I don't know exactly how that is dealt with. But, in addition to having a cutoff score of two or whatever, one could further stratify the randomization based on some measure of sensitivity where there is clear uncertainty on what the potential to respond is and the basis of background resistance.

That should include, actually, whatever is known about the experimental agents because many of

these drugs are coming into trial where the resistance profile is not fully characterized, particularly thresholds for sensitivity.

ACTING CHAIRMAN GULICK: Mr. Levin?

MR. LEVIN: I want to modify what I said this morning. So I'm glad we had all of the arguments and everything because I have had a chance to think about it and talk to some people. I do have some concerns, and I want to express them.

I am concerned about drug interactions, and I am concerned about making sure that every regimen in a salvage study is somewhat equally effective in being able to suppress virus, no matter which arm it is.

Having said that, -- and there may be some other concerns -- I don't want to rule out modified factorials. So I want to change my opinion on that and go on the record for that.

It does appear as though next year there will be at least four or five or maybe more new drugs. And so if you can find combinations that are somewhat equally effective, I think that would be important.

I don't think I have a problem with 16 weeks. I think that that's okay as long as there is an adequate antiviral activity, as said before, of at

least a half a log.

One thing I want to add here, though, is that we have not really spent a lot of time, adequate attention to identifying toxicities and side effects, including hepatotoxicity.

I would like to suggest that all of the companies and the FDA get together and find some way to maybe create a database for all salvage studies where we could collect toxicities, hepatotoxicities, and side effects and maybe come up with some information; in particular, lipodystrophy but also, in particular, what gets very little attention at this point is hepatotoxicity. We really need some data on people with hepatitis and HIV medications and what is really going on.

ACTING CHAIRMAN GULICK: Thanks for your comments. I would like to refocus us on the study design. Let's people quickly move from the add-on to the two-part hybrid study. Yes, Dr. Eron?

DR. ERON: The issue of bias in the use of a placebo in this particular study and just to get, in particular, the add-on study, to get people's comments because if there is an early pop-off and the patient knows that they're sensitive to two drugs, -- let's say that is the cutoff -- there may be somewhat of a

disincentive if it's a no treatment control, as opposed to a placebo control. I wonder if, Victor, other people have thoughts about that.

ACTING CHAIRMAN GULICK: So you're advocating for blinding, it sounds like, to deal with --

DR. ERON: Yes. I mean, it would make sense to me.

ACTING CHAIRMAN GULICK: -- the bias in that situation.

DR. HAMMERSTROM: There is a way or something we do in place of blinding. It's not as good as blinding, but there are instances where, like an injectable drug, it's basically unethical to inject somebody with saline solution as a placebo. Well, it's infeasible anyway for something like that.

What we want to see or at least what has been proposed now is we want very rigorous criteria that say this is what you should look like at week eight to be classified as a non-responder and to get into the expanded access.

I was given an example. It doesn't have to be the right one. If you have dropped half a log from baseline, if you've done that, you're a responder. If you haven't done that, you're a

non-responder and you're allowed to go to expanded access.

If you have dropped half a log or a little bit more than half a log but you want to go to the expanded access anyway, then that kind of thing raises a need for sensitivity analyses.

That's a sort of a biased differential dropout. We would not want to see that too much. We would basically like at the end those kinds of switches shouldn't be occurring frequently enough to be affecting the conclusion. So, at least up to now, that is the way we compensate for absence of blinding.

You have a very rigorously defined exit criterion. And the people who exit, even though they haven't met that criterion, have to be subject to some kind of sensitivity analysis to make sure that their switching isn't the reason you are finally ducting drug A is effective.

DR. ERON: Sure. And the concern that I have there, though, is that there might be differential adherence. So the people who are on the no treatment arm would maybe not be as adherent to their therapies such that they legitimately make that cutoff.

I suppose one way around that would be

whether potentially at drug levels --

DR. HAMMERSTROM: Yes. That actually came up on an IND we did. Did we come up with a solution for that? The incentive for someone to cheat when they're --

DR. JOLSON: It's a concern. It's part of the concern of doing an open label study.

ACTING CHAIRMAN GULICK: Okay. Any last thoughts on add-on before we move? Ms. Dee?

MS. DEE: You know, this would be maybe the most attractive to industry, but it's probably the least attractive to the patient population. So hopefully somewhere we can both give a little because if you're talking about the patient that the agency described this morning, maybe they won't get in your studies because they do have some other options and they do have some time before they really are "desperate," in quotes. So maybe they won't get on your study and you won't be able to accrue it.

ACTING CHAIRMAN GULICK: Just to be specific --

MS. DEE: The add-on, in other words, with the placebo, the add-on plus A versus placebo.

ACTING CHAIRMAN GULICK: Would not be encouraging to the patient --

MS. DEE: Right.

ACTING CHAIRMAN GULICK: -- because they may randomize to a placebo?

MS. DEE: Right, right.

ACTING CHAIRMAN GULICK: Dr. Cunningham?

DR. CUNNINGHAM: Just briefly. I had made a couple of comments about the add-ons. And then when Dr. Saag commented, he said he felt that there was a consensus about these being appropriate.

I guess I didn't want to imply that I thought that that's -- I was trying to point out some of the pros of the add-ons. I think the down side of the add-ons is that you get less information about drug interactions and you overall get less information than the other types of trial designs.

So I am not sure that I would say that I agree that there is a consensus that that is what we should do. I think that in certain circumstances, that is the appropriate trial design but not always.

ACTING CHAIRMAN GULICK: Okay. Yes, Dr. Blackwelder?

DR. BLACKWELDER: Another comment on the add-on. It sounded like there was a suggestion a few minutes ago that you could continue to follow patients beyond, say, eight weeks if eight weeks was the

primary time of evaluating and just keep the placebo group, whoever was left.

That is kind of problematic because you don't have the randomized study anymore. You just have a subset of them.

ACTING CHAIRMAN GULICK: All right. Dr. DeGruttola?

DR. DeGRUTTOLA: I don't quite understand that because I would think if you define a failure endpoint and you say that as soon as people reach that failure they can get access to the new drug or have some other strategy, then you still have a full randomized comparison between the two groups.

What you are doing is following patients until they reach failure. Then they are contributing their endpoint to the study. They are contributing their endpoint to the randomized comparison. Then after they reach failure, they can go on to another treatment. So I believe that this approach can be a full randomized comparison and not require subset analysis.

I do want to agree with Dr. Cunningham's point that while this may be the best design in some settings, in other settings where you are interested in looking at two new agents, the ability to study

them together and allow you to look at interactions, both in terms of toxicity and efficacy I think is important to consider.

Dr. Rooney mentioned that interactions can sometimes complicate interpretation, but I think if you are concerned about interactions, that is all the more reason to do a study up front, like a factorial or the so-called modified factorial, that allows you to evaluate those interactions in a structured way if you are talking about two drugs are ultimately going to be used together.

ACTING CHAIRMAN GULICK: Dr. DeMasi, the last word here.

DR. DeMASI: Yes. I just wanted to clarify the point about the eight-week potential switch and then looking at the eight-week and then subsequent 16-week, statistical analysis, to compare treatment groups.

I think that, echoing Dr. DeGruttola's comments that we're looking at an endpoint of cumulative virologic failures up through week 16. And a virologic failure that would occur prior to that which would allow patients to go on to a new drug would contribute to that week 16 analysis.

And for a more conventional or a change

from baseline type of a metric, if a patient switched at week eight, you could carry that observation forward in terms of the RNA result at week eight to week 16 when you did the week 16 interim analysis.

ACTING CHAIRMAN GULICK: Thanks.

DR. BLACKWELDER: You might want to do a comparison just of that subgroup with --

ACTING CHAIRMAN GULICK: Okay. We're moving on to the two-part hybrid, which many moved on to right away. Dr. Schapiro?

DR. SCHAPIRO: So regarding the two-part hybrid, by the way, Trip, I think regarding terminology, if "salvage" is based, I think "two-part hybrid" is a tough deal. I heard you snickering and saying something about me being a two-part hybrid when I got up. I think we should find a better name for that.

To summarize my comments on it, I think we did hear from some of the speakers. I would say we are able to get a lot of information from this type of study.

I think we do have two phases. Not to go over endlessly, there are differences. We are looking not only at the ability to get patients undetectable. We are looking at the ability to get them down. I

think Mike Saag and Steve Deeks made these points at sometimes. 0.7 log can be great. So we want to differentiate between the two of them.

I do think that initial number of weeks that we look at the drug itself does give us the opportunity to see: Is it potent, to what degree it's potent? And I think to some degree, we get some toxicity adherence also in that short phase. If a drug really is hard to take, we can sometimes tease that out in that short period of time.

And then the additional phase, which I'm just calling month or maybe a year here, we get the additional information about going undetectable. And we get some of the more CD4 adherence.

I think this is different for different drugs. I don't think you can give it as a number of weeks necessarily. I think it depends which drugs you're studying.

I think the key factor which was brought up earlier is that you don't want resistance to be generated in that time. I think for an NNRTI, this might not be appropriate. We have data that one or two doses can be enough.

So we can't make sweeping suggestions regarding this. I think the two-part hybrid, for lack

of another word for it, would not be appropriate for an NNRTI. The way we see them today, it probably would be appropriate for PIs and NRTIs. And we maybe could slide it between two or three weeks depending on what is being looked at.

A concept I think that we heard from Professor Vittecoq earlier which is being looked at is the degree of the slope. We don't have a lot of information on this, and we have to remember this will not predict necessarily long-term failure.

I think that John Mellors presented some data that long-term success is not necessarily dependent on just how potent you want. We know that there are other factors. If we want to see how potent the drug will be, the slope may be beneficial.

There is a study being done now at Stanford and in Holland with Anders DeLoupa being the PI on this where you take frequent measurements over a short period of time and possibly will be able to look at the slope. It may be that this would suggest a more potent drug than this than this.

I don't know if ultimately this will work out, but I think if we are going to be doing these type, I would encourage people to try to see if this works or not.

It may be that there is too much noise. It may be that it doesn't work out. But based on the pathophysiology of the disease and the impact that we have seen from some of the dynamic studies, there may be something here. And that might be something which we can use as another measure. Again, how potent is the drug?

Now, I think something we should also use that initial phase for is probably here. In this phase, we would want to try to delineate again with a specific resistance profile how much kick you get from that drug.

And I think that by taking a resistance profile at baseline, a genotypic and/or a phenotypic study, and then at the end of this initial stage looking at the drug level and also looking at the virologic impact, be it the reduction, the change, the slope, whatever we have, we can then try to tie in this correlation.

We will want a drug level which will tell us the exposure. And then we will want to put together these three parameters and come up with a statement which will tell us that if you have this genotype or phenotype and you obtain this drug level, then this will be your viral load response. We can

then take that and use it in other patients.

I think, as we discussed earlier, this is something that we can take home and we can use in different patient populations. When we see that phenotype or that genotype, we will -- and it doesn't have to be the exact same patient population. Anyone who has that, we now have data that with this dosage, you should get that response.

We don't necessarily have to repeat this study to all of the different populations because we have correlated a specific baseline resistance with a virological response.

True, it's short-term, but maybe if we look at it in a number of ways, it will be quite robust. And, again, if we add in here the drug level, we will be able to determine for that dose we can try to quantify how much effect we could get.

And I think Dr. Jolson made a good point earlier that we have to look at patients that are a little bit less advanced.

I think it's an excellent point. Thinking about it, I think the real difference this morning in the discussion, the difference between the deep salvage and I think the very appropriate group determined by the FDA is class resistance versus drug

resistance.

I think deep salvage means patients who have class resistance to all three; whereas, the patients you guys are defining I think in a very nice way are patients who may have resistance to drugs in the three classes but not necessarily class resistance.

Therefore, they may be appropriate to different degrees to look at this. We may be able to take this data, then, and look at it, even in the deep salvage.

ACTING CHAIRMAN GULICK: Thanks.

DR. SCHAPIRO: Thank you.

ACTING CHAIRMAN GULICK: Dr. DeMasi, you have another design that's also the two-phased, can we call it? How is that? Bridged phase?

DR. DeMASI: Thank you.

I just wanted to take this opportunity to present this additional design here, which I think summarizes or encompasses many of the points that have been addressed today and discussed this morning and this afternoon.

Essentially what I have done is I have focused this design in terms of a very specific phase of drug development, namely a Phase II study. It's a

randomized control trial looking at two doses of your drug compared to a placebo or no treatment background.

Essentially what I have here, I have called it the bridged phased I/III randomized control trial because I do think it contains characteristics of both Phase I, obviously the Phase II, and a Phase III study.

Essentially the basic study design is that patients are randomized to one of the two doses of the drug plus the optimized background regimen with either a placebo or no treatment. So everyone is able to get the optimized background regimen with additional investigational agents.

I think the importance of this study is because in today's discussion, I think there may be some under-emphasis of the importance of finding a correct dose for your drug to maybe take into a factorial design or a strategy trial or a pivotal Phase III study. And in doing this type of a design, where you can actually determine the dose, you can bridge some of the information you learn about your drug in Phase I to Phase III.

So more specifically, I have proposed here, similar to what has been proposed as a two-part study with a randomized comparison of viral load

response in the first one or two weeks. And the study is actually powered to detect differences in viral load. That estimate of the number of patients actually comes from your broad-based, dose-ranging study in Phase I.

The second phase is -- and I have specifically noted activity here in terms of the activity phase. The second phase is more of a comparative/efficacy phase of the study in which you maintain the randomized feature of the trial through some period; for example, 12, 16, 20 weeks.

And here in this example, I have noted 16 weeks. At that point, you could actually based on an interim analysis of the data roll patients into the optimum dose regimen, either A or B, and follow these patients to generate long-term safety data. In terms of building your safety database, that would contribute to an NDA submission.

I think a unique feature of this design because you have patients who are on the new drug plus additional investigational agents is that you can build into the fourth phase a randomized, potentially randomized, withdrawal study in which you randomize patients to either continue or withdraw from the investigational agent and you can look at the delta in

terms of an RNA rebound over one or two weeks from the RNA value at the time that he discontinued the investigational agent.

Just to complete the discussion, I want to summarize some of the points here. Obviously, also mentioned earlier, in terms of the features for the Phase I, for example, exploratory interim analyses of the two-week data could be conducted to look at correlation of genotype and phenotype by baseline with initial virologic response and also the PK/PD modeling, as was just mentioned.

I think in terms of some of the other features, the switch option could be built into the study here, for example. It's at eight weeks. So this would allow patients to switch either to the investigational agent or potentially an optimum or good dose of the investigational agent.

Additionally, I just want to conclude by saying based on this, I think, study, which could be done in a pretty fairly broad population, that additional exploratory and subset analyses could be used to refine the population that could be studied in a more definitive and larger Phase III study, although I believe that if you can demonstrate differences in terms of safety and activity during this phase of the

study, the randomized phase, through an intermediate time point, that this could be suitable for submission for a supportive study or a pivotal study.

ACTING CHAIRMAN GULICK: Comments? Dr. Fletcher?

DR. FLETCHER: I just want to say that I would be a strong proponent of this design. I think there are numerous attractive features to it. In particular, the ability to study early on, the pharmacokinetic characteristics in your Phase I move into two doses so that you can really, then, understand what is going on.

Certainly a variant, instead of two doses, could also be two concentrations. So the idea of a concentration-controlled study has come up. So, instead of randomizing to two doses, you could randomize to two levels of exposure as well and then from that work into your Phase III, where now that you have an understanding of your concentration-response relationship, you could then do a refined dose and probably simplify it into Phase III.

I think in my mind, a particular attractive feature of this is that if we continue a current approach in treatment-experienced patients of trying to determine virologic characteristics, whether

they be genotype or phenotype, but still continue to apply the same dosing strategies that we use in naive patients, we are doomed to failure.

If potency of a drug is some function of concentration to susceptibility, that ratio in a treatment-experienced patient is not going to be the same as in a treatment-naive patient. And so there has to be a method by which you can use that ratio or some other approximation of potency and learn about what it is in naive and begin to apply that into treatment-experienced patients.

One difficulty I have to mention in all of these, which has been brought up numerous times under the heading of drug-drug interactions, is we can't underestimate what a serious challenge that is going to be to have knowledge of drug-drug interactions before you launch into these types of studies.

If you think about, say, an optimized background, maybe it is not even the same among all patients. So maybe we allow six drugs to be used in an optimized background and then you wanted to add on two more drugs.

You now have eight drugs that can be taken five at a time, I think, if my math is right. And the amount of drug-drug interaction knowledge that you are

going to have to have becomes enormous.

Now, this design, the bridge one, does help in that because in that if you want that two-week lead-in phase, if you're committed to it, you could do an awful lot of learning about what drug-drug interactions are there and use that in a very fast approach to modify anything that is surprising or way out of the realm of what you would want.

ACTING CHAIRMAN GULICK: Dr. Mellors?

DR. MELLORS: Yes. Just a couple of points. I like the two-week lead-in phase because it's, again, clean and you can do PK/PD modeling, but there is risk. Rather than say a given class should or should not be applied to this design, like NNRTI might not be good or fusion might or might not or PI, it really depends on the characteristic of the drug *in vitro* and what the genetic and pharmacologic benefit is to resistance.

So you can make some estimations that a drug would or would not be a good candidate for this. So that is one caution. The other caution is -- and, Ralph, it is very ambitious, and it is really kind of pushing the envelope, but that is asking an awful lot of a trial to accomplish all three phases.

I would be happy if I got two phases

completed. The caution I have is that you want to make absolutely sure after the two-week mono therapy lead-in that you haven't done any damage to the response to the drug, namely you want to make sure that you haven't selected resistance and diminished activity.

So I am in favor of the design, not necessarily for registrational purposes because it is asking an awful lot of a single trial to go from Phase I dose-ranging mono therapy through Phase III registration because it may be successful if everything goes right, but chances are you will learn some information in the first and second phase that modifies the Phase III design.

ACTING CHAIRMAN GULICK: Dr. DeMasi, a response?

DR. DeMASI: Just to address a couple of the points, I think in terms of the suitability of the study design, I agree that it does depend on the particular drug that could be used in this type of study.

In the two-week lead-in phase compare an activity was presented as an example. And that could vary because of the resistance profiles of a particular drug to seven to ten days perhaps. And in

terms of the other coin, I agree that in terms of the logistics of the study, it would be a difficult study to conduct, but I do think it is feasible because if you look at the study design, the additional piece that you're carrying over that's there that's not in the two-part hybrid is the additional randomized phase beyond the two-week study period.

So I do think it is feasible to continue the study beyond two weeks and collect and look at the comparative contribution of the study drug in two different doses versus a background and no placebo.

ACTING CHAIRMAN GULICK: I'll take two more comments. Dr. Eron and then Dr. Falloon.

DR. ERON: In the FDA design that was put up, there is no control arm. Is that correct? Is that --

DR. HAMMERSTROM: Can we put up the slide that had our three trials up on it again?

DR. ERON: The problem with that is I don't understand how you have any certainty that the prolonged effect is not just due to the optimized background. In fact, Trimeris has already done this study.

DR. HAMMERSTROM: This is exactly. One of the questions that I have with this design is that if

the statistical analysis is straightforward, you've got three comparative arms. It's easy enough to do statistically valid comparisons.

What is going to happen between day 10 and week 24 is that at the end of week 24, you're going to have one sample. And you're going to look at that. You're going to look at change from baseline or change from day ten level and test: Is that change from baseline less than zero?

So what this amounts to, if you're going to conclude from that that this is evidence of durability, the effect, it amounts to a historic control data in which the historic control is, in fact, everybody's subjective impression of their case series that had they stayed, we know, everybody knows, without looking at any more data that if you stayed on a failing regimen, what would happen to you after 24 weeks would be that you would go up. If you get a statistically significant decrease, then that's evidence of effect.

I would like to know: Is that convincing to other people? What else would you need to this trial for that test to be convincing because if anyone does this at the final review, that's the question I'm going to have to or we're going to have to answer?

DR. JOLSON: The study that I'm thinking about is the T-20 study, where they gave exactly -- you know, it was 28 days of T-20. And there was a hiatus, but then people got optimized therapy and T-20 with no control.

Certainly every time I have presented that study people have said, "You can't make any conclusions about the impact of T-20 on that outcome" because I don't think there's any way to know whether the optimized background is what's driving the antiviral response. And it ends up being, "Well, I know that these patients in my clinic would have never done this well."

DR. MURRAY: First of all, I wouldn't envision this for a new class of agent. I would envision it where because the second part, the hybrid part of it, as you're doing a prospective observation or cohort or are you actually using the heterogeneity of the population in terms of their viral isolate sensitivity to make a conclusion about what was happening at the end, kind of similar to the 957 Kaletra study?

So it's randomized at the beginning, and you know that at the beginning if the new drug was contributing because if it wasn't contributing

anything, you wouldn't see any response at all probably during the first ten days to two weeks.

You know, the difference between that and maybe the T-20 study you're talking about is, too, that you're also seeing the initial contribution of optimized background. So it would be unlikely that a drug which was contributing, let's say, a log in the first two weeks was having absolutely no effect at the end. Of course, that's a bit of assumption.

But then I guess I would envision them tightening up the study at the end of 24 weeks with either to do a dose-response or you could do some -- at that point it becomes not uncontrolled but not randomized, an observational cohort, where you're using your baseline sensitivity.

I would agree this is a controversial design. We have not used it for registration in the past. We are trying to think of if you couldn't have corroboration, if you didn't feel comfortable about an add-on -- and I might add on an add-on, you do have to worry about using up your optimized backgrounds.

I mean, how many chances do you get with optimized backgrounds? So if you failed it, at the time you are ready to have your escape option, then what's your optimized background?

Anyway, so we're coming with maybe what would be a possible third option? And then we've gotten these. I think this is a bit similar, maybe to the EMEA proposal, and then we've seen some other proposals.

Then we got some I think pretty good information for lopinavir with a design somewhat similar to this, although they didn't have the initial mono therapy period in that trial, but they did in other trials.

So I realize it is controversial, but if it isn't, if that last part, the second phase of it, does not appear to be strong enough, we would like to hear that.

ACTING CHAIRMAN GULICK: Dr. Falloon?

DR. FALLOON: In going back to the question of what to do in the early part, there's a piece that all of these trials should define, that is, in essence, a mono therapy piece. An important piece of information for treating people with the package insert would be: What can predict who would respond? That's very hard to do when you have multiple complications. That makes some lead-in period attractive.

When we thought about how to look at

slopes and how to look at responses over a short-term period, the problem is complicated by where people start because they don't start -- when they start at drug-naive, they start at some sort of baseline set point.

When they're starting on therapy, they're starting generally not at some baseline because we're not here. You're not talking about people who have no drug options. So they're partially suppressed.

So what you do with their old regimen and what they have been on at the time you do your resistance testing has a major impact on what happens. And it's extremely confusing because you get shift, and I don't know how fast shift reverts.

So if you want to talk about over two weeks, you take somebody that you're going to look at a new NNRTI -- I don't even want to get into whether two weeks is too long for that, but they have had shift. They no longer have the 103.

Is that the question that you want to ask in that population? We have some plans for these trials, and those are some of the questions that we look at.

To my mind, they are not answered. And so a very short lead-in, while it's extremely appealing,

is very difficult to interpret.

ACTING CHAIRMAN GULICK: Ms. Delph, the last word. Then we're going to move on.

DR. DELPH: I just wanted to comment on the proposal, what I understand to be Dr. DeMasi's proposal. I couldn't see it very well from here. So I may have misinterpreted it. But I have a lot of concerns about the bridged Phase I/III trial of moving from a situation where you are still trying to do dose finding and come up with an appropriate dose into treating a salvage population.

I think before we embark on salvage trials, we need to have a good idea of what dose of drug is likely to be effective. I think we also need to have adequate interaction studies done, drug-drug interactions and not just two-way interactions but three-way and four-way if necessary drug interactions, so that we have a reasonable idea beforehand of what drug level, what drug dosage, is likely to be effective in these patients. They have a lot to lose, and they have a narrow therapeutic window. They are very susceptible to toxicity, and they have highly resistant virus.

Having said that, a lot of the PK studies and dose-finding studies are done in very small

populations and populations that I think that are unlikely to be typical of salvage patients, who are particularly susceptible to toxicities and who have a number of comorbidities as well often.

So I do think that these studies should also include some form of drug monitoring. Now, I don't want to open Pandora's box of whether we're doing C_{max} or C_{min} or AUC or whatever. I think we may be able to discuss that sometime later, but I think we need to do some sort of monitoring of drug levels in these patients because we really at the end of the day cannot predict what sorts of drug levels we're going to get with the kind of regimens that we are going to be giving these patients.

I certainly share Dr. DeMasi's desire to get drugs to patients quickly, but we do want to do that in as safe a way as possible. I think that the patients who often most need these drugs, who are in deep salvage, I agree totally with Dr. Mathews that we should not be requiring these patients to get into trials to have access to them.

I think that we need to ensure that expanded access programs are opened up to patients who have no other options available at about a time that we're going into these larger Phase II/III studies.

And that I think is what we need to ensure.

ACTING CHAIRMAN GULICK: Thanks.

Let's go on to consider the modified --

DR. WONG: Could I just make a comment? I haven't said anything.

ACTING CHAIRMAN GULICK: Okay. Dr. Wong, who hasn't said anything?

DR. WONG: I haven't had a chance to answer the question at all. So I think that one of the things that comes across to me from the whole day's discussion is that in different groups of patients with highly experienced patients with HIV, the question is different.

And I think that each of these study designs and others that have been proposed really are capable of addressing different kinds of questions. In those populations, the right design to pick is the one that addresses the clinically relevant question.

I would just urge that the agency -- and I think they do this now, but I think I would like to hear them kind of reassert that they would consider approving a drug for a narrow indication if the utility of that drug for that narrow indication was established and not necessarily demand that the studies address all or even multiple possible

utilities so that whether or not -- I mean, clearly all of these designs have advantages and disadvantages, but I think each of them is clearly capable of addressing a clinically relevant question in some population. And I would suggest that that be the key consideration.

ACTING CHAIRMAN GULICK: Okay. Let's move to the modified factorial design, which is the first one up there. So we're talking about new therapies A, B, and C with optimized background A plus B versus A plus C versus B plus C versus all three together. Victor, can you start us off?

DR. DeGRUTTOLA: Well, I think that one of the issues here is that: If you have multiple drugs that you're interested in studying, can you look at combinations of those drugs right from the start?

The advantage of being able to do that is to be able to investigate interactions between drugs and also perhaps to increase your efficiency in answering multiple questions.

This so-called modified factorial is actually not in a factorial layout, but as I understand it, what it allows you to do is use half of your patients who are enrolled to answer each comparison of A versus B, A versus C, or B versus C.

And that provides some of the advantages of a full factorial but not all of the advantages in that you can't study all of the interactions and also you don't have the full efficiency.

I would say that the kind of design that you would want to use in this context would depend on what treatment options were acceptable to patients.

Let's say you had two new drugs, just A and B. If the layout of a randomization of A to placebo and B to placebo led to treatments that were acceptable, clinically acceptable, to patients, that would have an obvious advantage, full power to look at each one of the comparisons, and reasonable power to look at interactions, or at least the best power you could get to look at interactions. But obviously that's only appropriate in settings where it leads to treatments that are acceptable to patients and physicians.

In cases where it would not lead to acceptable treatments because giving patients only one new drug would be inappropriate for reasons that John Mellors mentioned, then I think that this design does offer some advantages, ability to look at some interactions in the start and some improvement in efficiency in terms of doing the comparisons of A

versus B, A versus C, and so on, even though not the full efficiency of a factorial design.

So I think it would be very specific to the setting and the particular treatments whether they could lead to acceptable regimens for patients and physicians.

ACTING CHAIRMAN GULICK: Dr. Saag?

DR. SAAG: This is what I was saying earlier. I think in keeping with what Victor just said, that it really depends on the population. Keeping with the definition that the group wanted to use, two HAART regimen is an all three class exposure at a minimum. Then you can go to a resistance test.

This doesn't all have to be one study, but it could be where you take patients who meet that criteria, you do resistance analysis. It could be with or without PK, et cetera. You can make it fancy if you want. But if there are three reasonable options or more, then you do the A, B, C versus placebo.

This is the clean study we have been talking about. This is where drug companies can collaborate and you can have three different companies with each of their agents, A, B, and C, and get something done that has a lot of meaning.

In that same population, if you have fewer than three options, then the optimized background therapy is not satisfactory. Just add a single agent. That's what we have been talking about. And that's where you go to the two drugs or maybe even three together.

And so the point is that, no matter what, the patients are getting three or four drugs that should work based on at least reasonable activity. And this is something you could use, I would think, as a registrational study.

I mean, the big picture, throwing it all together in one trial, feasible maybe. Maybe you want to split. That's with the blue line. Maybe you have one study like this and one like that, but the attractive thing from a clinical perspective is that when I'm recruiting for studies, more times than not it's really getting frustrating as an investigator because we have a lot of studies on the board.

If somebody comes in and with the new requirements, if there are no exemptions for anything, a lot of patients we sell the study concept to. And we find there is one part that kicks them out.

So the notion of having a large study that can attract a lot of patients, you don't have to worry

about that. You let the entry criteria define them into their arms, rather than out of the study. That would be attractive as an investigator.

The other part because we're talking about industry collaboration, -- and this is maybe where the ICC comes into play -- another limitation we have as in aside that will help speed this study to completion is that some of our folks, believe it or not, still can't get access to optimized background therapy, at least easily.

And so while it is difficult to have a whole pharmacy's worth of drugs for the study to complete, if we could find a way to provide the entire regimen, rather than just the investigational part, that would speed accrual and get the answers faster in certain populations of patients.

ACTING CHAIRMAN GULICK: Other points on factorial? Mr. Hogan?

MR. HOGAN: With regards to what Victor was saying about it not being a true factorial, that's true. What we are calling it in the Coalition for Salvage Therapy is a minimum number of factors factorial because we basically have just taken out all of the smaller cells.

It strikes me listening to this whole

discussion that the biggest challenge all of us face in the room, collectively and individually, is there are gray areas in two specific places. One is the tension between public health and drug approval. And the other tension is between drug access and research.

Unfortunately, I keep hearing the conversation flip back and forth where people are talking about what is the best care for their patient, as opposed to what is the best research.

I am certainly not advocating Dr. Mengele's solutions where we do the best research at the cost of care to patients. I think it is very important to separate out those issues.

In particular, one place where I think there was a confusion with the other respect is when people mention if there is interaction, a factorial study is not viable.

I believe it was either Victor or Courtney who pointed out: Well, yes, that's true if you want a scientifically rigorous examination of the components, but if you want to know which drug strategy is worthwhile, then it doesn't really matter. Whichever is the winner is the winner. So there is that public health/regulatory tension.

So I guess, really, what I am saying is I

think it is very important that we pay attention as we have discussion to discriminating on both of those fronts because I do think that all of these issues are getting mixed together.

I know when I discuss these issues, I mix them together all the time: patient care, research. Public health/regulatory is a very hard discussion to have.

DR. SCHAPIRO: But it is mixed together, but that is exactly the issue. I think what Mike is saying is we are investigators and we are doctors and there is a true tension there. If there wasn't, I mean that that's a real issue. It's not separated because it is not separable.

MR. HOGAN: But the discussion needs to acknowledge that.

DR. SCHAPIRO: I think we all accept that that if we could -- leaving Mengele out of this, I think we all realize that there is a tension in each one of us to do research for the benefit of all but not do harm to our patient. And that is why this continually gets mixed up because it actually is. You have one patient to make a decision on, and each patient is one patient.

ACTING CHAIRMAN GULICK: Dr. Mellors, then

Dr. DeGruttola.

DR. MELLORS: I don't spend my day contemplating clinical trial design, but we have taken an eight-cell factorial and conveniently removed what appear to be unattractive arms.

I am not convinced in looking at this design whether we have the same ability to discern individual drugs' contribution to toxicity when we don't have an arm without drugs in it. Still, I would like some clarification from the people who do spend their day thinking about these issues whether we lose anything by lopping off half of the cells.

ACTING CHAIRMAN GULICK: Dr. DeGruttola?

DR. DeGRUTTOLA: Yes. You clearly do lose something by lopping off the cells. It is exactly the clarity of being able to examine all of the interactions.

The simple basic factorial design of A versus placebo, B versus placebo if it is acceptable to patient is the best way I think to investigate interactions, both in terms of toxicity and efficacy. You can examine all of the effects, the main effects as well as the interaction.

In this design, which is four of the eight cells from a full factorial design -- it's not a

fractional factorial. There is a way of taking four of the eight cells in something that is called a fractional factorial that would maintain many of the benefits of the full factorial.

This is not a fractional factorial, but it still does allow you to investigate some interactions. So it still may be useful as a structure design, even though it is not optimal. It is not optimal in terms of power, and it is not optimal in terms of studying interactions.

But, for example, if that design were AB; AC; BC; and A, B, and C, supposing that you saw toxicity whenever A was combined with B, so you saw toxicity in the AB and the A, B, and C but you didn't see toxicity in the AC or the BC arm, then you might be able to make some investigation of the fact that your toxicities don't seem to be associated just with A, just with B because you do have those two arms alone but only in arms that have A and B occurring together.

If, on the other hand, you saw the toxicity whenever you had B, whether or not A was there, then that might give you more reason to believe that it was B that was contributing to that effect. But the fact that you don't have the full factorial,

just as John Mellors states, does mean that you don't have the full power to investigate all of the interactions.

There is certainly a trade-off. The only reason for doing this design, less desirable statistically, is if it were the only one that came up with acceptable treatment options, you could still learn a lot from it.

MR. HOGAN: Just one point of clarification. Every drug is absent in a least one arm.

DR. DeGRUTTOLA: Well, that's different from the design. I was referring to this design here, which is the AB; AC; BC; or A, B, C. So I think you may be talking about a different --

MR. HOGAN: AB; BC; A, B, C; correct?

DR. DeGRUTTOLA: And AC.

MR. HOGAN: Right. So every drug is absent from at least one arm.

DR. DeGRUTTOLA: What is absent from the A, B, C arm?

MR. HOGAN: No, no. I am saying there is at least one arm from which every drug is absent.

DR. DeGRUTTOLA: Yes. That's right. That's right.

MR. HOGAN: So that, admittedly, you're not in a great power situation, but you always have a comparator to an arm that does not have that drug present.

DR. MELLORS: But if there are disease-related symptoms or signs that have nothing to do with the drug, you won't be able to discriminate that from drug toxicity.

ACTING CHAIRMAN GULICK: Okay. I think the points have been made. I think I would like to stop the discussion at this point. With apologies to a couple of other people who wanted to present other designs, I think we need to move on.

Just to summarize what we have said, as a group, we struggled once again with the population that we are considering, what constitutes experience, what to do about people with some options versus no options.

A very important point about viral and CD4 setpoints and how that is influenced by people taking their regimens going into one of our salvage studies.

In terms of optimizing background, some important practical points about access to therapies, how resistance tests are done, access to resistance tests, how they're interpreted, and then the important

suggestion that possibly stratifying by options that are identified by resistance testing might be an appropriate thing to do at baseline.

Also, we heard a plea again that before launching into any of these studies, that drug-drug interactions have to be known, particularly with the optimized background design given that multiple drugs may be used.

In terms of the simple add-on design, people were impressed that that's the simple, cleanest way to demonstrate efficacy and perhaps would be most attractive to industry. Also, it was noted that drug interactions would be relatively straightforward to define.

In terms of the negatives, essentially this is a design of functional mono therapy, which is of obvious concern, may predispose to early failure. And having a placebo control may be a relative negative for patient participation.

The second design we thought about was the two-part, two-phased, bridged, dimorphic, two-bit design.

(Laughter.)

ACTING CHAIRMAN GULICK: That's the compromised title.

In terms of the pros, people once again felt that this was a design which could cleanly demonstrate efficacy over the short term, also had the opportunity to do dose ranging, and to identify important pharmacokinetic properties. In addition, the second phase helps us to define longer-term safety in a larger number of patients.

In terms of negatives, the logistics was probably the top one that came up; also, the risk of resistance in the first phase of the study. And it was recognized that that would differ depending on the specific agent that was used.

In terms of the factorial design or the modified factorial design, the big plus here is using combinations of therapies which would be attractive to patients. That interactions between these drugs could be defined earlier on was another big plus.

In terms of the negatives, teasing out the specific activity of individual drugs and the specific safety issues of individual drugs, at least from a combination deign, might be challenging; finally, the very practical point about the number of drugs available for this kind of design.

And then some other interesting suggestions that came up in the discussion, we had a

discussion of blinding, whether that's appropriate or not. The nice option of an early escape or an early switch for predefined failure would be an attractive strategy to use. Concentration-controlled was another novel approach that came up; and, finally, a second randomization in a two-phased design, also a novel way of doing it.

I think I would like us to take a ten-minute break. And we will reconvene for the afternoon. Thanks.

(Whereupon, the foregoing matter went off the record at 3:28 p.m. and went back on the record at 3:38 p.m.)

ACTING CHAIRMAN GULICK: In conferring with the agency, they have assured me that they know the answers to Questions 3 and 4. And so we don't have to dwell on those. So we will move into the second part of the discussion today, which is a discussion on endpoint issues.

Our first speaker is someone whose career I have watched for a very long time. It is actually me.

Thank you, Mr. Chairman.

(Laughter.)

II. ENDPOINT ISSUES

RESPONSE RATES IN HEAVILY PRETREATED

HIV INFECTED PATIENTS

ACTING CHAIRMAN GULICK: I would like to speak about the response rates in heavily pretreated patients. This slide summarizes results from five clinical cohort studies, all of which were published in 1999. They're from throughout the U.S. and western Europe, cohorts from Amsterdam, Cleveland, Johns Hopkins in Baltimore, the Swiss cohort study, and UCSF.

What these cohorts have in common is that they are all taken from clinics, rather than clinical trials. In most cases, these were nucleoside-experienced patients who began their first so-called HAART regimen, usually defined as adding a protease inhibitor.

You can see that the numbers of patients involved are quite high. I have simply summarized what the percent of patients who started this so-called HAART regimen was above the limits of detection, which varied from cohort to cohort and the times in follow-up.

Just to make a long story short, what you see are clinical failure rates between 38 percent and as high as 63 percent in clinical cohorts over a

period of one to 2 years on our best first line therapies.

In looking at these five studies as examples of clinical cohort studies, there are predictors of virologic failure that jump out from study to study. And many have identified the same predictors of virologic failure.

Prior antiretroviral treatment, a small percentage, approximately 20 percent in each cohort, were antiretroviral-naive. That was an important predictor of virologic success.

The viral load level looking at either a higher baseline or a higher peak level was often predictive. Looking at the CD4 cell count, either a lower baseline or a lower nadir was predictive of virologic failure.

From study to study, the specific antiretroviral regimen used was important. In one study using a new nucleoside was a factor associated with a better response; in another, using a non-nucleoside in a naive population. And in a third study, the use of the protease inhibitor saquinavir was associated with a hither virologic failure rate.

Finally, in two of the five studies, more missed clinic appointments as a surrogate for

adherence was a predictor of virologic failure.

Moving to the two prospective studies of genotypic resistance testing, we can add several more factors that predict virologic response. From the GART study recently published in *AIDS in 2000*, we see that the number of active drugs picked in a salvage regimen as you move from four drugs is associated with a higher virologic response rate in terms of measuring change in HIV RNA over the course of the study.

Secondarily, what is noted is that more patients who had genotypic testing were able to come with a regimen with three or four new drugs, as opposed to patients who did not have the benefit of genotypic testing.

In a similar study, the VIRADAPT study, not only did they look at the use of genotypic resistance testing, but in a subset of patients, they also began to define drug concentration as an important predictor for virologic response.

They took three specimens at random clinic visits and simply measured the concentration of protease inhibitor in those three visits. They classified patients according to a dichotomous variable. Either they had SOC suboptimal concentration of their protease inhibitor or they had

an optimal concentration and then using a factorial design, really, patients, of course, who had been randomized either to genotypic testing or what was standard care at the time, which was no genotypic testing.

Breaking the patients down into four groups and relating these two factors to the change in viral load, you can see that the group that does the best in terms of the biggest decrease in viral load level is the group with optimal concentrations of their protease inhibitor and genotypic testing.

The group that does the second best is the group that simply had optimal concentrations of their protease inhibitor. And then the two groups with suboptimal concentrations, either with or without genotypic testing, had relatively less virologic suppression.

Now, there are limitations to looking at clinical cohort studies, particularly the ones that I have just considered. One thing is an issue that we have been struggling with all morning. That is heterogeneous patient populations. Some patients are naive. Some patients have taken nucleosides. Other cohort studies combine many different types of patients. And that makes it difficult to tease out

the most important factors in predicting virologic response.

The studies I have just shown you really reflect antiretroviral use in the years 1996 to 1998. So there is really a time bias already because in the last two years, things have changed, even since the mid to late '90s.

For instance, in that time, there were fewer antiretrovirals available. More complex regimens were routinely used, perhaps involving 12 to 15 pills a day and using Q-8 drugs routinely. And then sequential mono therapy was quite a common way to use the drugs as they were approved one at a time.

Another issue that I think came out with these initial studies was the fact that people were saying there are high virologic rates in the clinic. And that often led people to say that treatment failure rates are quite high in the clinic.

I think that simplifies what's really a relatively complicated concept. That is, is virologic failure, which is what is measured in those cohort studies, really the same or related to immunologic failure or ultimately clinical failure? And what do you really mean when you're saying "treatment failure," which one of the three of these? And,

really, that is what we are considering in this section of the panel discussion, the endpoint choice.

Well, of course, over the last several years, we have come to conclude that virologic failure does not necessarily mean either immunologic or clinical failure. This is work from Steve Deeks published in *Journal of Infectious Disease*. Looking at a cohort of 380 patients in San Francisco, they were classified according to their virologic response rates into one of four groups on therapy.

A responder group had consistent viral load levels less than 500. A partial responder group had levels above 500 but greater than a one-log decrease from baseline.

A transient transponder had at least a one-log decrease in viral load but then a rebound in viral load levels. And, finally, a fourth group, non-responders, essentially didn't meet any of those criteria, had no measurable change or no significant change in viral load levels.

What Dr. Deeks did is to relate in each of the four groups the virologic response to the change in CD4 cell count over the course of two years and follow-up.

What you can see is that three of the

groups, both the full responders, the partial responders, and even the transient responders, have significant increases in CD4 cell counts from 100 to 200 cells over baseline through the end of 2 years. The only group that didn't really have a significant rise was the complete non-responders.

So, even in the presence of transient or partial virologic response, there was an immunologic response, at least in terms of CD4 cell count change, from baseline demonstrated out to two years and follow-up.

The French cohort looked at a very much larger group of patients, over 2,000 patients, and began to relate both virologic and immunologic responses to clinical responses. So looking at all three endpoints, they classified the patients into one of four groups based on their response to their antiretroviral regimens at six months of time.

I should say that, once again, as in Dr. Deeks' cohort, about 80 percent of these patients were nucleoside-experienced upon starting their new so-called HAART regimen.

So they classified patients either with a virologic response, which in the case of the French study was either a viral load level less than 1,000 or

at least a one-log drop from baseline. That's considered a virologic response and then an immunologic response, which was considered to be a CD4 cell count, at least 50 over what their baseline levels are.

They divided people into four quadrants, basically either an immunologic response or a virologic response, positive or negative. And those are the four groups here.

In terms of patient numbers, of these roughly 2,000 patients, half were both immunologic and virologically responsive. And the other half were divided between the other three categories roughly more or less equally, so about a sixth of the population falling into each. That's at six months. What we are relating that to here is the percent alive and AIDS-free, so a composite definition in terms of clinical progression or death.

What you can see once again here is that the three groups who have either both an immunologic and a virologic response or one or the other have a better clinical response *in toto* than the patients who have neither an immunologic or a virologic response.

One of the largest cohorts to date to look at this issue and really tease out virologic response

and other immunologic and clinical endpoints is the EuroSIDA cohort. This was presented at the third Salvage Workshop earlier in the year 2000.

The EuroSIDA cohort studies well over 8,000 patients and now has described what their responses are. In terms of the cohorts, one group began their first HAART regimen, their second, and a third.

And what you need to know about this is how they define that. Moving from first HAART to second HAART does not necessarily mean failure of the regimen, but it means a change in therapy after at least one month on the first regimen. So that could be from failure of the regimen. It could also be from either adherence or toxicity, but it's a real world look at how successive HAART regimens do in terms of response.

In terms of virologic failure, which they define as a viral load documented at over 500 copies per mill, in terms of the first HAART regimen, 40 percent experienced virologic failure; the second HAART, 50 percent. And on the third HAART regimen, 67 percent experienced virologic failure on their regimen. And I should say this is at the two-year time point.

Looking at a combined endpoint using both an immune criteria, which was a return in CD4 cell to baseline levels, or a clinical failure, which in this case was an AIDS-defining illness or death, you can see the percentages, fewer than virologic failure but still increasing with successive regimens, going from 20 percent to 30 to 40 percent in the third HAART group.

And, finally, teasing out the specific clinical events, which are in some cases, of course, what we are most concerned about, what you see is only 5 percent over 2 years experienced a clinical event with their first HAART regimen, but this jumps up to about 25 percent in the second and third regimens.

So this is illustrative of the fact that many patients are experiencing failure, whether it is virologic, immunologic, or clinical, but the timing for the failure differs between the three cohorts as you progress in the number of HAART regimens that you have taken.

I think it is useful in thinking about so-called salvage treatment -- I, too, don't like the term, but we use it as shorthand -- to look at some examples of salvage studies that have been done and in some cases even published over the last couple of

years. This is a short list. What these trials have in common is they are not for the heavily treatment-experienced group but for first failures.

Next. Arguably, the first salvage study ever done was ACTG 333. This looks specifically at a saquinavir-experienced population. Patients that had taken at least 48 weeks of saquinavir hard gel, no other protease inhibitor. And, importantly, where we are in antiretroviral therapy for the design of the study, they were not allowed to change their background therapy two months prior to coming onto the study. Obviously we wouldn't design it that way today.

Seventy-two patients were enrolled and randomized either to continue the saquinavir hard gel, switch to the saquinavir soft gel, or switch to indinavir.

This particular study underwent interim analysis at eight weeks and looked at three different endpoints: change in viral load, the percent of patients less than 200, and the CD4 change.

Basically the two saquinavir arms showed no significant difference. With the exception of a 37-cell CD4 rise, the indinavir arm showed a .6-log drop in viral load change and a 37 percent drop in

less than 200 copies, which was the limit of detection on this study, but that was felt to be a blunted response.

In subsequent analyses, they related the pre-existence of resistance mutations to saquinavir and indinavir as being highly predictive of virologic outcome.

ACTG 372B is a good example of a salvage study in that it was designed when the parent study, which was ACTG 320, was still in progress. This study is notable because of the fact that people realize that salvage therapy options for patients failing first line regimens were going to be important to identify.

So patients were on ACTG 320. They received either AZT or d4T in combination with 3TC and indinavir and had virologic failure, defined here as a viral load level greater than 500 copies. You can see 84 patients participated.

In this salvage study, patients were given what at the time were all investigational agents, efavirenz, adefovir, and then they were randomized using a factorial design to receive either abacavir or a new nucleoside and then plus/minus nelfinavir, so several different questions trying to be answered

simultaneously in this salvage study.

The bottom line -- and I should say this study focused on the percent below detection, which again was state-of-the-art for therapy in naive patients and, thus, at least in the beginning, for salvage therapy trials, too. Overall a disappointing only 35 percent had HIV RNA levels less than 500 at the week 16, relatively short-term, time point.

Using the factorial analysis, this study showed no difference between adding abacavir or one to two nucleosides but did show a benefit of using nelfinavir in the salvage regimen over a matching placebo, 45 percent versus 24 percent, which was statistically significant.

Recently published by our group is ACTG 359, which was probably the first large salvage study, which sought to look at a particular patient population, the indinavir-experienced group. Patients were required to have taken 6 months of indinavir, had a limited viral load level between 2 and 200,000 were naive to other protease inhibitors and, importantly, non-nucleosides, nearly 300 patients were randomized into the study.

They were randomized to receive saquinavir soft gel in combination with either ritonavir or

nelfinavir. So this will be the first salvage study to really look at double PI combinations and then again in a factorial analysis, either the non-nuc to delavirdine, the nucleotide to adefovir, or both drugs together.

Bottom line and once again focusing in this trial on the percent below detection as being the primary endpoint, a disappointing 30 percent of patients reached less than 500 copies per mill at the primary study endpoint time, which was week 16.

In factorial analysis, there was no significant difference between the use of saquinavir-ritonavir or saquinavir-nelfinavir. However, patients who received delavirdine or delavirdine and adefovir had better virologic response rates than those who received adefovir by themselves.

In terms of longer follow-up, I can tell you that we will be presenting a poster at the upcoming antiretroviral meeting looking at the long-term changes in both viral load and CD4 cell counts on this salvage study.

Lastly in this group of studies is the Abbott 765 study, one of the first to really look at not just a specific protease inhibitor but any protease inhibitor experience. This is also an

example of a study that takes a new agent and seeks to look at its activity in a salvage-type population. They did this by being very exclusive about the entry criteria.

Patients could only fail one protease inhibitor, needed to be non-nucleoside-naive, and have viral load levels between 1,000 and 100,000. So this was really a pilot Phase II-type design.

Patients swapped lopinavir-ritonavir at one of two doses for the current protease inhibitor they were taking while keeping their background antiretrovirals constant for two weeks and then at the two-week time point all added novirapine and they were all NNRTI-naive and optimized their nucleosides; that is, they got to choose one or two new nucleosides. And they were followed for up to 96 weeks.

Next slide. So this led to a chance to look in the first two weeks at what the switch from the protease inhibitor that they had failed according to criteria to the new protease inhibitor, lopinavir-ritonavir. And what we were seeing over the first two weeks was over a log drop simply with that change.

So, in keeping with what we were discussing earlier, here is a design which looked at

the activity of switching one agent in the regimen for two weeks and showed significant antiretroviral activity. Then when the other changes were made, novirapine added, nucleosides changed, patients had a persistent suppression of viral load levels.

And, next slide, that translated to a significant proportion of patients dropping their viral load levels less than 400 in an intent-to-treat missing-equals-failure analysis. This is both doses, but they have very similar activity with roughly 60 to 65 percent reducing viral loads below the level of detection for up to 96 weeks of follow-up.

The next step from the first failure studies was to look at patients who had failed more than one protease inhibitor. And just because it has a very similar design to the study I just showed you, the Abbott 957 is similar in many ways except that it allowed at least two or more than two protease inhibitor experience with virologic failure as evidenced by a viral load level greater than 1,000.

Importantly in this study, once again, patients were required to be non-nucleoside-naive. Once again, they added lopinavir-ritonavir. And here is a good example of a drug-drug interaction, which was described while the study was going on and then

dose adjustments were made, a significant interaction between efavirenz lowering the concentrations of lopinavir in the presence of ritonavir and a boosting of the dose of lopinavir-ritonavir while the study was going on.

Once again, with all of those innovations, a significant number of patients dropping viral load levels, again, important to remember that they added efavirenz and were all non-nucleoside-naive but at the higher dose, in particular, shown in pink, 70 percent of patients reducing their viral load levels below 400 for up to 48 weeks.

Okay. But all of these patients don't really meet the definition that we have been considering. These are the first iterations of trials, but what we have been considering most of the day are heavily pretreated patients.

The definition which we have talked about earlier today is having a loss or lack of virologic response on at least two HAART regimens and being three-class-experienced. So what is the data for this particular group?

While we go back to the EuroSIDA cohort, the first thing to know is that in 1996 in this cohort, there were no patients who had three-class

experience. In the year 2000, 35 percent of the EuroSIDA cohort are patients who have had three-class experience with the current drugs.

This is looking at a subset of those, 266 patients who had 3-class experience and began another new salvage regimen. So this is addressing the fact in a clinical cohort: How do these three-class experience patients do?

Forty percent decreased their viral loads less than 1,000. And 30 percent maintained that decrease for as long as 6 months. Fifty-five percent had at least a one-log decrease and most of them, 45 percent, maintained this decrease at 6 months. Looking at it another way, 55 to 70 percent of these 3-class-experienced patients had virologic failure by the end of 6 months.

What about other endpoints? Over half, 55, decreased their CD4 below baseline. So if you use that as an immunologic criteria, you would say that they had immunologic failure and that was at the ten-month time point. But only five percent had a new AIDS event or death; that is, experienced a new clinical endpoint over the ten months of the study. That gives us some idea of how this patient group does in terms of those three endpoints: virologic,

immunologic, and clinical.

What were the predictors in the EuroSIDA cohort? In terms of virologic response, any prior viral load less than 500 was predictive of virologic response, less prior treatment, a higher latest CD4 cell count, and, for reasons that aren't clear to me, being a resident of Central Europe was predictive of a virologic response. Perhaps our European colleagues could comment on that.

In terms of predictors of immunologic and clinical response, having female gender was a predictor. A lower latest viral load, and fewer prior antiretrovirals were all predictive in this three-class experience group.

What about clinical trials in this group? There are two notable studies that have looked at highly treatment-experienced patients. The first was CNNA 2007. This has been presented nationally both by Drs. Eron and Falloon. It looked at patients who had at least 20 weeks of combination therapy with a protease inhibitor and a viral load level of at least 500. And there were 99 patients.

This is one of the most treatment-experienced populations ever enrolled into a clinical trial. Three-quarters had experience with

4 to 5 nucs, 44 percent with a non-nuc, and 60 percent had taken 3 to 4 protease inhibitors prior to going on to the study.

They received what at the time again were three investigational meds, now all approved, both in label, abacavir, efavirenz, and amprenavir. The primary endpoint for this study was antiviral effect at week 16.

Overall, once again, a disappointing result with only 26 percent, dropping viral load levels less than 400 at week 16. Also, a significant drug interaction was identified between efavirenz and amprenavir on the study, which may have contributed to that somewhat disappointing result.

And then in a subgroup analysis looking at NNRTI experience and viral load levels, not surprisingly, it was found that the subgroup that did the best were those who were NNRTI-naive. Recall that they got efavirenz and those with the lowest baseline viral load levels.

Probably the largest salvage study done to date is ACTG 398 presented by both Dr. Hammerstrom and Dr. Mellors at various meetings. This looked at people who had taken at least four months of up to three prior PIs, viral loads over 1,000.

They could have taken NNRTIs, although it wasn't mandated that they take them and nearly 500 patients enrolled in the ACTG. Once again, the approach here was to give all new drugs. Every one received open label amprenavir, abacavir, efavirenz, and adefovir together with in three out of four groups a second protease inhibitor, either saquinavir soft gel, indinavir-nelfinavir, or a matching placebo. And the primary endpoint looked at was week 24.

Overall, once again, a very similar and disappointing 31 percent dropped their viral load levels below 200 at week 24. And then using subgroups, NNRTI-naive subjects did better.

Patients who took two protease inhibitors did better than those who took amprenavir alone. And, surprisingly, patients who had experience with just one or more than two protease inhibitors did about the same.

Importantly also for ACTG 398 -- and I don't have the slide to illustrate it -- is that they focus not just on the percent below detectable but the change in viral load levels.

At the end of week 24, all groups had approximately at least a one-log decrease in viral load levels from baseline. That's important.

Particularly if you focus on this number, overly focus on that number, you might miss that important point.

Well, there are examples of just what we have been considering today. How do you test a new drug in a treatment-experienced population? I'm going to show you three examples.

The first is using the investigational nucleoside analog DAPD. These are results, once again, that Dr. Eron has presented. It looked at a study population who failed either ZDV or d4T plus 3TC, had specific viral load levels and a CD4 count above 50.

This is really on the pilot study level. It's a sample size of 24; however, once again, a very treatment-experienced group. Average number of antiviral is six, length of treatment four years. A hundred percent had taken nuc, 60 percent or more non-nucs, and over 80 percent had taken protease inhibitors.

They were randomized to receive DAPD at one of three doses. Three groups actually washed out of their antiretroviral seven days prior to going on.

One group added DAPD onto the regimen that they were already taking, so short-term virologic effect being demonstrated in a group of highly

treatment-experienced patients. And that is what is shown for you here.

So three doses in the patients that washed out. You could see at a higher dose, about a one-log drop. Surprising to investigators was the fact that the group that added on, as opposed to experiencing a seven-day washout, had a much more profound drop in viral load, about two logs below baseline, by the end of two weeks.

The next example is the investigational nucleotide analog tenofovir. This is the Gilead 902 study. This is an example of the add-on design that we discussed earlier, patients on stable antivirals with a viral load of at least 5,000.

Nearly 200 patients enrolled; once again, a very treatment-experienced group. Baseline mutations in nearly all for nucleosides, a third for non-nucs, and about 60 percent for protease inhibitors. They were randomized to add one of three doses of tenofavir or a matching placebo.

Here is a nice dose-response curve. Placebo is in blue with very little change through 24 weeks, the greatest change is at the highest dose of tenofavir. This is sustained at about .7 logs by the end of 24 weeks. At this point, the placebo group is

also offered in a crossover design the highest dose of tenofavir demonstrated.

The other endpoint looked at was CD4 cell count, for reasons that aren't completely clear, relatively low CD4 response by the end of 48 weeks.

The last example of using an investigational agent in a highly treatment-experienced group is the investigational fusion inhibitor T-20. This is the T-20 205 study.

This is actually a very novel design. The studied population was those who had T-20 experience in one of the earlier studies that the drug company had sponsored. So they were offered this study as a rollover study.

Seventy-one patients entered. Again, incredibly treatment-experienced, 80 percent were this specific patient group we have been considering 3-drug class-experienced.

They underwent baseline genotyping and were allowed to add T-20, 50 milligrams twice daily subQ plus the other antiretrovirals, which were chosen by their practitioners on the basis of history and the genotyping or what you might call optimized baseline regimen and then followed for 48 weeks.

In terms of results, 20 percent had less

than a .5-log reduction, so were essentially virologic non-responders. However, a third had at least a one-log reduction or reduced their viral load levels to less than 400 using this strategy.

This is actually an on-treatment analysis of change in viral load, rather than an intent-to-treat. It goes out to week 32 presented at the Durban AIDS conference. You can see significant virologic suppression sustained in this highly treatment-experienced group with this novel design.

In summary, in terms of salvage therapy, what we have been saying all day, virologic failure occurs commonly. Immunologic and clinical failure also occur at probably different rates and different times. And all need to be evaluated potentially as endpoints.

Predictors of response: not surprisingly, adherence, levels of viral load and CD4, resistance profile, number of active drugs and drug levels all may come into play.

Importantly, newer drugs with novel resistance patterns or mechanisms can and do demonstrate an activity, even in the heavily treatment-experienced patient population. And novel study design may demonstrate this activity and at the

same time provide benefit for the subjects, the conflict we have been talking about all day. And further clinical research is necessary. I'll stop there.

Thank you, Dr. Gulick, although you did run on a bit long.

(Laughter.)

ACTING CHAIRMAN GULICK: I would like to introduce Dr. DeGruttola from the Harvard School of Public Health to talk about statistical considerations.

STATISTICAL CONSIDERATIONS FOR ENDPOINTS IN HEAVILY PRETREATED PATIENTS

DR. DeGRUTTOLA: Well, given this morning's discussion and in the spirit of full disclosure, I want to let everyone know that I am myself a two-part hybrid because, in addition to working in AIDS research, I also teach statistical design for graduate students in my department.

I would be happy to report to them that the phrase "factorial design" has had almost as much impact here as "pregnant Chad" has in Florida. If anyone says this is just a narrow technical area, I think I will have a good rejoinder.

I'm going to talk a little bit about

choice of endpoints for salvage studies. And, just to review quickly the endpoints that have been used, as everyone is aware, there are clinical endpoints which we used earlier on in the epidemic, may make a comeback, AIDS-defining events, survival, quality of life; marker-based endpoints, like HIV RNA and CD4.

Endpoints, of course, for toxicity can be time to treatment discontinuation or targeted adverse events. Finally, there are composite endpoints that combine information across different endpoint categories; for example, time to treatment discontinuation, whether for virologic failure or intolerance.

We look first at the HIV RNA endpoints. Even within that group, there is quite a range. There are quantitative endpoints, change from baseline to week X; time to virologic failure variously defined; or we have binary endpoints. We could have one that is just cross-sectional, like either above or below a threshold at week X, or a more cumulative one: Have you failed by week X?

So if we look first at the cross-sectional endpoint, it is a snapshot. It is not affected by to-be-transient changes in HIV levels and frequent monitoring is not required. So those are advantages,

but a problem is that missing data at the time point where you're doing the measurement is especially problematic because you can't make use of the other information unless you carry values forward. And I will say a little bit more about that, as Dr. DeMasi mentioned.

Failure endpoints, of course, require assessments over time. They may be affected by transient changes in HIV RNA levels, and frequent monitoring is required.

Although you need to define your missing data strategies, you can more easily make use of partial information if you use a failure endpoint and also if after patients fail they can go on to some other treatment, that doesn't complicate an analysis in a time to failure because you have already got your endpoint; whereas, if you had a snapshot change by a certain period of time, then that would be a complicating failure.

Now, within the failure group, you can either do a time to failure or a cumulative proportion analysis. The time to failure, some of the concerns are the pattern of failure depends on the failure time.

In other words, we have definitions of

failure for people who fail within the first four or eight weeks of not having declined enough or later on perhaps of having a rise above nadir or rise above detection and so on.

What we mean by "failure" is different at different times. There are some inherent assumptions there, but the advantage of time to failure is it accommodates differential follow-up. And it is very useful if you are doing an interim analysis to be able to make use of the partial follow-up as well as, of course, at the end of the study.

The cumulative proportion has the advantage that you don't have to worry so much about the definitions of failure at each individual time point because you can sort of make use of the whole trajectory in making that consideration.

So, although you need failure definition whether you do cumulative proportion or time to failure, the cumulative proportion is not quite as sensitive to the exact failure definition as the time to failure might be.

The problem is that evaluation within an interim analysis is complicated. There are also some power advantages of time to event. Especially if the pooled failure rate between the arms is greater than

50 percent, time to event has appreciable advantages.

Just focus on the bottom line. If you had a study with a one-year accrual, six months of additional follow-up and a two-arm trial, then look at the bottom line with a 70 percent pooled failure rate. You get a 25 percent savings in sample size, which is considerable. So there are real power advantages of time to event if the events are quite frequent.

Some analysis issues. If you have moderate study withdrawal, time-to-event endpoint advantages increase further. These sample size advantages are even greater at interim analyses. You're much more likely to stop in an interim if you use time to event because you make better use of your information. There are also advantages for co-variate, evaluating co-variate, effects or flexibility in ending the study.

So the only real down side is that you are a little more sensitive compared to cumulative proportion on the precise definition of your failure.

And if we compare a purely virologic versus a composite endpoint, a purely virologic focuses only on the virologic response and allows tolerability and safety to be assessed separately. But, of course, you have to follow up for viral load

after treatment continuation.

A composite, which hasn't really been discussed a lot here -- so I am going to go through this quickly. A composite endpoint might combine virologic efficacy, tolerability, and safety. And, of course, it could differ substantially from the purely virologic if the toxicity rate is high and if you use that endpoint, you would at least want to do the virologic, pure virologic, as a secondary endpoint.

Some definitions, the issues in defining virologic failure, obviously you have to define an early failure, whether you're talking about insufficient decline, rise above nadir, the amount of time allowed before patients go below threshold, the choice of the threshold for suppression/loss of suppression.

Of course, what do you do about fluctuations due to treatment holes into current illness and so on? Patient stops drug, virus comes back, restarts, and they're doing well. Was that a failure or not, all those kinds of issues?

If you use a regimen completion endpoint, then you have all of the same issues, plus you have to worry about how many drugs need to be changed or added before you declare it was a failure? And are you

going to call all discontinuations failures or just those related to toxicity?

One of the factors that are going to affect the choice of endpoint is: What are the underlying clinical beliefs? In purely virologic endpoint, you believe that the effective therapies on RNA capture the essential information to define the role of the therapy. For the regimen completion, you believe the necessity to change regimens more closely measures tangible benefit. And the choices are obviously going to depend on the clinical beliefs.

The next slide just shows some of the types of endpoints that have been used in studies within the AIDS Clinical Trials Group. There have been, as you can see, a variety of choices, including the time to failure, regimen completion, the cumulative proportion failed by week X, and whether patients have gone below a threshold by a certain time point, and so on. There have been a range of different endpoints.

Now, a little bit more about the composite endpoints. They are going to be more numerous than a purely virologic endpoint, but they can dilute the effect of a treatment and especially concern if treatment discontinuation might be unrelated to the