U.S. DEPARTMENT OF HEALTH AND HUMAN SERVICES

FOOD AND DRUG ADMINISTRATION

CENTER FOR DRUG EVALUATION AND RESEARCH

ANTIVIRAL DRUGS ADVISORY COMMITTEE MEETING

Wednesday, August 20, 2003

8:05 a.m.

Versailles Ballroom
8120 Wisconsin Avenue
Bethesda, Maryland 20814

C O N T E N T S

C O N T E N T S (Continued)

C O N T E N T S (Continued)

P R O C E E D I N G S

Call to Order

DR. GULICK: Good morning. I'd like to welcome everyone to today's meeting of the Antiviral Drugs Advisory Committee for the FDA.

I am Trip Gulick from Cornell in Manhattan.

We would like to start by introducing the members of the Committee, so if each member could state their name and their affiliation.

We'll start with Dr. Brown.

Introduction of Committee

DR. BROWN: My name is Ken Brown. I am representing industry. I am on the faculty at the University of Pennsylvania.

MS. HEISE: My name is Lori Heise, and I direct the Global Campaign for Microbicides, and I am the Consumer Advocate.

DR. STEK: Alice Stek. I am an ob-gyn on the faculty of the University of Southern California.

DR. HAUBRICH: Richard Haubrich from the University of California at San Diego. I mainly do HIV clinical trials.

DR. PAXTON: Lynn Paxton. I'm a medical

epidemiologist at the Centers for Disease Control.

DR. FLORES: I am Jorge Flores, of the Vaccine Clinical Research Branch at the Division of AIDS, NIH.

DR. BARTLETT: I am John A. Barlett from Duke University Medical Center.

DR. WASHBURN: Ron Washburn, Infectious Diseases, LSU, Shreveport.

DR. MATHEWS: Chris Mathews, UC-San Diego.

DR. FLETCHER: Courtney Fletcher, School of Pharmacy, University of Colorado Health Sciences Center.

MS. TURNER: Tara Turner, Executive Secretary for the Committee.

DR. STANLEY: Sharilyn Stanley, Associate Commissioner, Disease Control and Prevention, Texas Department of Health.

DR. SHERMAN: Ken Sherman, University of Cincinnati, Division of Digestive Diseases.

DR. WOOD: Lauren Wood, HIV and AIDS Malignancy Branch, NCI.

DR. ENGLUND: Janet Englund, Children's Hospital, University of Washington Seattle.

DR. DE GRUTTOLA: Victor De Gruttola, Department of Biostatistics, Harvard School of

Public Health.

DR. FLEMING: Thomas Fleming, Chair, Department of Biostatistics, University of Washington, and Co-Director of the Statistical Center for the HPTN.

DR. BHORE: Rafia Bhore, Statistician, FDA.

DR. WU: Teresa Wu, Medical Officer, FDA.

DR. BIRNKRANT: Debra Birnkrant, Director, Division of Antiviral Drug Products, FDA.

DR. COX: Edward Cox, Deputy Director, Office of Drug Evaluation IV.

Conflict of Interest Statement

DR. GULICK: Thanks.

Tara Turner will now read the Conflict of Interest Statement.

MS. TURNER: "The following announcement addresses the issue of conflict of interest with regard to this meeting and is made a part of the record to preclude even the appearance of such at this meeting."

"The issues to be discussed at this meeting are issues of broad applicability. Unlike issues in which a particular sponsor's product is discussed, the matters at issue do not have a

unique impact on any particular product or manufacturer but rather may have widespread implications with respect to all topical microbicides for the reduction of HIV transmission and their sponsors."

"To determine if any conflicts of interest exist, the participants have been screened for interests in topical microbicides for reduction of HIV transmission and their sponsors. As a result of this review, it has been determined that no reported interests present a conflict of interest or the appearance of such at this meeting."

"In the event that the discussions involve any other issues not already on the agenda for which an FDA participant has a financial interest, the participant's involvement and exclusion will be noted for the record."

"With respect to all other participants, we ask in the interest of fairness that they address any current or previous financial involvement with any firm that is developing or studying a topical microbicide for the reduction of HIV transmission."

Thank you.

DR. GULICK: Thank you.

Now we'll turn to Dr. Birnkrant for some opening remarks.

Opening Remarks by Dr. Debra B. Birnkrant

DR. BIRNKRANT: Good morning. Before I get to my opening remarks, I would like to take this time and opportunity to thank some members of our Committee who will be rotating off.

The first person the Division would like to thank is Dr. Courtney Fletcher, who has served on our Antiviral Drugs Advisory Committee through many complicated meetings, and he has served the term from March 2000 until October of this year. We want to thank him for his contributions to the Committee.

Next, I'd like to thank Dr. Sharilyn Stanley, who has also served from March 2000, and her term ends October 31, 2003. We want to thank her for her comments and help during many complicated Advisory Committee meetings.

Thank you very much.

And lastly, I'd like to thank Dr. Chris Mathews, who has served also on the Committee since March 2000. We are happy to have him here today as he ends his term as of October 2003.

Thank you.

[Applause.]

DR. BIRNKRANT: With that, I would like to welcome our Advisory Committee members, guests, and consultants to today's meeting on topical microbicides. This is a landmark meeting because this is the first time we are bringing this topic to the Committee in a public forum--although actually, we have been working on this area for more than 10 years as an agency.

This tells you how complicated the field is. Another example of how complicated the field is relates to the history of N-9. Nonoxynol-9 is the active ingredient in over-the-counter spermicides, and although it has shown activity against HIV in vitro and in animal models, we now know, many trials later, that it is not an appropriate candidate for a topical microbicide because of its nondiscriminating surfactant properties.

So, why are we here today?

One of the main reasons why we are here today to discuss topical microbicide drug development is because we are receiving Phase 3 clinical trials from sponsors, and we want to be able to provide them with the best possible advice.

So we convened this meeting of experts to help us help the sponsors.

To have a productive discussion today, I would like to lay out a background of topical microbicides, beginning with the definition that we developed.

[SLIDE]

It is a drug or biologic product that is being developed for the reduction of transmission of HIV or other sexually-transmitted infections, and given its name, it is applied topically.

It comes in various formulations that can be used with or without a device, such as a sponge or applicator. Formulations range from cremes, gels, et cetera.

It may or may not have spermicidal activity.

It is applied prior to intercourse, intravaginally or to the rectum.

And for the purposes of today's meeting, we will be focusing on female-controlled, intravaginally-applied topical microbicides for HIV reduction.

[SLIDE]

What are some of the ideal characteristics

of a topical microbicide?

It should be non-irritating in that the normal vaginal defenses should be maintained as well as the epithelium and the natural flora that reside there.

It should be discreet in that it should be odorless, tasteless, and colorless.

It should be stable in most environments, because the hope is that it will be used worldwide to reduce transmission of HIV.

And, although the FDA does not get directly involved in pricing, it should be affordable to reach as many people as possible.

These are the ideal characteristics, but we also need a topical microbicide to be safe and effective. Although this is the standard for the U.S. FDA, it should also be the standard for developing countries as well as developed countries.

[SLIDE]

There are a number of classes of drugs in the pipeline that are being considered as topical microbicides. Broadly, there are surfactants, buffering agents, chemical barriers, entry inhibitors, and nucleoside and non-nucleoside

reverse transcriptase inhibitors.

Why is there such an urgency today to discuss this pertinent topic?

I can think of three main reasons why we should be discussing topical microbicides in a public forum at this point in time. One, there is no vaccine on the market for HIV prevention. The second reason why I think there is an urgency is that it is difficult for women to deal with the condom issue. And lastly, HIV/AIDS remains an infectious disease of epidemic proportions.

[SLIDE]

This is seen on this slide, which is taken from the UNAIDS WHO database and shows adults and children estimated to be living with HIV/AIDS as of December 2002. And what is remarkable here is that of the 42 million, almost 30 million are living in Sub-Saharan Africa. But Eastern Europe, the Pacific, Latin America, and North America are also significantly infected and affected.

[SLIDE]

We take this data from UNAIDS and WHO and look at it in a more tabular format. What is remarkable in this slide, in addition to the numbers of people living with AIDS and HIV--and

that is the main mode of transmission of HIV. So throughout the world, particularly in Sub-Saharan Africa, North Africa and the Middle East, North America, et cetera, heterosexual transmission remains one of the main modes of transmitting HIV/AIDS.

[SLIDE]

In this slide, we see highlighted the number of women infected by this infectious disease. This is a global summary as of the end of 2002, and looking at the three categories--number of people living with AIDS; people newly infected with HIV in 2002; and AIDS deaths in 2002--you can see that women, highlighted in yellow, make up almost 50 percent of this epidemic.

So it is hoped that with rational drug development, we will be able to develop a marketed microbicide that will help to decrease the numbers of new infections.

[SLIDE]

The United States has not been spared. This is a CDC estimate of AIDS incidence in women and adolescent girls as of 2001. What you can see on this pie chart is that heterosexual transmission accounts for 66 percent, made up of the two

categories, sex with injection drug user, 16 percent, and sex with men of other or unspecified risk, 50 percent.

[SLIDE]

So what will we be discussing at today's meeting to help sponsors develop Phase 3 clinical trials that will be successful?

We will be discussing trial design issues primarily, and our speakers today will be presenting information on different types of trial design, namely, Phase 2/3 run-in versus traditional types of trial designs. We will be discussing the virtues of a single trial versus two adequate and well-controlled trials. We will also be asking the Committee to comment on control arms in three-arm and two-arm clinical trials and discuss the criteria of FDA of a "win" in a clinical trial.

In addition, we will be asking you for your opinion on trial duration, the goal of which is to capture not only efficacy endpoints but assess durability of treatment as well as long-term safety.

[SLIDE]

Today we have a number of outstanding speakers, some of whom have traveled great

distances to be here today, and we greatly appreciate that.

Our first speaker will be Dr. Salim Karim from South Africa. He will give the global perspective on the urgent need for an efficacious microbicide.

He will be followed by Dr. Lut Van Damme, who is the principal investigator in the COL-1492 clinical trial of nonoxynol-9 vaginal gel.

Then, Dr. Teresa Wu, a Medical Officer in the Division of Antiviral Drug Products, will be presenting a regulatory perspective on considerations for topical microbicide Phase 2 and 3 clinical trial designs.

This will be balanced by an investigator's perspective from Dr. Andrew Nunn from the UK.

Then, we will have a presentation on statistical considerations by Dr. Tom Fleming, and we will have the regulatory perspective by Dr. Rafia Bhore.

Thank you very much.

DR. GULICK: Thanks, Dr. Birnkrant.

So we'll jump right in and start with our speaker presentations.

Our first speaker is Dr. Salim Karim, from

the University of Natal in Durban, South Africa.

HIV and STIs in Women:

The Urgent Need for an Efficacious Microbicide

Dr. Salim S. Karim

DR. KARIM: Thank you very much.

I'd like to start by thanking the organizers for inviting me. What I hope to do in the next 15 minutes is to give you a very personal perspective, but I also want to share with you data that come from one of the potential trial sites for some of the microbicides that are going to be tested in Phase 2 and 3 trials soon.

So I am going to try to address the issue of capturing the main issues in the epidemic, particularly the epidemic as it affects Sub-Saharan Africa, and I want to make the case for an urgent need for a safe and efficacious microbicide.

Dr. Birnkrant has already touched on the issues of the global epidemic and the way in which women are particular infected, so I am going to skip over the first two slides. Just to make the point that within the entire global epidemic, the epidemic is particularly affecting Sub-Saharan Africa, where we have close to 30 million of the 42 million infected individuals.

[SLIDE]

Within that context, the country that is most affected is the one I come from--South Africa--where we have some 5 million infected individuals. So I want to share with you some of the data from this epidemic to show the way in which this epidemic is affecting women in particular.

[SLIDE]

Let me start by sharing some data from the national antenatal surveys. These are done by the Government of South Africa each year, and they plot out the way in which the epidemic has been steadily growing in South Africa.

So if we look at the period prior to 1990, we had almost no HIV infection in the general heterosexual population, and it picked up, as you can see, the first period of the epidemic, where there was a slow and steady increase. And that was followed in about 1994 with a period of very rapid rise in infection, and over the last few years, we are seeing some degree of evening off within this epidemic curve.

[SLIDE]

Now let me go to one particular site, and

this is a rural community in a part of the country just 3 hours north of the city of Durban. I want to share with you data that come from this particular community in Hlabisa and show you how the epidemic has grown in this particular community.

In 1992, the prevalence of HIV infection was 4.2 percent. A year later, it had grown to 7.9 percent, and 2 years later to 14 percent, to 27 percent--and you can see in the latest data we have from 2001, the prevalence of HIV infection in prenatal clinic attendees is 36.1 percent.

Data on incidence, which we have calculated through a mathematical model, show how incidence has also grown concomitantly, driving the increase in prevalence. The latter estimates of incidence have also been corroborated with estimates calculated through the D-2 [inaudible].

[SLIDE]

But this epidemic is not affecting both men and women equally. HIV in South Africa is a highly discriminating virus. It has a certain gender distribution and age discrimination, and let me try to capture this.

Although these data come from an early

point in the epidemic, they are still applicable today. So if you follow with me the yellow line, you can see how the prevalence arises in men and achieves a peak in the age group 25 to 29.

If you compare that to the situation in women, we have a situation where the prevalence starts rising in the young teenagers. So we even have close to the peak of the HIV prevalence in the age group 15 to 19.

So what we have is a situation where young women are particularly affected by the HIV epidemic in this community in Hlabisa.

[SLIDE]

Let me for a moment look at the cohort effect, and what I want to do is present data to you that is AIDS-specific from Hlabisa. So let me start by just asking you to focus on 1992.

If one looks at the data for 1992, the prevalence in 20 to 24-year-old women was 6.9 percent. And if you look as you go to the older age groups, the prevalence steadily declines.

If one looks at how the epidemic has grown over the period 1992 to 2001--that is the 10-year period involved--we see that the prevalence has grown from 6.9 percent to 21.1 percent to 39.3

percent to 50.8 percent. This is nothing short of a catastrophe. And what we are seeing in these young women is an epidemic that is growing explosively in these three intervals.

Let me now ask you to cast your eye to the diagonals. What we have is because we have three differences in these periods of measurement, the individuals in this particular cell, a large number of them, will be in this cell some 3 years later, and so on.

So if we follow this particular birth cohort, if we think about it as the "class of '92," these women experienced this epidemic growing from 6.9 percent, some 3 years later to 18.8 percent, to 23.4 percent to 36.4 percent.

So what we are seeing in this setting is a rapidly growing and explosive epidemic.

[SLIDE]

And if we look at the incidence rates that we have been able to measure--and we have been able to measure them in 1998 and in 2001--what we see is not only that we have a growing prevalence rate, but we are seeing that the incidence rates continue to remain high. So that from the period 1998 to 2001, we continue to see high incidence rates.

[SLIDE]

One of the studies that we have where we have long-term follow-up data--not data where women have only been followed up for a year--comes from the COL-1492 trial. I have just collapsed the data for both arms of the trial in this particular slide. And if you look in this particular population--and these are sex workers who work at the truckstops in the midlands or the middle region of the province of Kozulu Natal [phonetic]--you see that in the period 1996 to 1997, the incidence rate was 16.8 percent per annum. In 1998, a year later, it had gone up to 18.2 percent, and in 1999 had gone all the way up to 20 percent per annum.

Some people ask me, how can you even get an incidence rate of 20 percent. Well, these are data that come from the follow-up of these women, and what we are seeing is the way in which this epidemic continues to rise, not only being driven by high incidence rates but even growing incidence rates at this level.

[SLIDE]

In the same group of sex workers, let's for a moment look at the incidence rates of STIs. For trichomonas vaginalis, the baseline prevalence

on enrollment in the study was 36.1 percent, and by the end of each year on average over the 3 years of follow-up, a woman was being infected with trichomonas more than once. So we have an incidence rate of 114 percent per annum. And you can see again here the HIV incidence rate of 18.2 percent.

So what we are seeing in this particular population is incredibly high incidence rates of STIs and HIV.

[SLIDE]

If we go back for a moment to the rural community of Hlabisa and try to understand in a little bit more detail one of the key issues regarding the way in which STIs are distributed within this community, let me for a moment present data that come from a collection of various studies that we have undertaken.

In this particular community, we estimated that there are about 56,000 women age 15 to 49 years. So in the reproductive age, we expect that there are about 56,000 women. Right now as I am speaking to you, we estimate that about 25 percent of these women have at least one STI. And I am referring here to the five major STIs in this

particular community.

Of these women, of these one out of four women who have an STI, we estimate that only half of them have some kind of symptom. The symptoms would be pain or burning on miturat ion [phonetic]. And of these symptomatic individuals, only 2 percent of these women will recognize these symptoms and seek treatment. And of those who seek treatment only 65 percent, or 2 out of 3, will be adequately treated. The other one-third of the patients will either go for traditional healing or would be treated incorrectly in the private or public sector.

So w hat we have is a huge burden of sexually-transmitted infections in a community like this.

[SLIDE]

You might ask have we not been able to make any dent on this epidemic. What of all the prevention programs? Let me present some data that show that within South Africa, we have had a growing use of condoms in both males and females.

Let me start by presenting some data on the male condom. In 1994, before the Mandela Government took over, the Government of South

Africa distributed approximately 8 million pieces of condoms each year. In the first year of our democracy, that went up to 97 million. And you can see in the year 2000 that we distributed 250 million, and that went up to 267 million in 2001. I don't have accurate data for 2002, but these are national government estimates, and they estimate that they will be distributing some 358 million pieces of condoms.

[SLIDE]

If one looks at the situation for female condoms, one can see here again--and female condoms are made available publicly through the government clinics--we distributed 600,000 pieces in 2000, and that has grown to about 1.3 million pieces, and they estimate that that will continue to grow to about 2.5 million pieces last year.

So in the presence of this kind of epidemic, what we are seeing is an increasing use of condoms, both male and female.

[SLIDE]

Just to give you some idea that these condoms are not merely being taken from clinics and thrown in the bin or being used as balloons at children's parties, we did a study where we

followed up 384 condom recipients, and these were at six clinics throughout South Africa. These 384 individuals had received 5,528 condoms. We then revisited these individuals at 5 weeks, and we undertook an assessment to look at how many of the condoms had been used, how had they been used, and what remained.

What we found was that 43.7 percent of these condoms had been used, that 21 percent had been given away, 8.5 percent had been lost or discarded, and 26 percent were still available for use. That enabled us to get some estimate that our wastage in condoms at 5 weeks remains still below 10 percent. So if we extrapolate the use of condoms in South Africa based on this, we were talking about 87 million condoms.

So there is o question that condom use is already increasing and we have high levels of condom use in certain parts of South Africa.

[SLIDE]

What I would like to show is that what we have in this particular epidemic as it affects a community like Hlabisa is that the condom is of little use to the particular women who are at highest risk in this community. Why am I saying

that?

If one looks at the women in Hlabisa, many of the young women have partners who are migrant workers. A woman of let's say 20 years will have a partner of around 30 to 35 years, and that man will be a migrant worker either in the mines or in the city of Durban. When he comes home, he is coming home to his girlfriend or to his wife. She is looking to have his children. There is no possibility that the condom would even feature in that kind of equation. But when he is in the city or when is at the mine, he has a town wife or he is using visiting sex workers, so we have a situation where the very person that she wants to have unprotected sex with is the person who is infecting her.

We see this over and over again in this particular setting. When I was working in Hlabisa Hospital, I remember a young woman coming to me with her newborn baby--the baby was about 8 months or so by then--and the child had severe diarrhea and really looked emaciated. We did an HIV test, and the child came back positive.

I was involved in counseling this young women and explaining to her that the child does

have HIV and that she should also be tested. So when we tested her and the result came back, she was also HIV-positive. And I was trying to explain to her how one gets HIV, and she explained to me that she doesn't sleep around; she has been faithful to her husband.

So it is not a question that she has any of these risk factors, and it is very hard to explain to her that in fact it is the very person that she is having sex with--her husband--who is the one who infected her.

We are looking at a setting where young women are really powerless to use these condoms, so the condoms that are being used are not being used in those particular age groups of young women where they could have maximum benefit. What we need in this particular age group are methods that women can use and control.

So, what happens when prevention fails, as we have in our setting?

[SLIDE]

Let me show you again from this community in Hlabisa the prevalence of tuberculosis--or, actually, it is the incidence, the number of cases of tuberculosis in this particular community.

By the year 1990-1991, we had TB very much under control in this community. We have a superb DOT program in Hlabisa District. And at that point, Hlabisa Hospital had one TB ward for women and two TB wards for men. And if we look at the way in which the numbers with TB have increased, we can see that it has moved up from about 400 in 1990-1991 to a situation where we have a four- to five-fold increase, with a peak in 2001 of over 2,500 case of TV. We have had one whole section of the hospital that has been converted to TB wards, and we now have four female TB wards and two male TB wards.

It just shows you again how this epidemic is growing particular in women and particularly in young women.

[SLIDE]

If one looks at our teaching hospitals, this is a study done in 1998 in medical inpatients. So these are patients admitted to the medical ward. Fifty-four percent of the patients were HIV-positive, and 84 percent of them met the criteria to be regarded as AIDS cases.

We have more women being admitted than men, and that 56 percent of the HIV co-infected had

tuberculosis. What is striking is if you look at the case fatality rates, where we have 22 percent of HIV-positive patients admitted to medical wards leave the hospital in a hearse compared to 9 percent for HIV-negative patients.

[SLIDE]

Let me end by sharing with you some data on mortality since these data tell the real crux of the story of the epidemic in South Africa.

I need to explain briefly how to read the data on this particular graph. This point, the reference point of 100 or 1, is the average mortality rate in men during the period 1985 to 1990. So we have used that as a reference point.

If one looks at the period 1996 to 1998, we see that the mortality rate in young men around 25 to 29 and 30 to 34 is starting to rise, although much of this is simply noise.

If one looks at the mortality rate in 1999 and 2000, one can see a clear upward rise. So what we have is an increase in the mortality rate in men about one-and-one-half-fold in the age group 30 to 34 years.

So what we are seeing is about half as many more men dying during this particular period.

[SLIDE]

Now let's look at the situation for women. What we see here--again, remember this is the baseline of 100--is in the year 1999 to 2000, what we are seeing is a three-and-one-half-fold increase in the mortality rate in young women. And this particular peak occurs in women 25 to 25 to 30 years of age.

So what we are seeing is an epidemic that is growing particularly rapidly where incidence rates continue to remain high against a setting of a high prevalence of other STIs, and we are now starting to see morbidity and mortality taking its toll, particularly in young women.

[SLIDE]

In conclusion, the epidemic in Sub-Saharan Africa with South Africa gives us one picture. We are experiencing five parallel effects. First is the continuing large numbers of new infections, and with the high prevalence of HIV in young women, this is the group that is also most reproductively active, so we have a growing number of both orphans and infected young children.

We have rapidly rising mobility, and we can see its impact on our health services. And

with that is the rapid rise in the number of deaths and an increase in the number of orphans.

What it highlights to us is that although we have been making this plea that we must have treatment, we have got to avert this crisis of the growing mortality. Treatment on its own is not going to be good enough. We have to be looking at prevention and treatment.

And lastly just to say that women are more severely affected by this epidemic and that condom uptake and use continues to increase, but there is still within that context a clear need for a woman-controlled method and that within this epidemic which is affecting young women, microbicides have the real potential to influence the course of this epidemic.

Thank you.

[Applause.]

DR. GULICK: Our next speaker is Dr. Lut Van Damme, who is from the Contraceptive Research and Development Program in Arlington, Virginia, and was the PI of the COL-1492 study.

Lessons Learned from COL-1492,

A Nonoxynol-9 Vaginal Gel Trial

Lut Van Damme, M.D., M.Sc.

DR. VAN DAMME: Good morning. I will present the lessons learned from the COL-1492 trial for the design of future microbicide Phase 3 trials.

[SLIDE]

UNAIDS was the main sponsor of this study. COL-1492 is marketed in the United States as Advantage S and is a vaginal gel containing 52.5 mg of nonoxynol-9 in a bio-adhesive carrier.

The placebo that we used in all the trials is a vaginal moisturizer also on the market under the name of Replens. This is very similar to COL-1492, although a little bit more viscous and a slightly lower pH.

The study was two-arm, randomized, blinded, placebo-controlled study. And I want to draw your attention to the fact that we did a Phase 2/3 trial. Women who were enrolled in the Phase 2 in which we performed colposcopy could stay in follow-up while we awaited on our DSMB decision to continue with the Phase 3, and those women were all contributing to the main analysis of the study.

[SLIDE]

Before starting on a Phase 3 study, we decided to test the product for its safety. First,

we tested it on low-risk women who used the product once a day for 14 days. In this safety study, we did include a no-treatment arm, and there was no difference with regard to the incidence of lesions with an epithelial breach in the three arms, and this incidence was also very low.

Based on these results, we started our Phase 2/3 trial and started enrolling women in the Phase 2 part of the study. This is a study population at high risk of infection, using the product as much as they wanted because there was no set maximum, and also here, the incidence of lesions with an epithelial breach was low, and it did not differ between the two treatment arms.

[SLIDE]

Back to our Phase 3 trial and the main results. The main analysis was done under intent-to-treat principle. There were a total of 104 seroconversions, 59 of which occurred in the COL-1492 arm, giving a 15 percent incidence of HIV, compared to 10 percent in the placebo, and this difference was significant.

[SLIDE]

These are the issues I would like to briefly discuss with you during my talk. Some of

them are a direct consequence of the COL-1492 trial results as the placebo and the no-treatment arm. Others are more generally linked to Phase 3 microbicide trials.

[SLIDE]

When the COL-1492 results became available, the placebo that we used was questioned as to its ability of protecting women from HIV infection. We cannot completely answer this question since we did not design a trial for measuring the placebo effect. However, our explanatory analyses do point toward a toxicity of COL-1492 use.

But it is indeed correct that an ideal placebo should have no impact at all on HIV infection, be it by lowering the vaginal pH or coating the vaginal walls or having an impact on the flora. And it should also be indistinguishable from the experimental product to allow blinding of the trial. However, if we cannot completely blind, it's better to partially mask than to have no masking at all.

Based on discussions with colleagues from CONRAD and Vita H. Petty [phonetic] and Tom Lynch from Reprotect [phonetic] have now developed the

ideal placebo which is a HEC-based gel and which should have no effect at all on HIV.

Currently, this product is being tested for safety in the clinical facilities of CONRAD in Norfolk.

[SLIDE]

Another often-made argument is that if we had included a no-treatment arm in our trial, our data interpretation would have been much more simple. That is correct on first glance, but when you look more closely at the issue, it definitely is not.

Suppose that we have a no-treatment arm which has an equal HIV incidence with the placebo arm. What does this mean? Is it indeed that we have found the ideal placebo which has no effect at all on HIV, or are we looking at the differential behavior change between the two groups?

This differential behavior change may go in two directions. We could imagine that the women who are assigned to no treatment are adhering much more to the safe sex counseling guidelines than the women in the treatment arms, and thus they increase their condom use, and thus, the equal HIV incidence that we see is in fact women in the no-gel arm

using more condoms and thus masking the protective placebo effect.

However, we cannot predict if this change will go in the direction I just pointed out. It could also go in the opposite direction, and that is that women who are assigned to a gel are much more motivated to keep to the trial procedures, and trial procedures do include safe sex counseling, and thus women increase their condom use more so than women in the no-gel arm.

So we cannot exclude that with a no-gel arm treatment there will be a differential behavior change. That's one thing. Two, we cannot predict in which way this behavior change will go. And three, if it happens, we cannot predict the magnitude.

The randomization takes care of baseline characteristics but does not correct for prospective bias happening because of differential behavior change after randomization. This prospective bias is a very big threat to our data interpretation.

There also would be an impact on the loss-to-follow-up. It may well be that women who are not assigned to the gel arm are not so

motivated to stay in the trial for the period of length that we are testing and come to the clinic on a regular basis, and thus, you are introducing a differential loss-to-follow-up among the gel arms compared to the no-gel arm--again, making our data interpretation much more difficult.

Some investigators feel that there may be an impact on recruitment potential, since for many people, if you are part of a study, it means you will have to use a study product, so when they hear they can be assigned a no-gel arm, this may make them lose interest in trial participation.

And we should not forget that there may be a tendency that women who are assigned to a gel arm would be inclined to share their product with women, often their friends, who are assigned to a no-gel arm.

Besides those factors, there is also the impact on the real conduct of the trial. If we have to implement a three-arm study with two control arms, are sample sizes per definition increased? I am sorry--I don't know why that sign is there; it should be a double arrow. This makes the sample size bigger, much more difficult to recruit, a much more expensive trial, logistics

more difficult to handle, and it will take much longer to finalize a trial.

[SLIDE]

We should also not forget that any experimental product which has less effect than a placebo, even if this has a low effect, will not have a tremendous effect on HIV prevention on a worldwide scale. Some of those products are already there, and this might just reduce the looming HIV epidemic.

Another challenging thing is what about the behavioral data collection. One could argue that since we do all the Phase 3 main analyses under the Intent to Treat principle, we do not really need to collect those data since we do not use them for doing our main analysis.

However, they may prove very useful if we want to better understand trial results and do exploratory analysis as we find out with the COL-1492 trial. Only these data allow us to better understand what was happening in the trial.

We assume that the [inaudible] would be equal in the two arms since both were assigned to a gel, and the trial was blinded.

How best to collect those data is not

known today. We started with a simple coital log chart which we then changed to a more detailed coital log chart. This had been piloted before, with success. However, in the big trial, it was not all that good. Also, the counting of all those different sexual acts, with or without gel and with or without condom, was a huge burden to the staff. So we changed the procedure and asked them direct questions on their most recent sexual acts.

Some say--and this may indeed be true--that women are inclined to report behavior that they think the researchers would like to hear and thus over-report safe sex behavior. This may be correct. Therefore, some researchers [inaudible] the older, computer-assisted self-interview. This would decrease the desirable behavior tendency, and it would also decrease the intensity that goes together when you talk directly with women on sexual behavior issues which are still sensitive and sometimes a tabu issue.

[SLIDE]

And then, what to do with the safety trials. In our safety trials with COL-1492, we did not detect any toxicity that worried us despite that in the second safety trial among high-risk

women, they could use the product as much as they wanted. In the Phase 3 data, however, we saw a strong association between having a lesion with an epithelial breach and the HIV seroconversion. This risk was twice the risk among women who had never had such a lesion.

Should we disregard all the safety trials because probably what we see is that the sample size in a safety trial is too small to detect any significant effect? I would say no. One, if there were a major toxicity, we would detect it. Two, the COL-1492 trials show indeed what we thought--a lesion with a breach increases a woman's risk of HIV infection. We can detect those lesions.

The problem, however, today is that we do not know the threshold of an acceptable incidence of lesions, and this today can only be assessed in a Phase 3 trial where the sample size is big enough to detect any significant effect because a product which has limited toxicity may prove to be protective against HIV.

A third reason for doing the safety trials is to detect any systemic toxicity that the product may have.

Currently, investigators are looking at

different ways of addressing and assessing the safety of a product beyond colposcopy. Today, it would probably be best if you could put all the data together of cytokines, neutral fields [phonetic], and so on, but today again, you cannot link the results of this extra testing to the risk of a woman becoming HIV-infected.

[SLIDE]

Enrolling sex workers has also often been criticized by saying that a sex worker is not representative of women in the general population, and thus we cannot generalize study results to a general population setting.

But what is a general population? If we go to women in stable relationships who have an average of two acts per week, can we say she is representative for a young girl in her early sexual debut and who goes out on the weekend and has multiple acts?

We should also not forget that by generalizing results from a trial, we always have to be careful, because once a product is on the market, it will be used in a different way than when it was in the trial, since the pressure of being in a trial and regular contacts with study

staff will be gone.

We should also keep in mind that women who enroll in a trial do show an interest in that product, or else they would not volunteer to participate in the trial. Today we do not know, since there is no effective microbicide, if that interest in using a product is really generalizable to the general population.

Another argument against sex workers has been that we may be withholding a potential beneficial product because those women are using the product multiple times a day, thus triggering its toxicity, and this may be correct.

However, we should not forget that most women will use a product at one time or another multiple times a day. The COL-1492 results show clearly that it is very important to know what happens if women are using this product multiple times in a short period of time--and this can happen not only in sex worker populations but in every general population, especially among the young women, who are very vulnerable to HIV.

[SLIDE]

And then, the p-value. This is not directly linked to the COL-1492 trial results.

However, it is very high on the current agenda since the FDA requires a p-value of .001.

On this side, you can see the impact that the p-value has on the sample size, and thus, you may see that the p-value of .001 doubles the required sample size compared to a p-value of .05.

We do indeed not want to erroneously decide that a product is effective when it is not. However, the .001 value is, I think, too high a threshold. There is an urgent need to find a method that women can use to protect themselves, so it is very important that we can do the trials in a timely fashion. By using a .05 p-value, we do not do any harm to the quality of the science.

[SLIDE]

So, based on the COL-1492 experience, based on discussions with colleagues in the field and choosing to do high-quality science which can be done in a timely fashion due to the urgent need for a female-controlled method, CONRAD has assigned on its Phase 3 design as shown on this slide.

It will be a 2-arm trial, randomized placebo control with an 80 percent power and a two-sided .05 significance level. We assume a 50 percent effectiveness of the product, a one-year

retention of 80 percent, and we will ask women to stay one year in the trial.

The one-year retention rate is based on real life data, and the one-year follow-up is based on what we think is feasible to implement in the field.

[SLIDE]

I will now briefly discuss some ethical issues and can go quickly over this slide, because for once, there is consensus in the field.

We are all aware that obtaining informed consent is not a "once and for all" event and that we have to repeat our information to trial participants, since women tend to forget what has been told them.

At the end of the session, when we obtain a woman's consent, we ask her a set of questions on the basic principles of the trial--for instance, randomization and blinding. We repeat this set of questions throughout the trial, and whenever she does not remember certain aspects, we repeat the information.

No matter how long and how often we repeat some information, there are beliefs which are very difficult to change--for instance, "What every

doctor tells me is good for me."

[SLIDE]

And then, last but not least, there is the issue of providing treatment. There is no discussion at all on providing STI treatment for all women in the trial at screening and during trial participation. However, providing antiretrovirus [phonetic] is a different issue. Some say that we should continue to refer to the local standard of care, whatever that is; others feel that we should make ART available to women who seroconvert while they are participating in the trial.

CONRAD has not made a final decision yet, and we will discuss it with AID, one of our main sponsors, and with investigators in the field. In-house discussions pointed toward that we would try to make a fund available for investigators so that they can use this fund whenever women who seroconvert during the trial need to go on ART. We probably would set a pre-set limit on the period of time that this ART would be sponsored, and of course, we would also sponsor and pay for the prevention of opportunistic infections.

These are the things I wanted to discuss.

Thank you.

[Applause.]

DR. GULICK: Thanks, Dr. Van Damme.

The next speaker is Dr. Teresa Wu, from the agency.

Considerations for Topical Microbicide Phase 2 and 3 Trial Designs: A Regulatory Perspective

Teresa C. Wu, M.D., Ph.D.

DR. WU: I would like to firstly thank the two previous speakers for nicely explaining why there is a real need, a real global and urgent need, for developing a safe and efficacious microbicide.

My name is Teresa Wu, and my charge this morning is to present considerations for topical microbicide Phase 2 and 3 trial design from a regulatory perspective.

[SLIDE]

What I plan to accomplish in my presentation is to firstly summarize for you the types of microbicide in the pipeline or in clinical development. Then, I will describe the regulatory tools in existence provided by the U.S. FDA that may facilitate and expedite review of a microbicide application.

I will then describe the Divisions current recommendation on how to develop a microbicide from non-clinical to Phase 1, 2, and 3 trials.

For Phase 2 and 3 trials, which are the focus of today's meeting, my colleague, Dr. Bhore and I have selected the following topics--design, populations, endpoints, controls, effect size. And Dr. Bhore will later discuss the statistical issues such as study duration, single trial, sample size.

[SLIDE]

To reiterate what Dr. Birnkrant showed in her introduction, the types of microbicides are grouped by mode of action. One group is detergent-like chemicals which are capable of destroying pathogens nonspecifically. The second group of chemicals provide natural acidity of a normal vaginal environment and therefore maintain vaginal defenses against infection. The third group is based on mechanisms targeting attachment of pathogens to target cells. The fourth group is based on specific mechanisms targeting HIV at either entry or replication steps.

There are still potential microbicides with mechanisms of action unknown, such as herbal agents.

[SLIDE]

In a survey conducted by Alliance for Microbicide Development, approximately 60 products are currently in the pipeline. About 20 of these are either planned for or are entering human testing. There have been 9 applications filed with FDA, and four of them are presently planned for Phase 2 or 3 human trials.

[SLIDE]

What are the regulatory tools? Given the urgent need for an efficacious and safe microbicide, our present goal is to guide promising candidate microbicides to quickly move into Phase 2/3 trials.

[SLIDE]

Under the regulation, topical microbicides are eligible for the so-called Fast-Track Drug Development Program because they are intended to prevent a serious or life-threatening condition, and development of a microbicide will have the potential to address unmet medical needs.

[SLIDE]

Sponsors can apply for a Fast-Track application any time after the IND submission. Under the Fast-Track Drug Development Program,

there are several regulatory tools that can expedite the review process. Before an IND submission, sponsors are highly recommended to have early contact with FDA through pre-IND consultation. After IND submission, sponsors are entitled to request regular meetings with the Division, such as Phase 1, end of Phase 1, end of Phase 2, pre-NDA meetings, to discuss and achieve agreement on critical issues.

When the NDA is submitted, FDA may consider to review portions of a marketing application before the complete NDA is submitted. This is the so-called rolling submission.

The review clock will not begin until the applicant informs the agency that a complete NDA has been submitted. A priority review will be granted after FDA determines the fileability of the application. The review time for a priority review product is 6 months as compared to a standard review time of 10 months.

[SLIDE]

There are two recently-published guidelines which summarize a consensus developed by participants from academic, pharmaceutical, and regulatory organizations including FDA at two

separate workshops. One was sponsored and then issued by the International Working Group on Microbicides, or IWGM, in 2001, and the other was sponsored by the Rockfeller Foundation in the year 2002. Both publications are complementary to each other.

Despite these two published guidelines, there are still issues unresolved on the development of topical microbicides. This is why we are having today's meeting.

As a regulatory agency, our recommendations on how to develop topical microbicides are in large part consistent with these two published guidelines.

In the remaining slides, I am going to summarize FDA's current recommendations.

[SLIDE]

Before a microbicide product can be administered to humans, vigorous nonclinical studies are required. These include in vitro antiviral activity, cytotoxicity, mode of action, resistance and cross-resistance activities, impact on pathogens causing sexually-transmitted infections.

Today, the animal models used for

demonstrating microbicide antiviral activity have had limited utility in helping to decide which compounds should go forward into clinical trials.

Nonclinical studies to assess local and systemic, general and reproductive toxicity and pH should be conducted.

Microbicide products should meet the standard chemistry and manufacturing control expectations in terms of their proper identification, stability, purity, and strength.

[SLIDE]

Phase 1 trials of topical microbicide typically are conducted in about 200 subjects. The primary objectives are to assess local and systemic safety; selection of dose, formulation and initial product acceptability; usually, the microbicide is given once or twice daily for 7 to 14 days; in HIV-negative women, first including women to be abstinent during the study, followed by enrolling sexually active women.

[SLIDE]

Conventional Phase 2 trials commonly enroll several hundred women, are designed to collect local and systemic safety data and acceptability than a larger group of women, and

also to evaluate microbicide activity as proof of concept study.

[SLIDE]

However, in microbicide trials, since there are no known clinical correlates available, proof of concept for HIV prevention can only be measured in studies with very large numbers of participants.

[SLIDE]

This is because of two factors. Number one, low HIV incidence rate in high HIV-prevalent regions--for example, one study showed that in India, Zaire, and Rwanda, among commercial sex workers receiving condom counseling, the instances were three to five per 100 person-years. In another study in Cameroon, where the HIV prevalence rate was very high, the rate was reported to be seven per 100 person-years.

These numbers are lower than those presented by Dr. Karim due to the considerable variation in HIV prevalence between different regions in Africa. A 5-per-100 person-year rate has been commonly used by sponsors for calculating trial sample size.

[SLIDE]

The second reason is that HIV is a fatal and incurable disease. It is ethically necessary to promote condom use and provide safe sex counseling to all participants. Here, I am referring to male condom use only. Therefore, high levels of condom use will likely further reduce HIV incidence rates.

[SLIDE]

Both the IWGM and Rockefeller Foundation initiatives have suggested a hybrid design for combining Phase 2 into Phase 3 design. A subgroup of participants will enroll in the Phase 2 component and undergo monthly visit evaluations, more intense safety evaluations, including expanded local safety testing. Moreover, a subset of this group will undergo colposcopy examination for vaginal epithelial abnormality.

Phase 2 participants will continue follow-up and the first 3 months Phase 2 data will be reviewed by DSMB.

Concurrent with the follow-up portion of the Phase 2 component and the time required to complete the Phase 2 data review, accrual of Phase 3 participants will begin, and the earlier Phase 2 participants will uninterruptedly be phased into

Phase 3. Examination will be quarterly. HIV seroconversion will be tested quarterly as well.

This design allows for a more intense safety evaluation in the Phase 2 component before a large number of women exposed to the candidate microbicide. I should point out that the Phase 2 component is not designed to address the proof of concept.

[SLIDE]

Who should be studied?

It is generally accepted that the ultimate goal is to make a microbicide product available to women at risk at all levels. The study population will be women in regions with high HIV prevalence; they are HIV-negative, sexually active, and non-pregnant and at risk for sexually-transmitted infections.

Such high HIV prevalence rates occur predominantly in developing countries such as Sub-Saharan African countries.

Some sponsors have proposed a study exclusively in commercial sex workers because of higher instance of HIV infection. Given their potentially high rate of product application, which might enhance the rate of vaginal irritation,

results obtained from commercial sex workers may not be fully representative of a product's safety and efficacy among other groups of women.

Therefore, we generally recommend that women at varying degrees of risk for STI infections be included.

One important group which should be particularly mentioned is adolescents. Adolescents represent a very high-risk population for acquisition and spread of STIs. A safe product in adults is not necessarily safe in adolescents given adolescents' maturing anatomy and physiology and risk behavior.

However, due to legal and cultural constraints, including adolescents in clinical trials may be logistically difficult.

[SLIDE]

Because most topical microbicide trials will be conducted in developing countries, and sponsors have expressed an interest to seek marketing approval for their product in the U.S., studies conducted in foreign countries will likely become the major if not the only basis for most microbicide applications.

When foreign data as the sole basis for

marketing approval is sought, one of the requirements is that "data are applicable to the U.S. population and U.S. medical practice."

[SLIDE]

Since most microbicide trials will be conducted in developing countries, we think the easiest way to meet this requirement is to have a U.S. bridging population as part of the package for a candidate microbicide application.

U.S. population is primarily for determining the safety profile and acceptability under the condition that the duration of microbicide usage will be comparable to that of non-U.S. participants.

There are a number of options the sponsors could choose from by including a subset of U.S. participants in Phase 2 run-in Phase 3 trial, or by using data from a separate contraceptive trial if the microbicide is also a spermicide, or by using data from STI prevention trials other than HIV, such as chlamydia prevention in U.S. women.

[SLIDE]

The primary goal is to measure the rate of HIV acquisition and safety of the product, depending on the adequacy of the diagnostic

facility available at the study site and the prevalence rate at the site. The study should include but not be limited to STIs such as chlamydia, gonorrhea, syphilis, trichomoniasis, and reproductive tract infections such as BV, vulvovaginal candidiasis as a secondary endpoint.

To include STIs as secondary endpoint is based on the fact that STIs have been considered cofactors in HIV acquisition. In particular, ulcerative STIs have been shown to promote HIV acquisition and transmission.

The potential to increase susceptibility to one or more STIs should be assessed.

[SLIDE]

The selection of controls is a complicated issue for the topical microbicide. As I mentioned earlier, a microbicide trial, all participants should receive condom promotion counseling. We have recommended some sponsors to consider using two parallel controls--a placebo and a no-treatment arm. We prefer the term "no-treatment arm" over "condom-only arm" because in developing countries, condom use rate are very low despite condom counseling.

[SLIDE]

Placebo is the logical comparator at a time when there is no approved microbicide. Placebo remains the gold standard for providing blinding, maximizing unbiased estimate of efficacy and safety of the candidate microbicide.

[SLIDE]

In the case of microbicides, some components of the vehicle of the candidate microbicide, for instance, carbomer, have shown anti-HIV and anti-bacteria activity. Thus, more and more sponsors have turned to using a totally unrelated gelling compound as a placebo for the microbicide trial--the so-called "universal placebo." This term has gained popularity recently.

Because this universal placebo is not a vehicle, we have required sponsor to conduct limited nonclinical and Phase 1 studies prior to being used in Phase 2/3 trials. The universal placebo has been shown to have no in vitro activity against HIV and bacteria. However, some uncertainties still remain.

[SLIDE]

What are the uncertainties? The universal placebo gel itself is a physical barrier while

intravaginally applied. Thus, placebo may have an unknown level of efficacy. Equally unknown, a placebo may contribute to some level of local toxicity. Even if the placebo shows no vaginal toxicity in a small number of participants in Phase 1 studies, the safety profile in a large number of women still has to be established in a Phase 2/3 trial.

[SLIDE]

Thus, the advantages of having two parallel control groups are: blinding; validate the interpretation of efficacy and safety data obtained from the candidate microbicide arm; since the placebo may have some level of efficacy and/or toxicity, the inclusion of a no-treatment arm is to validate interpretation of the efficacy and safety data obtained from the placebo arm.

However, we are mindful of the disadvantages associated with the inclusion of a no-treatment arm. The no-treatment arm cannot be blinded, and as a result, participants may drop out of the study, resulting in differential dropout rates. Participants' risk behavior may change, either more or less motivated to use condoms. This would likely create a bias between groups.

Another potential effect could be gel-sharing, which will be very difficult to document. And regarding the control arms, my colleague Dr. Bhore will discuss further this issue in her presentation.

[SLIDE]

In a setting where condoms would be used consistently and correctly, condom alone can offer 85 percent protection against HIV transmission. However, low rate and incorrect condom use have been the norms in most developing countries. The microbicide community has generally accepted that even if the first product approved is shown to be only modestly protective, that is, relative to the consistent and correct use of condoms, one can still expect a significant public health impact on the reduction of HIV transmission.

Measuring the level of efficacy of microbicide in the present design is to measure incremental benefit offered over imperfect or actual use of condom use alone. The range of effect size expected for the first generation of microbicides in conjunction with imperfect or actual use of condoms is between 30 to 50 percent, as most experts in the field have agreed.

We acknowledge that this range is arbitrary; nevertheless, it was based on clinical judgment.

[SLIDE]

In summary, we recommend a Phase 2 run-in Phase 3 trial design; population enrolled should be generalizable, and data should be applicable to the U.S. population. Endpoints include HIV incidence, safety, STI incidences. We prefer two parallel controls, and effect size would be 30 to 50 percent in the context of condom promotion.

Thank you for your attention.

[Applause.]

DR. GULICK: Thanks, Dr. Wu.

Our next speaker is Andrew Nunn, from the Medical Research Council, London, UK.

MR. NUNN: Mr. Chairman, ladies and gentlemen, I would like to begin with a couple of introductory remarks, first of all to thank you very much for the invitation to speak today; secondly, to indicate that although what I'm saying is very much a personal perspective, it does reflect the views of those of us involved in the UK-based Microbicide Development Program, which is actually involved right now in the development of a

protocol for a large Phase 3 trial which we hope will begin next year.

[SLIDE]

I have been given 20 minutes, and in 20 minutes, it is likely that 100 women will have been infected with HIV. Most of those women are in the developing world, and most of the women will probably have had little opportunity to prevent that infection to protect themselves.

How many of those infections could have been prevented by the use of an effective vaginal microbicide?

[SLIDE]

We may differ in respect to a number of points that we are discussing here today, but I think we have a common goal that we will all agree on: We need a microbicide which is effective, safe, acceptable, and affordable.

[SLIDE]

There is a particular link between safety and efficacy which is almost unique in this situation, because local adverse events, some of which may actually be very minor in effect and may not even get reported, such as minor inflammation, may be closely linked to an increased risk of

infection and thus reduce the effectiveness of a product.

Clearly, the experience gained in the COL-1492 study which we heard about briefly earlier has alerted us to the need for a new level of vigilance concerning possible adverse effects from products under study.

[SLIDE]

What is the most urgent priority today? These are al priorities, but what is the most urgent--a highly effective product, a licensed product, or proof of efficacy?

[SLIDE]

I would suggest that in fact proof of efficacy is particularly important, because funders will only go on funding for so long, and if we reach a point in time at which they say, "We don't have much evidence of efficacy," they may lose interest and not be willing to continue funding.

Now, effectiveness of a microbicide will depend on the extent to which that microbicide is used. Use will depend on acceptability. And acceptability is likely to vary considerably between populations.

Heterogeneity of populations may provide

us with the best chance of demonstrating proof of efficacy. I shall return to this point a little bit later on.

[SLIDE]

In an ideal world, our trial design would be something like this. We would have several promising products to look at, and we would test them in one trial. The products would be outwardly indistinguishable from each other and from the placebo. The placebo would be completely ineffective, and behavior would be unaffected by participants taking part in a trial.

[SLIDE]

In reality, things are often different from that. Products may not be indistinguishable from each other--it may be necessary to have a placebo for each product. And sometimes one has to have dummy placebos in certain contexts, two placebos to each individual--but not in this particular context.

Placebos may have some protective effect, as has already been alluded to, and behavior will change. In fact, I would suggest that in a trial, behavior almost always does change, because of course, it's not a very real situation.

So, as a consequence of points 2 and 3, any such trial would not mirror what happens if microbicides were to be introduced into a real life situation.

[SLIDE]

So the question has been raised, would a second control arm help. Two control arms have been proposed--a conventional matched placebo control an a condom-only, or what I prefer to call a no-gel arm.

[SLIDE]

The no-gel arm has, it would appear, certain advantages. It would eliminate problems associated with a placebo which might have a protective effect, and it would reflect real life.

But I would ask the question: Are these advantages real? Would it really reflect real life?

[SLIDE]

What are the disadvantages of a no-gel arm? I believe they come under two headings. First of all, differential behavior change within the population,a nd secondly, difficulty in achieving a uniformly high follow-up.

[SLIDE]

What are the behavior change issues, first of all? In a randomized clinical trial, participants usually behave differently to how they would outside the trial. They are being seen much more frequently, they are being counseled regularly. In a microbicide trial, they will receive regular counseling about safer sex.

Within the trial, behavior changes are not so important when comparing indistinguishable treatments if we want to look at the relative effects of two treatments. However, as we have already heard, a no-gel arm clearly unblinds participants and almost certainly results in differences in behavior change. Women allocated to receive no gel may choose to share the gel with those allocated no-gel. I mean, many women are actually going to help recruit others to the trial. Women will recruit their sisters, their cousins, their friends--and the reality is that most women will hope to be receiving gel. They will be very disappointed when they don't get it, however well we try to counsel people otherwise.

Consequently, what may well happen is that one woman will say, "Don't worry, I'll get a bit more gel, and you can have some of mine." And that

may be very difficult to measure, but the reality is it is likely to happen.

[SLIDE]

Could we allow for these problems, these issues, behavioral issues, in our analysis?

Sexual behavior data such as partner change, frequency and type of sexual intercourse, use of condoms are inherently very difficult to ascertain accurately. We could never be sure of the true differences between the distinguishable treatment arms. Consequently, interpretation of differences, I believe, would be impossible.

There are also, as I said, follow-up issues. However good our consent process, it's almost certain that many women will enroll into a trial, as I have already said, in expectation of receiving gel.

Women requested to attend for regular follow-up who receive no gel are likely to be less adherent--unless they manage to get it from another source--than those who receive the gel.

Without coercive incentives, women allocated no gel are more likely to default from the study than those receiving gel. And of course, the longer the study, the more likely that is to be

the case.

[SLIDE]

So I would say that at this point, we could conclude that the no-gel control arm would make the study impossible to interpret. Results from a study including a no-gel arm are likely to be, at best, of interest but at worse will be seriously misleading.

[SLIDE]

I want to return to the issue of collecting accurate sexual behavior data. Although, as I have already alluded, it is very difficult to collect, I believe it is very important to attempt to obtain accurate data--as accurate as we can obtain--in order to be able to better understand the results of our study.

For example, if we see no effect in one particular site, but we see effects in other sites, could that be explained by what we term "condom migration"--that is, women who are receiving gel, who have been using condoms, actually using condoms less because they don't think they need them.

[SLIDE]

How do we use the sexual behavior data? I believe that if a gel shows evidence of

effectiveness in most but not all of the sites in a trial, this may be due to differences, for example, in acceptability, difference in adherence and/or sensitive behavioral factors such as the frequency of anal sex--which we may have little evidence on as to whether it is being practiced unless we have good behavior data for our populations.

We need to know why we are getting different results from different sites, and I think it is extremely likely that there will be variation in results from sites if we have different sites from different parts of Africa, different populations, urban and rural.

[SLIDE]

So I come back to a point I alluded to a little bit earlier, and that relates to heterogeneity of sites. Is it a good thing or is it a bad thing?

You could regard it as bad insofar as it could reduce your change of demonstrating overall effectiveness. That would be true, of course, if you had actually been fortunate in identifying a site where you expected to actually be able to demonstrate an effect--but I don't think we are in such a fortunate position.

Alternatively, since a product may not be universally acceptable or effective, variation between sites could increase the chance of demonstrating an effect on the primary endpoint or at least explaining reasons for lack of an overall effect if we see variation in effect between sites.

And again here, this is where the sexual behavior data becomes important, too.

[SLIDE]

There has been some discussion, too, and it has been referred to by earlier speakers, about how long the Phase 3 trial should be. Both adherence to gel use and regularity of follow-up are likely to be influenced by the duration of the trial design.

Even persons who are under treatment for active disease, in such populations, we know that maintaining adherence is very difficult. I have a background in tuberculosis, and in fact in the days before short-course chemotherapy, there were very dramatic findings of how populations dropped off with time in terms of collecting their drug. Even though they were populations where the patients knew the seriousness of their disease and the importance of actually receiving it, by the time

you got to 12 months, the proportion of men and women who got TB who were picking up their drug could be as little as 25 percent of those who had been originally enrolled.

The problems have also been demonstrated, I think, in some of the HIV therapy trials in recent days as well.

Maintaining good adherence with preventive therapy can be even more difficult, and it can become increasingly difficult with time.

[SLIDE]

So we could ask the question, well, how short could the Phase 3 trial be.

Shorter designs of maybe six or nine months are more likely, I believe, to demonstrate proof of efficacy than studies requiring participants to be adherent, shall we say, for periods up to 24 months.

Long-term safety data could be obtained from such studies by following a subgroup of women for longer periods of time. Not all women would actually just stop being followed at six or nine months. We could go on following women beyond that time to collect long-term safety data.

Long-term effectiveness, because it will

be dependent on adherence, is likely to improve once proof of efficacy has been demonstrated, and we can say to women that we have good reason to believe that these products are going to be beneficial. We cannot say that at this point in time.

[SLIDE]

One of my final points relates to population selection. Proof of efficacy will be more difficult to achieve in certain circumstances--such as, if we include participants who are unlikely to benefit from microbicides--for example, those who are regular condom users or those frequently practicing anal sex. We would clearly make our work more difficult to actually identify an effect in a population.

However, restrictive inclusion criteria prevents subsequent generalization of our findings, and we must always bear that in mind as well.

The reality is that site selection and to a lesser extent, the study personnel that are conducting our studies are likely to be important in determining the outcome of our studies. You could even say it depends on who your friends are, which sites you have actually chosen, the ones that

you have experience with, which will have quite a major determinant on what the results of the study may actually turn out to be.

[SLIDE]

So in conclusion, if we are to reduce the number of new infections, we need a flexible approach to study design which will maximize our chance of achieving proof of efficacy and reducing the number of women likely to be infected in the next 20 minutes.

Thank you very much.

[Applause.]

DR. GULICK: Thank you.

What I would like to do is hold questions until we hear the two statistical presentations.

Let's now take a 20-minute break. We'll reconvene at 9:55.

[Break.]

DR. GULICK: Welcome back. We are ready to resume the meeting.

Our next speaker is Dr. Tom Fleming from the University of Washington.

Statistical Considerations for Topical Microbicide Phase 2 and 3 Trial Designs: An Investigator's Perspective

DR. FLEMING: Thank you, Dr. Gulick.

[SLIDE]

I am pleased to be here. The discussions that we have already heard have certainly pointed out that there are many challenging issues that we face with the design of topical microbicide studies.

What I would like to do is try to touch on a few of these key issues, and I will be talking about choice of controls, required strength of evidence, and what to do after Phase 1.

[SLIDE]

So let me begin by addressing further issues we have already discussed a fair amount today, that is, the role of blinding.

It has long been understood in clinical trials, particularly when you would have, let's say, a subjective endpoint such as pain that bias can occur if the treatment that the participant is taking is known to the evaluators--for example, where their judgment could be influenced by their being unblinded--it is known that if it is known to the participant or patient, there could be placebo effects. And if caregivers are unblinded, in those

settings where the endpoint, such as hospitalization, is one actually influenced by the caregiver, then the unblinding could introduce some bias.

[SLIDE]

If we look at the potential mechanisms of action of an intervention, using a placebo control as a comparator to the active microbicide would be an ideal approach to be able to estimate the antimicrobial effects of that intervention.

It has also, though, been recognized for a long time that there are controversial issues in some settings with the use of blinding. Pocock has addressed a number of these many years ago.

We look first of all at the practicality issues. Treatments or interventions need to be of a similar nature and cannot induce obvious side effects, so for this reason, a large fraction historically of comparative trials in the oncology setting, for example, have been unblinded trials.

Ethical issues are also important. Blinding should not result in harm or risk. So it wouldn't be ethical to try to induce within a blinded control in an oncology setting an intervention that would induce nausea, vomiting,

stomatitis, alopecia, et cetera, in order to achieve the blind.

There are a number of other important issues that really are key to consider when you are thinking about blinding in a microbicide trial. One of the issues is how serious is the risk of bias without blinding, as Pocock mentions. These risks are more serious with subjective endpoints. Fortunately, dealing with an HIV infection endpoint, it is a more objective endpoint such as survival would be in an oncology setting, and that reduces some of the risk of bias that would occur in an unblinded setting.

The importance of understanding efficacy and effectiveness is also critical. A microbicide intervention is by its nature not only made up of its antimicrobial components but also involves behavioral components, and understanding the global aspect of the effect of the intervention is critical, so understanding efficacy and effectiveness is important.

[SLIDE]

And it is also key to have adequate evidence to establish that the placebo is truly inert. So if we return to this consideration of

the potential mechanisms of action of a microbicide intervention, not only are those mechanisms antimicrobial effects, but the microbicide might also provide protection through physical barrier effects, lubrication effects, and other effects.

These components may in fact also be carried by the placebo. So a simple comparison against the placebo may actually be even underestimating efficacy.

In contrast, a comparison of the active microbicide against the unblinded control would incorporate not only the antimicrobial effects but also all of these other effects and would also be able to incorporate effects on risk behavior, being able to look, then, at a global estimate of effects or in essence on effectiveness.

[SLIDE]

Let me consider half a dozen specific circumstances to get a little bit more insight into what we might learn in a trial that would in fact have both a placebo control and an unblinded control.

To explore this, in each of these six settings what I am presenting on this slide is the annual risk in the active arm as well as the

placebo arm as well as the unblinded control arm.

In the lower left-hand side, we would have a situation where the annual risk is 3 percent in each of these groups, and we would clearly have a setting in which we would have established a microbicide with this particular mode of delivery in this population as being ineffective.

A more ideal circumstance would be where we would have a one-third reduction in transmission rate relative to both the placebo comparator group and the unblinded control group; and clearly we would have a positive circumstance there.

What I have presented in the upper portions in the right-hand column are settings where we still have a one-third reduction relative to the placebo control, but in this setting, we have about a 20 percent relative increase in risk-taking behavior in the blinded arms; here, a 50 percent increase in risk-taking behavior in the blinded arms.

When we would then look at the comparison not only against the placebo but against the open-label control, we would see that we still have evidence of effectiveness here, although there would not be net effectiveness in this setting.

In the left-hand column, we have two circumstances where we still have a one-third reduction relative to the open-label unblinded control. In this setting, we have a situation where we have about a 20 percent relative efficacy as estimated against the placebo, but by having the open label, we see a more complete sense of the true treatment effect, which is in fact potentially somewhat missed by a placebo that in fact is itself carrying some of the benefit.

This is a circumstance where we in fact have a one-third reduction carried by the placebo, but there is no additional antimicrobial effect. And in fact this is not hypothetical. In the past year, in another setting studying an antimicrobial where the FDA had urged the sponsor to have both a no-treatment open-label as well as a placebo, this is exactly the circumstance that arose in that setting.

How would we interpret results? What conclusions would we draw in each of theses settings?

What I would like to do is come back to that question after taking a moment to consider the issue about required strength of evidence.

[SLIDE]

A standard that has long existed within FDA for regulatory approval is to have two adequate and well-controlled trials. Essentially, statistical significance for each trial would be based on the strength of evidence by obtaining a one-side p-value less than .025--or in essence, if we have evidence where the result is sufficiently favorable that this result would occur by chance alone if there were no true treatment effect would only be 2.5 percent, that's the standard for strength of evidence of a single positive study.

When we have had major clinical endpoints, the FDA has been flexible to consider a single trial situation, a single pivotal study. These could be situations where the endpoint is death, stroke, loss of vision, or HIV infection. And in particular in these settings that are also involving resource-intensive trials, the FDA has considered applications based on single pivotal studies, and what I have noticed, a fairly consistent terminology that they use is that the strength of evidence for that single pivotal trial needs to be "robust and compelling."

When sponsors have asked, "What does that

exactly mean in terms of a p-value?" the FDA has correctly said, it's not so simple as a single p-value. The ultimate judgment about approvability of an intervention needs to take into account not just the primary endpoint, which is critical, but all relevant data--data on secondary endpoints, data on safety, external data and, importantly, data on quality of trial conduct.

My sense is that a proposed guideline for strength of evidence, then, when you are planning such a study might be to target a strength of evidence that might be midway between the strength of evidence of a single positive study and the square of this, which would be two positive studies--essentially, to be in a position that one would have sufficiently robust and compelling results even in the event that there may be certain irregularities that show up in the trial.

[SLIDE]

One study that is under design right now is the HPTN 035 trial, and I'll use this briefly to illustrate some of these concepts.

This is a study that is in fact planning to look at both the placebo control and an unblinded control, and we will be looking at two

active microbicide interventions.

It is targeting 33 percent effectiveness with 24 months of follow-up.

The question is with this particular design, for any of these pair-wise comparisons that may be made of active against control, how big does the study have to be; what does this actually mean in terms of events.

[SLIDE]

In Scenario 1, if we were looking at building a study to have strength of evidence, that is, the traditional 2.5 percent false-positive error rate, if we were trying to detect 90 percent power to detect a 33 percent effectiveness, that would take 256 endpoints. And essentially, in a setting that we are looking, about 4,000 participants per pair-wise comparison, or 2,000 participants per arm.

In Scenario 2, where we might be building for essentially a strength of evidence midway between that of strength of evidence of a single or two trials, again, if we are looking at 90 percent power to detect 33 percent effectiveness, essentially, it would take--as you might expect--about one-and-a-half-fold, or about 405

events, or about 3,000 participants per arm.

[SLIDE]

Essentially, what would the estimated effect have to be in these two settings? So, in Scenario 1, where we are essentially targeting a traditional 2.5 percent false-positive error rate, what I have plotted here in yellow is what the percent reduction in HIV risk may be in these trials, and essentially in this setting to achieve the strength of evidence of a single positive trial, your estimate would have to be about a 21 percent relative reduction. Strength of evidence of one-and-a-half trials, if you in fact achieve the 29.5 percent estimate reduction and a 33 percent would be the strength of evidence of two trials.

Not surprisingly, in Scenario 2, where we are actually looking at 405 events per pair-wise comparison, powering it in essence to the strength of evidence of one-and-a-half trials, it would take a less impressive estimate to achieve the strength of evidence of a single study--17 percent--and roughly 24 percent estimated reduction for a strength of evidence of one-and-a-half studies.

[SLIDE]

Now, in a setting where you have dual controls, what might in fact be a general guideline for strength of evidence against these two arms?

My proposal for illustration would be a setting where essentially, we require the .0025 for one of the comparisons, where the other one would just need to be at the traditional .025 level.

So specifically, then, if we obtained a compelling result against placebo, the strength of evidence against the unblinded control might only need to be supportive; or if the result against the unblinded control is in fact compelling, the result against the placebo may only have to be supportive.

[SLIDE]

With this as an illustration for targeted strength of evidence, then, what might the conclusions be in a trial where you had a comparison to the placebo and the unblinded control?

Let's return to these six circumstances here. Clearly, in the lower left-hand circumstance, we would conclude that it is a negative trial, a trial that has ruled out benefit. In the lower right-hand side, we would have clear

evidence of efficacy as represented by both the comparisons against the placebo and the open label.

In these middle scenarios, on the right-hand side, we would have compelling evidence against the placebo control and supportive evidence against the open label, which I would argue would also be a positive circumstance. Or, on the left, we would have compelling evidence of effectiveness and supportive evidence in the comparison against the placebo.

The illustrations up here on the top are illustrations where, on the left, we have essentially evidence of minimal effect of the antimicrobial components of the microbicide; and on the right, we have minimal evidence of effectiveness.

It has been argued by some that when you add the unblinded control, the end result is simply to make it more difficult to conclude benefit--and in fact, I would argue that that is not true. There is really symmetry here. I have underlined here the two situations where the unblinded control would give you a different conclusion than you would have had if it didn't exist in the trial.

And certainly in this setting where you

have evidence of no effectiveness, it does lead you to have concerns about approval of this intervention. But in this particular circumstance, if you would just look against the placebo control, you would have had an estimate of only a 20 percent reduction in transmission rate, whereas when you have added this additional insight from the open label, you are getting a clear indication that you may have in fact underestimated the efficacy by missing components of benefit that in fact were also carried by the placebo.

[SLIDE]

I'd like to spend a little bit of time talking about issues that relate to where do we go after Phase 1.

If you have in fact completed a Phase 1 trial with on the order of 100 participants, what would be the next proper step? Traditionally in clinical trials, we have gone to Phase 2 studies, and Phase 2 studies provide many important benefits.

[SLIDE]

One of the key areas of benefits of a Phase 2 study is it provides invaluable insights to allow us to design an improved Phase 3 trial. For

example, by conducting a Phase 2 study, we are able to learn a great deal about how to achieve timely enrollment of participants, high-quality study implementation, and high-quality data including retention. To achieve interpretable unbiased results, it is going to be extremely important to keep loss-to-follow-up rates low. We really should be targeting for 12-month follow-up 95 percent retention.

Phase 2 studies are going to give us important insights about how to improve our ability to retain patient participants in trials.

Adherence will also be critical, and Phase 2 studies can also provide important insights. We are not dealing with a vaccine that may require a one-time implementation. To achieve the full benefit of microbicide, we are going to need to have consistent adherence. How can we in fact improve the behavioral element of this intervention to maximize the adherence to the active microbicide, and also to maximize the adherence to condom use and other approaches to reduce risk of transmission.

So these are all insights that will be invaluable to the design and conduct of a Phase 3

trial that comes out of a Phase 2.

Traditionally, of course, as well, Phase 2 trials give us important additional clues about safety that will be important to have in hand before doing Phase 3 trials and, in addition to that, plausibility of efficacy by using biological markers and establishing effects on those markers.

Unfortunately, in settings such as topical microbicides, there aren't in fact biological activity measures that we can use to assess plausibility of efficacy. So what might be an approach to take rather than launching immediately into a full-scale Phase 3?

[SLIDE]

One additional approach to consider that I'll talk a little bit about would be a Phase 2B trial, or we might call it an intermediate trial. So in the setting of the 035 trial, if it is in fact conducted as an intermediate trial, the primary endpoint would in fact be the HIV infection rate itself, but essentially, we might be looking at a much smaller version of the study; rather than the 400 events per pair-wise comparison, we might be looking at a third to one-quarter that size--for example, 100 endpoints per pair-wise comparison.

The goal, of course, would be to estimate the true percent reduction in HIV infection risk, and the estimate of that, I will denote by delta hat.

[SLIDE]

So what we see on this slide is the nature of the evidence that we would obtain in an intermediate trial versus the full-scale Phase 3. So let me start with the full-scale Phase 3 trial.

[SLIDE]

In this particular setting, with 400 events per pair-wise comparison, we would have considerable precision--basically, our two standard errors would be plus or minus 17 percent--and recollect that we said earlier that when there were 405 events, a p-value of .025 would be obtained if you had essentially a 17 percent estimated efficacy; a strength of evidence of 1-1/2 trials, if you had an estimated 24 percent.

So what we see down here is that if in fact there truly is a one-third reduction, then you would have high probability, about 97.5 percent, of achieving strength of evidence of at least a single trial and about 90 percent chance of obtaining an estimate of 24 percent or higher.

[SLIDE]

Now, if instead you embarked on the intermediate trial, which would be about one-quarter the size, it would have roughly twice the variability. So that essentially you would have to observe now a 33 percent efficacy to be able to have the strength of evidence of a single trial.

Suppose we took the following approach, basically, a multiple-decision outcome. If you see 15 percent estimate of efficacy or less, you abandon the intervention. If you see 15 to 33 percent, you have encouraging evidence that would require confirmation in a Phase 3 trial. If you have basically 33 to 44 percent, you have at least the strength of evidence of a single trial, and 44 or better would in fact be conclusive evidence of benefit.

If in fact there truly is 33 percent efficacy, this is a strategy that has the desirable properties that you have only one chance in eight of abandoning the regime; you have three chances in eight, basically, of having evidence that would require additional confirmation; and you would have about a 50 percent chance of actually in this trial

achieving evidence that would be at least the strength of evidence of a single positive trial.

Another benefit of this approach is for an intervention that doesn't provide benefit. You have about an 80 percent chance of getting a more efficiency answer to that question without having to spend as much in resources.

One of the benefits of this is that if you do obtain evidence that is encouraging but not conclusive, a follow-up trial could in fact be smaller. It would only have to be a study that would provide the traditional strength of evidence of a single positive study.

An appropriate question, though, is if you get encouraging but not conclusive evidence, can you in fact validate that result; is it practical to do so?

[SLIDE]

To illustrate this issue, I would like to move to another setting that in fact in certain circumstances is very similar to what we are confronting today with microbicides. It is the surgical adjuvant therapy setting for colorectal cancer.

This is a setting where a surgeon can make

a complete clinical en bloc resection of the disease, but minimal microscopic undetected residual disease exists. It leads to the very significance risk of a 50 percent mortality within 5 years. For 20 years up to 1980, there had been repeated efforts of looking at adjuvant chemotherapy to try to reduce this risk, without success. So there was a very serious unmet need for survival hazards of 50 percent in this population.

The particular trial in hand was looking at 5-FU levamisole and levamisole, and this study, the North Central Cancer Treatment Group study, was basically a 2B trial looking at about 100 events per pair-wise comparison. This study showed very encouraging evidence--a 33 percent reduction in death rate--from both 5-FU levamisole and levamisole alone.

[SLIDE]

In spite of the fact that there was a serious unmet need for survival in this setting, it was recognized that confirmation was necessary. A cancer intergroup study was done of approximately four times the size.

[SLIDE]

So this is at least an illustration that confirmatory trials of promising but not conclusive intermediate trials can be performed successfully. It also illustrates the value of confirmatory trials because they can reveal both true positives and true negatives, and to look at this more closely, 5-FU-levamisole had a 33 percent reduction in death rate. That was exactly confirmed by the cancer intergroup trial. However, levamisole alone also had had an estimated 33 percent reduction in death rate, but the much larger, more reliable trial showed that in fact that was a false-positive conclusion.

So with this suggestive evidence of benefit of levamisole, it was actually proven to be an unreliable lead.

This confirmatory trial was extremely important because it provided much more reliable evidence so that people in fact were able to be treated with a regimen that in fact was beneficial rather than a potentially somewhat less toxic regimen but in fact one that was established to not be beneficial.

[SLIDE]

The question is could an intermediate

trial itself provide compelling results. An illustration of this could be provided by the HIVNET 012 trial that was looking at mother-to-child transmission of HIV, looking at two short-course regimens. And again, this was a study that had approximately 100 events per pair-wise comparison.

This study showed results that were in fact statistically very compelling, on the order of the strength of evidence essentially of two positive trials.

Well, this in fact arose by essentially having an estimate of a 47 percent reduction in transmission. So in this trial, we were right here at a 47 percent reduction that does in fact translate into compelling evidence of benefit.

[SLIDE]

In conclusion, just returning to the three points, for blinding, certainly a blinded control often is the gold standard. But we need to have reliable evidence that the placebo itself is inert, and might the physical barrier or lubricant or other effects that the placebo itself carries lead us truly to underestimating efficacy if we simply look at a placebo control.

Furthermore, in this setting, efficacy and effectiveness are relevant. Microbicide regimens in fact have both an antimicrobial component as well as a behavioral component. Understanding the global effect that this intervention would have in the real world setting is important.

There could be flexibility here, though. That is, certain trials such as the HPTN 035 trial could be studies designed to look at dual controls. It doesn't mean necessarily that all studies would have to have dual controls. If certain studies provide a foundation to understand more globally both the comparisons against placebo and against the open label, it is entirely possible that other studies could be designed by other sponsors that would simply have the placebo control.

Secondly, relating to standard of care, FDA has shown flexibility in allowing single trials in some settings. When they have allowed single trials, they have consistently asked that data be "robust and compelling." I believe sponsors would be well-advised, then, when planning single-study applications, to target strength of evidence that would be between that of one and two trials.

And just as a simple example of these

irregularities that can arise, in 2001, the Anti-Infective Drugs Advisory Committee was considering Zigris [phonetic] for another compelling unmet need setting, which is improvement of survival in severe sepsis patients. And in that particular trial, the results were in fact somewhat stronger than the strength of evidence of a single trial, one-sided .025. But there were, as is often the case in trials, irregularities. There were concerns in interpreting the data about inconsistencies in subgroups, about changes in the regimen, et cetera, and ultimately, that committee was left with a 10-10 vote, a split vote of uncertainty as to how to proceed.

In fact when we are dealing with a single trial, it is advisable to be targeting stronger evidence to provide results that are robust and compelling.

And finally, after Phase 1, particularly in settings where there is no biomarker for Phase 2 plausibility of efficacy, what is the right step? And I grant this is a very difficult issue. The HPTN in thinking about this issue had major jumps in jumping from roughly 100-person Phase 1 studies to a $100 million, 8,000 to 12,000-participant

four-arm Phase 3 trial. And in looking at this, those concerns were in part based on the fact that we don't have Phase 2 proof-of-principle biological markers to establish plausibility of efficacy and because, even though there had been extensive preparedness studies done to provide assurances that we could provide timely enrollment, high levels of retention, high levels of appearance, to be able to do so in the context of a 10,000-person study was something that the group was very uncertain about and much more comfortable moving into a 3,000-person study.

Ultimately, this is a decision that each sponsor will make. It may be in the judgment of sponsors appropriate to jump into a Phase 3 trial.

In closing, I would simply say that the goal is not specifically to get into a Phase 3 trial as soon as possible. The goal should be as soon as possible to complete Phase 3 trials that have robust and compelling evidence of a favorable benefit to risk.

Thank you.

[Applause.]

DR. GULICK: Thank you, Dr. Fleming.

Our final speaker of the morning is Dr.

Bhore from the Division and the Agency.

Statistical Considerations for Topical Microbicide Phase 2 and 3 Trial Designs: A Regulatory Perspective

Rafia Bhore, Ph.D.

DR. BHORE: Good morning.

My name is Rafia Bhore. I am a statistician in the Division of Antiviral Drug Products at the FDA.

Today I will be giving the FDA perspective on the statistical considerations when designing a clinical trial of topical microbicide for the prevention of HIV infection.

[SLIDE]

In this talk, I will first give an example of a Phase 2/3 clinical trial design of a topical microbicide in prevention of HIV infection.

Next, I will discuss the issue of whether such a trial will include two arms or three arms. I will also talk about the p-value that is conventionally required, whether it is a single large trial or two trials, and the criteria to declare that the clinical study or studies are successful.

I will also mention the statistical power

considerations in designing a clinical study, give estimates of sample sizes as well as mention other considerations that will be important to ensure the success of a clinical study in preventing HIV infection.

[SLIDE]

In this hypothetical example, the objective of the clinical trial is to establish the safety and effectiveness of an investigational microbicide in preventing HIV infection.

This is a three-arm study design. Test group participants are randomized to use the microbicide in conjunction with a condom for every sexual act. Control Group 1 will be randomized to placebo in conjunction with the condom, and Control Group 2 will only use the condom. This third arm has previously been referred to as the "no-treatment" arm by our FDA speaker, Dr. Teresa Wu. for the rest of this presentation, we will use these two phrases interchangeably.

In such a design, we recognize that it will only be possible to blind the test group and the Control Group 1. Control Group 2 cannot be blinded and so will be open-label.

[SLIDE]

Should this study really have two arms or three arms? If two arms, then which control group should be included? Remember that the goal of this study is to establish the safety and effectiveness of the microbicide being investigated.

[SLIDE]

First of all, why is inclusion of a placebo arm necessary? The placebo arm will provide a means to blinding investigators and participants as to which product is being assigned, whether it is the investigational microbicide or the placebo. This kind of blinding, as we know from clinical trials, in general maximizes the likelihood of obtaining an unbiased estimate of efficacy of the drug that is being investigated.

In a microbicide clinical trial, can we assume that the "placebo" is inert? In most cases, we do not know about the presence or absence of the antimicrobial activity of placebo, or it has not been proven in a clinical setting.

So the question is: Is the effect of the placebo a protective effect or a harmful effect? If a placebo has a protective effect, then the investigational microbicide will have to be proven to be better than a placebo that is protective.

This will make it more difficult to prove the efficacy of the microbicide.

So far, some of our speakers have not considered the possibility that if a placebo is harmful, then a microbicide that is shown to be better than a harmful placebo may at worst be harmful itself, or the microbicide could have a neutral effect in preventing HIV infection. So we would like the Committee to keep this issue in mind during the discussion. Or, at best, the microbicide could be beneficial.

Therefore, ideally, we want a placebo that is inert, and the placebo should have a neutral effect.

[SLIDE]

Next, why is the condom-only or no-treatment arm necessary? We know that the use of condoms is an established gold standard for the prevention of HIV infection. This arm is necessary because it will provide the real-world effectiveness of the microbicide in preventing HIV transmission. It will also provide data on the sexual behaviors associated with the use and non-use of microbicide products.

Thirdly, recall that this is the single

component of the other two arms that contain a gel and a condom--gel being either the microbicide or gel being a placebo. We need to know what is the contribution of this gel component in preventing HIV transmission. This arm is therefore also important in order to help validate the safety and efficacy data from the placebo arm.

[SLIDE]

Now we will talk about the level of significance needed in designing such a clinical trial. In statistical jargon, "level of significance" is the probability of making a Type 1 error. A Type 1 error is the error of incorrectly declaring that a drug is effective when it is not. So this is the error of getting a false-positive signal.

In order to prove the effectiveness, we want a p-value which is based on the actual data to be smaller than the predefined probability of getting a false-positive signal.

In simpler words, let's say, for example, a p-value less than .05 means that there is a smaller than 5 percent chance of declaring a drug to be effective when in fact it is not.

[SLIDE]

So, conventionally, when designing Phase 3 clinical trials, one trial that is designed at a one-sided .025 level, or a two-sided .05, which is double of that, provides the evidence of one trial. We look for a p-value based on the data to be smaller than this number .05.

In the regulatory environment, we conventionally require two adequate and well-controlled clinical trials each at a two-sided .05 level. Accordingly, two trials, each at the same level as before, will have an overall alpha of .00125, and hence, two trials will provide evidence worth two trials.

So if one considers designing only a single large trial instead of two adequate trials, we would still require the overall level of significance to be the same as two trials, which is p-value less than .001. And that is the same as the previous line.

In other words, the level of evidence with a single trial will need to be the same as that of two trials.

Some sponsors have proposed to us in terms of designing a smaller single large trial that will provide the evidence worth one-and-a-half trials.

This is a novel concept, and the Division of Antiviral Drug Products at FDA is open to discussing such alternative possibilities.

[SLIDE]

As I mentioned earlier, the conventional regulatory requirements for approving a drug for a single indication are two adequate and well-controlled clinical trials. So historically, this has been translated as follows.

Each of the two trials will need to show a two-sided p-value less than .05. And if there are two separate microbicide trials, then the question is will they be run in parallel, or will they be staggered in time. If they are staggered, one needs to think about how much gap in time there will be. Since this is a prevention of HIV indication, one may not be able to do a second trial after the first trial is completed and the results are known.

[SLIDE]

Alternatively, if a single trial is conducted to show prevention, this single trial will need to show as strong and robust evidence as to separate trials. It may not even be repeatable due to ethical concerns.

One trial will therefore need to show a two-sided p-value less than .001. And we showed the calculation of this number, .001, two slides ago.

[SLIDE]

Additionally, suppose that if one were to conduct only a single trial first, if we want to confirm the results of a single trial--that is, if we want to replicate the results of the study in the future--then, one important question is what is the probability of observing a statistically significant result--for example, p-value less than .05--if this clinical trial were to be repeated.

So, assuming that the effect size that we observe in the first trial is the true effect, and if the first single trial has a p-value less than .05, then the probability of getting a significant result in the future is only 50 percent.

Instead if the observed p-value the first time is .01, then the chances of seeing a significant result in the future, whether it is a future trial or in the actual environment, are higher. And in this situation, it is 73 percent.

If this p-value is even smaller, and it is .001, the chances of seeing a significant result

are much higher and increase to 91 percent.

[SLIDE]

Therefore, based on this discussion, when we consider the overall evidence of a single trial, a p-value that is less than .001 would be considered convincing; but a p-value that is greater than or equal to .01 would be inadequate. A p-value that falls in the gray area between .001 and .01 would be possibly adequate, provided that the results are consistent across various subgroups. This is also referred to as "internal consistency of the data." In addition, if the p-value is in this gray area, we would need to see other supporting evidence that is strong.

In the case of two trials, the collective evidence will be evaluated in a similar manner.

[SLIDE]

So if a three-arm clinical study is planned, what should be the criteria to declare that such a clinical study is successful?

A win here means that the investigational topical microbicide is proven to be effective in the reduction of HIV transmission. We would declare a win if the HIV infection rate in the microbicide-containing arm is less than that in the

placebo arm, and the HIV infection rate in the microbicide-containing arm is less than that in the condom-only arm. Each will need to show a p-value less than .001.

And because there is no need for multiplicity adjustment, the overall Type 1 error, which is the probability of observing a false-positive signal, is maintained at .001.

[SLIDE]

But why do we need superiority versus the placebo arm? Let's look at two scenarios where the microbicide wins versus only one of the two controls, but it does not win over the other control.

In the first case, if the HIV infection rates in the microbicide arm are lower than that in the condom-only arm, which is good, however, the rates in the microbicide arm are similar or could be even worse than that of the placebo, does this mean that the placebo is as good as the microbicide? This does not prove the efficacy of the microbicide.

[SLIDE]

In the second scenario, the HIV infection rate in the microbicide-plus-condom arm are lower

than that in the placebo arm, but they are similar or even worse than the condom-only arm. What does this mean? It implies that the use of microbicide in combination with the condom did not provide any additional protection than a condom alone would provide. So the microbicide is not shown to be effective.

[SLIDE]

Therefore, in order to prove that the microbicide is effective in preventing HIV infection, it needs to be proven that the microbicide is better than both placebo and condom-only.

[SLIDE]

Now we will show some examples of estimates of sample sizes for a three-arm clinical design. The sample size of such a clinical trial will depend on a number of factors. Firstly, it will depend on what is the background rate of HIV sero-incidence. We will assume that this is the rate of the sero-incidence in the control arms. As mentioned in the FDA background document, we have seen numbers as low as .5 per 100 person-years in the United States to numbers varying from 6, 7, and 9 in countries outside the United States.

Sample size also depends on the effect size. What is effect size? Effect size in simple terms means compared to the control, how effective is the investigational product. In the case of topical microbicides, sponsors are proposing that a new microbicide will further reduce the HIV sero-incidence rate by 33 percent to 50 percent. We will show some examples in the next slide to clarify what does it mean by a 33 percent reduction or a 50 percent reduction in actual numbers.

Thirdly, sample size will depend on the length of follow-up of participants--whether they are followed for 12 to 24 months exactly for each participant or whether the study continues until the last participation completes 12 to 24 months.

Since statistical power is directly related to the number of events observed--that is, number of HIV seroconversions--the more events are observed, the greater will be the power to detect the treatment effects. Therefore, it is advantageous to follow each participant until the end of the study so that the maximum number of events are observed.

Thus, longer follow-up will maximize the power of the study without having to add more

subjects.

And finally, statistical power is also an important factor affecting sample size. We will discuss that later.

[SLIDE]

Here are some examples of sample size estimates. In these examples, we have assumed that the endpoint is timed to HIV seroconversion. Duration of the study is assumed to be 24 months, and the power for comparison versus each control is 90 percent. These are estimates for a single large trial conducted at the .001 level.

Suppose the rate of HIV sero-incidence in any control group is 6 per 100 person-years--and for simplicity, we will call this 6 percent. A 33 percent effect size means that the number 6 percent is reduced by one-third, so two-thirds of 6 percent gives 4 percent. This will give a total sample size if 12,520. This is the total sample size.

Similarly, a 33 percent reduction from 7 percent rate of background infection means that the rate of HIV infection in the microbicide arm will be 4.67 percent. Or, if it is a 50 percent effect size, then a 50 percent reduction from 7 percent means a 3.5 percent rate of infection in the

microbicide arm.

As you can see, if the expected background rate of HIV infection in the study population is higher, then the sample size is decreasing. Also, if the effect size is higher, the sample size decreases as well.

However, if an unrealistically large effect size is assumed when in reality the microbicide has as small effect side, then there is a risk of underpowering the study. So the larger the effect size is assumed, the greater will be the risk of getting an unsuccessful study due to underpowering.

Sample size is also dependent on the length of follow-up. Shorter study durations will require larger sample sizes, while studies with longer follow-up will have smaller sample sizes. So we encourage the sponsors to collect data with longer follow-up, which will likely require a lesser number of participants.

[SLIDE]

Because we want to ensure the success of the trial, we must take into consideration the statistical power when designing a study. Statistical power is a concept that is opposing to

the concept of p-value. Statistical power is related to Type II error while p-value is related to Type I error.

Statistical power is one minus the probability of Type II error, so Type II error is different from Type I error in that it is the probability of incorrectly declaring that the microbicide is not effective when in fact it actually is. So Type II error is also called "probability of a false-negative signal." We want to minimize this probability of a false-negative signal and hence, we want to increase the power.

[SLIDE]

To determine power, we need to know the hypothesis to be tested. First, we want to test whether the microbicide-plus-condom arm has a lower HIV infection rate than placebo-plus-condom. Second, we want to test whether the microbicide-plus-condom arm has lower rates than condom-only. If we assume that the statistical power for each test is 90 percent, and we are seeking a 33 percent reduction in HIV infection from condom-only, then what is the overall power of getting a win for this study?

We define a "win" if the microbicide wins

against placebo and wins against condom-only.

[SLIDE]

This is a plot of the overall power of the study versus varying rates of risk reduction from placebo. When a background rate of HIV infection of 6 percent is assumed, we assume that this is the rate of HIV infection in the presence of the availability of condoms.

And since we do not know or have not proven the activity of the placebo, HIV infection rate in placebo is a moving target. At point zero, the microbicide is identical to the placebo, which is this vertical line. And as you move right, the microbicide has higher HIV infection rates than the placebo. So placebo is better as you move to the right.

And as you move to the left, the microbicide is much better than the placebo. When microbicide is much better than the placebo--that is, 33 percent reduction, 50 percent reduction, 67, and so on--then the statistical power of the study is at least 81.5 percent.

In other words, the chances of the study to be successful are greater when the effect size of the microbicide is equal to or better than the

placebo.

However, if the placebo is as good as the microbicide, or if the placebo is much better, the statistical power of declaring a win will drop dramatically.

[SLIDE]

I will also mention a few other important considerations in order to ensure the success of the study. First of all, we recommend that the study be continued until the last subject enrolled completes at least 12 months on study.

We also strongly recommend that the study personnel and sponsor be proactive in following the participants. This can be done by actively pursuing and identifying reasons for dropouts and continuing the follow-up after study drug discontinuation. If a participant is not followed after premature discontinuation of the study or study drug, this may raise a flag whether there are any drug-related safety issues.

Given that the first generation of microbicides will be used for a long period of time, we have a number of points to clarify regarding long-term follow-up versus short-term follow-up.

It is likely that most of the dropouts in a clinical study will be observed in the first year of follow-up, so participants who stay in the study through the first year will likely stay longer in the study through the second year.

Additionally, long-term follow-up will help collect more person-years of data because of long-term exposure.

And finally, if one observers higher loss-to-follow-up rates in long-term follow-up compared to a short-term clinical trial, this does not necessarily mean that the rates of loss-to-follow-up adjusted for time are higher with long-term than they are with short-term.

[SLIDE]

The second important consideration in design is monitoring the use of the condom and the microbicide. We recommend collecting data on the use of condoms as well as other barriers or drug use, because the evidence of efficacy is closely tied with the compliance of the product. There are four possibilities here: sexual acts with condom and with microbicide, without condom and with microbicide, and the other two are without microbicide and with or without condom.

We suggest that the sponsor frequently collect information on the number of sexual acts with or without the use of condom and number of sexual acts with or without the use of microbicide so this recommendation will also hold for the placebo arm.

[SLIDE]

Finally, another consideration when determining the overall power for such a three-arm study design is the allocation ratio. Allocation ratio is the ratio according to which the total number of subjects are distributed or randomized to each study arm.

Standard practice in clinical trial design is to allocate equal numbers of subjects to each group. This is called an allocation ratio of one is to one is to one.

Alternatively, one could choose to assign unequal numbers of subjects to the three arms. For example, one may choose to assign 1-1/2 times as many subjects to the microbicide group than the control groups. So in this example, more participants are exposed to the microbicide, but the control groups have the smaller number, and both controls have the same number.

This issue has been brought up because our preliminary analyses show that the alternative schemes of allocation ratios could likely maximize the power of a study to detect differences in HIV rates between test group and control groups. Also, such alternatives are proposed so that more safety data on microbicides could be collected. This alternative approach could be particularly applicable to the U.S. data where the goal is to maximize the amount of safety data that is collected in microbicide arm.

[SLIDE]

In summary, based on statistical considerations, I have discussed why a 3-arm design will ensure the effectiveness of the first microbicide ever for prevention of HIV and that such a study is studied appropriately.

A single trial for the development of a microbicide in prevention of HIV is acceptable. However, in the interest of maintaining regulatory standards, a single trial will need show the same level of evidence as two separate trials. And this was reflected by the need to show a p-value less than .001.

We also showed an example with estimates

of sample sizes for a 3-arm single-trial design. Clearly, we know that sample size will depend on the number of assumptions, such as the background rate of HIV infections, the effect size of the topical microbicide, the length of follow-up, the level of significance, and the statistical power of the study.

Topical microbicides are products that will potentially be used for the lifetime of a woman. Hence, an adequate length of follow-up of participants in a clinical trial will be extremely important in not only studying the safety of the product but also observing HIV infection rates due to long-term exposure.

[SLIDE]

I want to thank Dr. Teresa Wu and Dr. Debbie Birnkrant for their input in this thought process, and finally, thank you for your attention.

[Applause.]

Questions from the Committee

DR. GULICK: Thanks, Dr. Bhore.

We now have about an hour to entertain questions from Committee members.

Our first four speakers are in the front row, and there is a mike there which they can

respond to. Please come up to the front row, Dr. Van Damme. And then, Drs. Fleming and Bhore are at the table.

So we will entertain questions from the Committee or points of clarification, and as usual, let's try to refrain from actually beginning to discuss the issues, because we have the whole afternoon to do that.

Who would like to start us off?

Dr. Mathews.

DR. MATHEWS: I have a question for Dr. Van Damme. I was struck by the failure of the Phase 2 trial that you talked about to show the toxicity associated with the nonoxynol-9 preparation in terms of breach of the cervical vaginal mucosa, and I am wondering if it is not so much a sample size issue as a use condition issue in terms of frequency and so on, and if the problem is not necessarily solved by increasing the sample size but designing the Phase 2 trial in such a way that the use conditions would approximate what you would expect to see in a larger trial with a more heterogeneous population.

DR. VAN DAMME: I do not have a definitive answer to that. I do think the sample size is

important to enroll in that study where we considered the Phase 2 data for 320 women on which we had colposcopy events. The women who were in that study, as I said, were really out of Phase 3 study population, so they could use the product as they were going to use it into the Phase 3. Indeed, for instance, a center in Bangkok was part of that Phase 2 study, which had a much lower rate of use than other populations, but the biggest center in the Phase 3 trial and also driving the results which was observed is a center where the women who were in Phase 2, were in Phase 3. So those are driving the data, and those women were there from the very start.

DR. MATHEWS: And their behavior didn't change over the--

DR. VAN DAMME: Not that we could document, no--do you mean from the Phase 2 to the Phase 3?

DR. MATHEWS: Yes.

DR. VAN DAMME: No.

DR. MATHEWS: So what do you think actually explains the difference, then, why it was detected--

DR. VAN DAMME: I do think sample size.

We don't have enough power to detect an effect. You need a huge sample size to detect such an effect, which we never do in a Phase 2.

DR. MATHEWS: But the point estimates in the Phase 2 trial--did they even suggest a difference?

DR. VAN DAMME: A difference between the lesions in the two arms?

DR. MATHEWS: Yes.

DR. VAN DAMME: No.

DR. MATHEWS: So if the point estimates didn't even make the suggestion of a difference, it strikes me that it is not just a matter of sample size.

DR. VAN DAMME: Can you repeat your question?

DR. MATHEWS: What I'm saying is that that Phase 2 trial had something like 800 patients in it.

DR. VAN DAMME: No. The data for the Phase 2 includes 320 women on which we did analysis.

DR. MATHEWS: Okay.

DR. VAN DAMME: And that could also be indeed one visit into the trial.

DR. MATHEWS: All right. Thank you.

DR. GULICK: Dr. Wood, and then Dr. Sherman.

DR. WOOD: This question is for anyone from the FDA. Multiple presentations have all reinforced that any study of microbicides is going to be done in the background and the setting of condom use.

Is there any requirement for looking at whether or not the placebo gel vehicles or the microbicide itself has any effect on condoms in terms of stability, breakdown, chemical interactions, those kinds of issues?

DR. WU: That is a very good question, and at FDA, we regularly recommend the sponsor to conduct a condom compatibility study with both placebo and a microbicide to be tested.

DR. GULICK: Dr. Sherman?

DR. SHERMAN: Thank you.

The question initially will be for Dr. Karim, although others may choose to address this as well. It has to do with the data that you showed on condom use since that is part of the assumptions and the background of any of these studies that there is going to be a baseline level of condom use

and that everyone is going to be counseled to use condoms. You had indicated that 43.7 percent had been used after 5 weeks in the analysis you did.

Can you expand on that in several ways--first, what is the generalizability of these data to different populations around the world? Second, when you talk about use, is there any more specific data--was it used 43.7 percent of the time by 47.3 percent of women every time they had a sexual contact, or is there considerable variability where one woman uses it 47 percent of the time--because those things make a big difference in how we interpret that background protection.

DR. KARIM: Let me answer the first question on the generalizability of those results. The sites were chosen in terms of both rural and urban areas. So I would imagine that the data are reasonably representative of South Africa but probably not representative of anything more than that. I wouldn't want to presume that 43 percent of condoms taken from public health services in any other country would be used within 5 weeks.

But let me address your second question, which is the critical one, and it is probably

better to go to the COL trial. You heard that the Durban site had the largest sample size contribution, and it is certainly true that the patients or the subjects in the COL trial had high levels of product use.

When they were enrolled, we measured the condom use in the last sexual acts, and on enrollment in these sex workers, condom use varied from between 10 and 14 percent of sexual acts. Now, within that group, we have documented quite extensively that there is a very small subgroup who insist on condom use fairly routinely. But even in that group, they do not have 100 percent condom use.

Then there are others, particularly the newer women coming into the truckstops, who simply haven't yet learned how to get condom use from the truck drivers, so they have very low levels of condom use.

So the enrollment figure of 10 to 14 percent reflects that variability within the sex worker population. Upon enrollment, when we look at condom use within the first 4 months, it goes up to almost 40 percent. So there is no question that when you bring people into a trial like this, and

all you do is you keep telling them about the importance of condoms and you keep giving them condoms all the time, they do increase their condom use. But what we do notice is that that is not sustained, and certainly we did not see women being able to implement 100 percent condom use to any significant degree.

DR. GULICK: Dr. Barlett?

DR. BARTLETT: My question is directed to Dr. Van Damme, Dr. Nunn, and perhaps also Dr. Karim.

You have expressed concern that the follow-up rate in the condom-only group may be lower and that that may be a reason to have some apprehension about randomization to this strategy. I am wondering if there is any evidence from clinical trials that the follow-up rate would be lower, so ideally, if you have an evidence-based answer, and if not, do you have experience that would make you feel this way?

DR. VAN DAMME: I'm not sure that there is indeed evidence that women would leave the trial sooner or more than when they are assigned to the no-gel arm.

I think Mark River [phonetic] from

[inaudible] can report more accurately on their own trials in Cameroon where indeed there was a no-gel arm. This fear is based mainly on when we talk with investigators worldwide about how they feel the study population they will be recruiting will be looking at it. I myself was involved in a no-treatment arm in Antwerp, and I could already see that it was indeed more difficult to recruit women in the trial, but it is the fear also that when women are in a trial and are not using a product, it seems to then, "Why am I in a trial, and what am I contributing?" And we may counsel as much as [inaudible]--some things which, despite intensive counseling and explaining of the procedures--it is difficult to keep the women motivated and strictly to the science. Science is not always as easy to grasp.

So it is mainly based on a feeling that is expressed by investigators in the field.

DR. BARTLETT: Is the investigator from the Cameroon trial here?

DR. VAN DAMME: The statistician is here.

DR. BARTLETT: Would you mind addressing that question? I'm sorry I don't know your name. If there is data, that would be great.

I'm sorry, I can't hear you. Do you want to come up to the mike? Thank you.

DR. DOMINIK: I am Rosalie Dominik, and I am with FHI. We have the paper, and you specifically asked about the follow-up rates in the two groups, right?

DR. BARTLETT: Right. I think that's what we were referring to.

DR. DOMINIK: I was going to try to find it right in here--but maybe we can come back on that.

DR. GULICK: Yes, sure, we can come back. I'm sorry to put you on the spot.

DR. KARIM: I don't have a direct answer to your question. We have never done a trial like this before; that's why there is all the debate. But I can tell you that in trials that we have done, we have been able to maintain generally very high levels of follow-up. And certainly in the COL trial, we have had very high levels of follow-up.

We also have very high levels of follow-up in our regular cohort studies. We have several cohort studies where there is no intervention, and we are able to maintain follow-up.

I think it is very difficult to

extrapolate both of those to a setting where some people are getting product and others are not. So I'm not sure if it helps, but I am just giving you the chronology information that we do have.

DR. NUNN: If I might just very briefly answer the previous question, because the question was asked as to whether in fact as well, if there was evidence from other areas about condom use.

We are currently actually looking at the condom use in other countries--Zambia, Tanzania, and Uganda--and in fact we are finding much lower rates than in South Africa. Indeed, even after intensive counseling, it is not changing much. But there is a very different pattern according to what type of partnerships people are in, actually, as to whether they are using condoms or not.

To the question about follow-up rates in the context of a no-treatment. One of the problems here--and I am not thinking specifically about this sort of trial because we haven't conducted a microbicide trial before--but in other trials in other areas of infectious diseases and so on, we have always had a treatment of some kind, a placebo of some kind, in fact in order to be able to reduce biases. So actually, it is in part based, as one

of the other speakers said, on the perception of the local investigators about their concerns, particularly as the women are looking forward in anticipation to something which they can use apart from condoms which actually might be valuable to them. And I think their concern is if they were getting nothing, they would feel there was nothing in this trial for them.

What I would say also is that in fact in a preventive therapy study for opportunistic infections in HIV-infected patients, a large study is going on in Zambia at the moment where we have noticed as time goes on that there is a tendency to drop off. The women in a post-natal women's study we are doing, as they are followed up for one year, two years, three years, are less likely to come as time goes on. They just begin to get fatigued within the study and lose interest, too, despite encouragement to continue to continue to come.

DR. FLEMING: Before leaving this point, might I just add some evidence-based experience? I think Dr. Bartlett's question is very appropriate--what do we actually know from experiences? The HPTN has nearly finished one major trial that might provide some insight into

this. It is a 4,000-person comparative trial looking at an intensive behavioral intervention against a standard, and it is unblinded, open-label, as I said, 4,000 participants. And interestingly, in this experience, the challenge in retention has been much greater in the active experimental arm. We actually have a higher retention in the control arm. And as we have been monitoring this study, we have been having to work extraordinarily hard to actually bring the retention rates in the experimental arm u to the level in the open-label control arm.

The second interesting point about this is in fact, I think, consistent with the point that I think Rafia was making in her presentation, and that is, this is a study in which the participants are followed for 3 to 4 years, and the loss-to-follow-up rate was much higher in the first 6 months. We probably lost 5 to 8 percent in the first 6 months. Out to 3 years, the cumulative loss-to-follow-up rates are only about 12 percent.

So a large fraction of those that were lost over 3 years of follow-up were actually lost in the first 6 months, which provides some additional incentive for the fact that as you

follow longer in time, you get a lot more events without in fact correspondingly have a lot more additional loss-to-follow-ups in that particular trial.

But it is interesting in this one experience that the reverse of what we are hearing being predicted actually occurred in this 4,000-person open-label trial.

DR. GULICK: Dr. Fleming, I have a follow-up question to that. Can you tell us what the intervention was in that study?

DR. FLEMING: Yes. It's called the HTPN 015 trial, and it is a randomization in MSNs, men who have sex with men, looking at standard behavioral intervention against a more intensive behavioral intervention to try to reduce risk-taking behavior and improve protection against transmission.

DR. GULICK: And what is the interpretation for the differential rates of follow-up? How do you explain that?

DR. FLEMING: Well, it's always speculation as to whether or not people are leaving for various reasons. I think the best speculation in this setting might be that it is a more

intensive, burdensome involvement to be involved and active, and that could in fact be influencing. But I have to say it is not perfectly clear what all the factors would be.

DR. GULICK: Thanks.

Dr. Paxton, a follow-up point.

DR. PAXTON: Just a question--you said that most of your loss-to-follow-up occurred in the first 6 months. Was that group substantially different in terms of their risk behavior when you looked at them?

DR. FLEMING: It's a very good question as well. It's always very important to do everything possible to fully retain people, because in most instances, missing-ness is informative, i.e., those people who aren't followed are different from those who are.

This particular trial, this 015 trial, had a series of eight to nine behavioral interventions over a 6-month period. There is a striking relationship in that those people who were predominantly going through all of the intervention were in fact then retained. Those people who were dropping out of the intervention early in fact were also much less likely to be retained and, when we

looked at their baseline characteristics, were in fact associated with characteristics that typically would characterize them as being at higher risk.

So there is loss of events and hence there is loss of efficiency when you have missing-ness, but of much greater concern is the bias that is induced if there is differential loss to follow-up in people who are leaving being different from those who are being followed. And some people have said, well, we'll correct this--let's say there is 20 percent missing-ness--we'll correct this by increasing sample size by 20 percent. And I say, well, that gives you a more precisely biased estimate. Your only true correction is to really ensure that we have procedures in place to minimize loss-to-follow-up.

DR. GULICK: A couple of follow-up points--Dr. Bartlett and then Ms. Heise.

DR. BARTLETT: So, Dr. Fleming, the HPTN 015 trial is being done in MSMs in the U.S. and Western Europe, and the loss-to-follow-up rate at the greatest is about 12 percent.

DR. FLEMING: Yes. The overall retention through 3 to 4 years is about 88 percent, so there is an annual average retention rate of about 97

percent annualized.

DR. BARTLETT: But it would be fair to say that that's a really different population than what we are going to be talking about.

DR. FLEMING: Indeed it is a different population.

DR. GULICK: Ms. Heise?

MS. HEISE: I think the field has very little experience to go on. I believe you have the only experience that you will share with us in a moment. But I do think there has been behavioral and social science data done at these sites are part of the preparatory work. And I think the concern is less about whether or not you can enroll people in let's say just a condom promotion study and follow them successfully, but what happens when you have a group of women who are very interrelated and one thing that everyone wants a lot and the others do not.

In these trials, there is up to a year or more of preparatory work done in the community about the trial coming, and education on microbicides, and the possibilities. And it does create--which is very difficult to counterbalance--this real desire--these are

desperate women, and they desperately want something to try to use because they already have the experience that condoms don't work.

So we have to work, or at least investigators have to work really hard to try to counterbalance the notion of the hope that something will work. When you have that strong a hope, and you have some groups of women who are getting the hope and some who are not getting the hope, that's what creates the problem, I think--at least that is the fear. And I think in your trial, it actually wasn't borne out, if I recall correctly.

DR. DOMINIK: Well, the study statistician actually isn't here. But there were 1,200 participants in this trial where participants were randomized to either the gel-plus-condom arm or the condom-only arm, and this was only a 6-month study, so it is a little different from some of the studies that we're talking about, but there was an extremely high follow-up rate achieved in this study--in fact, there were only 20 participants lost to follow-up, but 13 of those were in the condom-only arm and 7 in the gel-plus-condom arm.

Also, with respect to reported condom use,

in the condom-only arm, participants reported using condoms in about 87 percent of acts versus 6 percent less often condoms were used in gel-and-condom group. Of course, that is just reported condom use. We don't really know true use.

DR. GULICK: And did I understand correctly, just to clarify, that this is really the best data we have right now to try to answer this question?

DR. DOMINIK: Somebody else would have to answer that.

DR. VAN DAMME: Yes. As far as I know, it is the only microbicide trial which has been done. I talked about effectiveness with the no-treatment arm; as I mentioned, we did a no-treatment arm in the safety study before.

DR. GULICK: Dr. Flores and then Dr. Haubrich.

DR. FLORES: In addition to the potential differentials in lost-to-follow-up, I think we have to be very concerned about potential differentials in actual behavioral impact of being in an active arm versus being in a control no-treatment arm. And I would argue that we could expect that in the

placebo arm, the effect on placebo is zero, both in efficacy and safety if that is equal to the no-treatment arm.

If this were a vaccine trial where you compare to vaccines, and one had no effect, I would argue that you have the tendency to combine the two control arms and therefore have an impact on the power of the study to analyze. I'm not suggesting to do that, but I think if we really feel that it is possible to have a control/no-treatment arm that would be somewhat a surrogate of a placebo in addition to placebo, then we need to make sure that in addition the potential share of product, the potential differential in follow-up rates, and behavioral impact are going to have to be an important factor to take into account.

DR. GULICK: Again, let me suggest that we try to avoid getting into the discussion at this point and stick to questions. Those are important points that we'll get back to in the afternoon.

Dr. Haubrich and then Ms. Heise.

DR. HAUBRICH: My question is for our statistical presenters. One scenario that in Dr. Fleming's talk I didn't see addressed would be if, in the no-treatment control, the condom-only arm

actually ended up doing better than both other groups because in the gel receivers, there was a reduction in condom use because they perceived that the gel was better--it's a new thing, they don't need to use condoms, they can get more money from their clients, et cetera--so the two questions are how do we deal with that, because you could say you could try to look at that by looking at condom reported use behavior, but if reporting of condoms or sexual acts is anything like adherence to antiretroviral therapy, we have solid data now, based on MEMSCAPS [phonetic] that it is notoriously underreported.

So how would we deal with that, and what would be the outcome if a study showed in fact that the treatment was better than the control, but both were significantly less than the no-treatment condoms alone?

DR. FLEMING: I'm just trying to best understand the exact scenario. It sounds similar to what was in the six scenarios I gave the upper left-hand scenario where the condom arm was definitely better than the open-label, but the microbicide arm didn't show up as being better than the condom-only arm; is that essentially the

circumstance you're talking about?

DR. HAUBRICH: Well, unless I'm looking at the wrong slide, it looks like the condom arm and the treatment arm are the same--

DR. FLEMING: Yes--2 percent, 2 percent, 3 percent.

DR. HAUBRICH: --and the control.

DR. FLEMING: So if you just give your scenario in terms of percents, what setting are you asking us to--

DR. HAUBRICH: No. It's similar to that except that, say, the treatment is better than the control, so 2 percent, 1 percent, 3 percent.

DR. FLEMING: Well, in fact if that occurred, which is even a more extreme example, what is evident when you would compare the placebo to the open-label is that either the placebo itself is extremely beneficial or adherence to the blinded arms are very much higher so that the risk levels are much less. In that setting as well the one that I gave that is less extreme, you would come away with a clear indication that the antimicrobial effect of the intervention is not adding, so you certainly wouldn't be marketing that microbicide, although it could give clues that other elements of

the intervention carried by the placebo, specifically, the physical barrier, the lubrication effects, et cetera could be in fact protective.

And I mentioned briefly that there are many other settings other than topical microbicides that the FDA has considered with sponsors the merits of having both placebo control and open-label control in settings where there are uncertainties about whether the placebo is inert and in settings where understanding where globally effectiveness is important in addition to efficacy. And in one such setting in the past year, this very scenario is what arose. There was no additive effect of the antimicrobial agent, but the placebo was much better than the open-label.

DR. GULICK: Dr. Bhore has a follow-up.

DR. BHORE: Yes. I want to clarify the question asked by Dr. Haubrich.

Are you trying to ask about a scenario where the no-treatment arm shows greater reduction in transmission than the other two arms? Is that what you are asking?

DR. HAUBRICH: Yes.

DR. BHORE: Well, if that happens, then let's give an example in terms of numbers. Let's

say the infection rate in microbicide is 3 percent, placebo is 3 percent, but for condom-only, it is only 1 percent, so condom-only or no-treatment--

DR. HAUBRICH: What I actually meant is let's say 4 percent in the treatment--2 percent in treatment, 4 percent in control, and 1 percent in the condom-only. So that essentially what happens is people stop using the condoms in the two gel arms so their--

DR. BHORE: So that is an example of the scenario I showed where I said the microbicide turns out to be better than the placebo arm, but it is almost the same as condom or it is worse. So 2 and 1 percent, we don't know if that's statistically significant, and in that situation, then, you have to ask the question: Well, microbicide is showing to better than placebo, but we don't know if the placebo was harmful. Is that why it showed placebo had higher rate, or whether truly the microbicide is good? So if the microbicide is showing 2 percent and no-treatment is showing 1 percent, the question is what is the microbicide adding to the condom-only, to the condom component. So that raises a dilemma.

But of course, we would have to look at

the collective evidence if such kind of data arises, because we would look at consistency of the data internally and whether there is any other supporting evidence. So this could become a review issue when we look at the data.

DR. GULICK: Ms. Heise and then Dr. Washburn.

MS. HEISE: I have two questions, and I direct them to whomever might have data to address them.

One is we have talked about threats to validity in terms of loss-to-follow-up, and I heard Dr. Bhore say that if we go longer, we get more events and whatever. But I'm wondering what we know about rates of pregnancy in these cohorts. My assumption is that in many cases, the women who become pregnant during the trial, so over the 2-year rate, would go off product and then be lost to a potential event. And my experience is that even women who say they will use contraception and are not necessarily desiring to have a pregnancy in the 2-year, that many women within the developing country settings that we are working in actually do become pregnant.

So I was wondering if anyone could comment

on whether there is any data about the potential impact of pregnancy on follow-up rates and how that would influence shorter follow-up times versus longer follow-up times.

That's the one question.

DR. NUNN: I'll give a partial answer to this question and also just make a brief comment on the previous one about the condom use.

A point that I hoped to have put across earlier in my presentation was that we do get tremendous variation between different sites in Africa. I mean, condom use in South Africa compared to condom use in places like Zambia and so on is very, very different in rural areas of Zambia. We are talking about a situation where getting people to use condoms is actually very difficult.

As far as pregnancies are concerned, in the early data that we have actually gotten from our feasibility studies we are conducting, we are showing, for example, in a site in Johannesburg that in fact we are getting very, very few pregnancies because they are using effective contraception in that population. But the data that we are getting from Tanzania and from Zambia

is quite different, where in fact they are not using the same level of contraception, and we anticipate that in a trial context, quite a high proportion of women will become pregnant in the course of a trial. And of course, the longer the trial goes, the greater the chance that that will be the case.

In Tanzania, we actually asked the women about their intention to become pregnant in the next 12 months, to look to see whether we could exclude those who intended compared to those who didn't. We actually found that those who intended to become pregnant were less likely to become pregnant than those who didn't, so it didn't actually work.

[Laughter.]

DR. VAN DAMME: I'm not sure I really understand the question. In COL 1492, we did tests on pregnancies, yes, quite a lot.

MS. HEISE: And did they continue on product, or were they lost--I mean, did they stop product?

DR. VAN DAMME: We did not consider them lost. They were discontinued from product. They could stay in the follow-up trial, but they were

discontinued from the product, yes--unless a woman expressed--may I say this here--unless a woman expressed that she wanted a termination of pregnancy.

DR. KARIM: I don't remember the exact pregnancy rate in the COL trial, but I can tell you in one cohort where we followed young women age 18 for about 2-1/2 years, close to one out of four became pregnant during that period--and these are very young women who are in their most reproductive period, and the use of contraception in that group is quite low.

I do think that that is a major consideration, that these women when they become pregnant remain at risk of HIV, but they are not using product anymore. And in the intention-to-treat analysis, of course, that pushes down our ability to show a difference.

So it is a major consideration when we have very long follow-up periods.

DR. GULICK: And data from the Cameroon study?

DR. DOMINIK: The earlier Cameroon study that was a one-year study of an N-9 film versus a placebo film that also had about 1,300 women, there

were only 5 women overall who became pregnant during that study, but that was a sex worker population.

It was also a very small number of pregnancies in the 6-month Cameroon trial, but I don't have those exact figures.

DR. GULICK: Thank you.

Dr. Washburn and then Dr. Englund.

DR. WASHBURN: This is a question for any of the presenters who might have any information about this. Commercial condoms that are available in drugstores, many of them have lubricants on them. Is there any evidence whether those lubricants affect HIV transmission?

We recommend to our patients that they use condoms to prevent HIV transmission outside the context of these studies, so one would hope that those lubricants are at least neutral--so an idea comes up that we can talk about this afternoon.

DR. GULICK: Dr. Birnkrant, do we have any data?

DR. BIRNKRANT: Well, there is, I believe, a lack of data with regard to N-9 impregnated condoms. That is, it is not really known whether N-9 impregnated condoms are any better than condoms

without N-9 in them.

With regard to more inert lubricants, I don't think we have that type of data to show that the lubricated ones are more effective than the non, except when it comes to breakage rates, perhaps.

DR. GULICK: Someone is signaling me from the audience. If you have some data, we would be happy to hear it--and please introduce yourself, too.

DR. FARLEY: I am Tim Farley from the World Health Organization.

I don't have data which addresses this directly, but I can tell you the most common lubricant in condoms is just a silicone oil. I am not aware of any information that indicates that that is protective against HIV.

The other issue which is a concern, of course, is if people are using N-9 condoms, but as far as I know, all the studies that have been in the field and are thinking of going in the field are specifically going to be providing non-N-9-lubricated condoms.

So I think we can be reassured that the lubricant in the condoms which are used is not

active in any way.

DR. GULICK: Thanks, Dr. Farley.

I have Drs. Englund, Stanley, and then Paxton.

DR. ENGLUND: I pass.

DR. GULICK: Okay. Dr. Stanley.

DR. STANLEY: I am just trying to get a handle on the behavioral aspects, and I guess perhaps Dr. Nunn or Dr. Karim--can somebody summarize for me what we know about changes in condom use behavior upon enrollment in all the clinical trials that we have been talking about and particularly when they are getting something additional? We really need to get a handle on understanding that in this population because that is where these studies are going to be done, and I am just having a hard time getting a grasp on that--I mean, if people looked at before enrollment and then after and things like that.

DR. KARIM: I can only reiterate some of the data which we know from the COL study. The COL study used coital logs in order to measure condoms. And we actually determined later on that it wasn't a very accurate measure in that women were sometimes seen filling out the logs while they were

waiting in the waiting room.

So we do have that as a genuine measurement problem. What we do know from the COL trial is that condom use on enrollment--and in fact, we had done several studies before this cohort was enrolled looking at condom use--we know that condoms were used in aggregate in about 10 to 14 percent of sexual acts. It varied over the years that we measured it.

However, we do know that when we put them into the trial, in the first 4 months when we looked at it, condom use did go up very substantially. Whether that is because they thought we expected them to say that they had used the condoms that we had just spent all this time trying to tell them they should be using, I can't answer that, but I would be surprised if condom use didn't go up. However, it was not sustained, and that was the other part.

DR. GULICK: Others? Again, I'm sorry, I don't know your name. Please introduce yourself at the mike before your follow-up comment.

DR. STEIN: Dr. Stein, Columbia.

I had some data also from the sex workers in the COL 1492 which I haven't discussed. I have

this from Dr. Gita Ranjee [phonetic], Joanne Mantel [phonetic], and Linda Mayer [phonetic], who did a follow-up series of focus groups with women who had been on the COL 1492. They had been told the results of 1492, which was negative, and they had also been told repeatedly that the microbicide was different from the placebo and that they were to use a condom. And I have actually some of the conversation--I was going to enter into this later--some of the conversation in those focus groups.

DR. GULICK: I'm sorry--could you speak right into the mike?

DR. STEIN: They felt that the condoms were cleansing and probably kept out what was harmful in the semen, and that so good did it feel that they rejected the male condom in favor of the gel. And they had, of course, been strongly and repeatedly counseled against doing just that.

So we do have some information that after being on the gel for some time, they said, "Good," which is very good, of course, for the future of microbicide testing, but is problematic in terms of the trial.

DR. GULICK: Was there any data from the

Cameroon study? I'm sorry we keep coming back to you--but in terms of changes in condom use before and after enrollment into the study.

Dr. DOMINIK: At baseline in the original, the 1991 study, about 45 percent of participants said they had used a condom during their last act; and condom use during the trial was reportedly sustained at a very high level of around 90 percent.

DR. GULICK: In both arms.

DR. DOMINIK: Right. But that was a blinded study.

In the study where we had an unblinded arm, about 60 percent of participants reported that they had used a condom during their previous act at baseline; and then, during the trial, in the condom-only arm, there was about 87 percent condom use, and in the other arm, the N-9, 81 percent condom use was reported.

DR. GULICK: Okay. Is this a follow-up comment?

MS. HEISE: This is more data.

DR. GULICK: More data. We like that.

Ms. Heise?

MS. HEISE: Unfortunately, it is not here,

but there have been two global reviews, one by UNAIDS and one by the London School of Hygiene and Tropical Medicine, that specifically look at all of the data both in terms of condom use rates pre-intervention and condom use rates different types of interventions.

And one thing--even across widely differing scenarios, I think there are two truths that come out of both of those studies. One is that the rate of consistent condom use that you can achieve is most defined by the type of partner that you are talking about. So that, for example, the very same people in this very same intervention done trying to get people to use condoms with a casual, a new, or a paying client achieve much higher rates of consistent condom use than where it is being introduced with a regular partner.

So for example, even in these rates where you have sex workers who are achieving 90 or 80 percent consistent condom use with clients, they aren't using them with their boyfriends or their husbands.

So when you talk about condom rates and what can be achieved, you have to think about who you are enrolling and what type of partner they are

talking about. And that is consistent across every, single study.

The other thing you see is that people over-report condom use, especially in the context of trials. So you have lots of examples where people are saying they are using them 100 percent of the time, but they are getting pregnant or they are getting STDs. So we know that overreporting of condom use in terms of social desirability in this trials is a problem that is very difficult to manage. And I can give the committee any of those reviews if you are interested.

DR. GULICK: Dr. Van Damme, a follow-up?

DR. VAN DAMME: Yes. I can confirm with Lori that also in the COL 1492 trial--again, these are self-reported data--that indeed condom use with clients was achieved at a much higher level than we could achieve with what we call regular partners in the trial.

DR. GULICK: So just to clarify this point, and then I am going to come back to my list, I promise--your question, Dr. Stanley was how much data do we have on condom use before enrollment into a study and then after enrollment into the study. And if I understood, the data from the

Cameroon study was that rates went up, but they went up in each arm.

Is that correct? You said it was about 60 percent of baseline and then on the study, it was 81 to 87 percent in the two arms.

DR. DOMINIK: Yes. That is true for COL 1492, too.

DR. GULICK: Thank you.

Waiting patiently--Dr. Paxton?

DR. PAXTON: Actually, I have a question, and I'm not sure to whom to address it, but it's about the potential for gel-sharing.

I personally find the theoretical arguments about how this might occur and the rationale behind it to be quite compelling. But I was wondering, for example, from the world of antiretroviral treatment in resource-poor settings, is there any data that we have from that showing that people might share their drugs? I remember when that was starting several years ago, people would said people will take their drugs and give them to somebody else they know who is infected.

Did that in fact occur?

DR. GULICK: Dr. Haubrich has some data.

DR. HAUBRICH: I have no data, but I have

anecdotal experience from our training with African military groups where the availability of antiretrovirals is extremely limited, and they said sharing is quite common.

DR. GULICK: Okay. Dr. Nunn?

DR. NUNN: I just wanted to say--it wasn't an antiretroviral situation; it was actually antibiotic prophylaxis where women had been enrolled into a study, their partners discovered in fact that the women were in the study, and they didn't like it at all, and they either said, "I'm going to have some of that drug, or you aren't going to be in the study," or in fact they actually told either women to get out and leave home.

So in fact there was the sort of feeling--this was men and women, of course--but there was the sort of feeling of why should some people have it and not others. I know in some studies now with antiretrovirals, we have to look very carefully, like giving antiretrovirals to children without giving it to their parents, so in fact the design actually makes sure that we are incorporating the parents and getting them treatment as well, because you can't realistically expect them to say we are giving you what could be

effective treatment, but we're going to deny it to another member of the family.

So I think we are aware of the problem, but I don't think there is any other data on antiretrovirals from recent experience.

DR. GULICK: Dr. Barlett, a follow-up comment.

DR. BARTLETT: Just a historical comment to Dr. Paxton's question. We were involved in the original Phase 2B/Phase 3 study of AZT, and indeed, there was some sharing of drug among study participants in that trial that was done in the U.S. And if anything, the bias that is introduced is to diminish the difference between groups. So with regard to the U.s. context, we saw that as well.

DR. GULICK: Dr. Englund, a follow-up?

DR. ENGLUND: Two things. First of all, I think there is good documentation that there is drug-sharing. In pediatric studies and studies ongoing right now in Kinshasha [phonetic] in Zambia, we will only treat children when the parents are simultaneously being treated because of documentation of drug-sharing. So that is well-known.

And that brings me to my question for perhaps our honored guest, and that is are there age differences. We are hearing good data showing that our younger girls are the ones who are getting infected, and that certainly is what I see in inner city Chicago as well as in Africa. And certainly who we would aim an intervention at potentially, I saw your study enrolled girls down to age 16, which doesn't quite capture it, but it's getting down there.

What are we seeing in terms of age differences in condom use and the pressure that these younger girls may be getting?

DR. KARIM: I actually don't have data from Hlabisa on condom use in young girls, but I have data from Wulanladla [phonetic], another rural area closer to Durban where we have been following girls as young as 12. These are girls who are coming in either for family planning or they are coming in as pregnant women for antenatal care. And we have been following them up now for the last 8 months or so.

Condom use in this young age group is negligible. It is so low that we are only occasionally finding them using condoms. So

although we are now using hundreds of millions of condom pieces, my suspicion is that most of those are being used in concordant sexual acts and largely in older groups.

The big problem that we have with these young girls is that they are having sex with much older men, where they are really quite powerless in terms of their ability to insist on condom use. There is also a tendency in this group for slightly more violent or more aggressive sexual behavior as well.

DR. GULICK: Dr. De Gruttola, and then Dr. Brown.

DR. DE GRUTTOLA: I have a couple questions for Dr. Van Damme or Dr. Karim or anyone else who may have the information.

Dr. Karim mentioned that following pregnancy, the product may be discontinued in the course of one of these studies, and that would lead to an attenuation of the effect, potentially. I also wonder if there are issues about following women who are pregnant if it is more or less difficult to follow. Obviously, if there were effects of the intervention on pregnancy as well as on transmission, differential follow-up could

complicate interpretation. So I just wondered what the experience was in Dr. Van Damme's study or anyone else in terms of following women who are pregnant and in terms of continuing use of product during pregnancy.

DR. VAN DAMME: In the trial, they were not allowed--as far as we could control it--to continue product use once they were pregnant. So I don't think we can speak on that.

DR. DE GRUTTOLA: How about follow-up of the women after they became pregnant?

DR. VAN DAMME: That was more difficult since women who are pregnant, there was [inaudible], since we discontinued their product, of staying in the follow-up of the trial.

DR. DE GRUTTOLA: So did you have a sense that you were losing the majority of them to follow-up of the pregnant women, or--

DR. VAN DAMME: I do not have [inaudible].

DR. DE GRUTTOLA: I see.

DR. GULICK: Dr. Karim had a follow-up.

DR. KARIM: Just to comment--we were one of the sites and the largest site in that trial. The one big problem we had was once the women became pregnant, they left the truckstop, and that

was the way in which we maintained the follow-up. So that was a real big problem for us to keep them in the study.

However, they do eventually come back to the truckstop, so we would have some blood at some point in those subjects, but they haven't been using product for quite a while in the meantime.

DR. DE GRUTTOLA: But it would certainly help in terms of completeness of follow-up, as you point out.

DR. VAN DAMME: A lot of the pregnant women also choose to terminate the pregnancy, so they come back into the trial.

DR. GULICK: Dr. Wu had a follow-up comment, and then we'll come back to your next question.

DR. WU: Yes, I would like to make some comments regarding pregnancy and being retained in the trial.

DR. GULICK: Can you speak up?

DR. WU: Typically, for any drug, for any microbicide, before being administered to humans, they have to undergo a reproductive toxicity study. There are several stages. Usually, the first stage is for fertility, the second stage is to check

embryo toxicity. And most topical microbicides have to go through this test before they can be given to women of childbearing age.

However, if they are willing to go all the way up to the third stage, that is, perinatal toxicity testing, also conducted before getting into human trials, then pregnant women can be given this microbicide, because in animal toxicity, all three stages have been cleared in terms of toxicity.

However, most sponsors only conduct up to two stages and leave the third stage sometime during Phase III clinical trial. Then they do concurrent animal testing. Therefore, once the woman becomes pregnant, the woman would discontinue drug administration, but once the child is born, after a certain period of time, they are allowed to come back. Some sponsors have used this type of clinical trial design, and FDA is supportive of it.

DR. GULICK: Thanks.

Back to you, Dr. De Gruttola.

DR. DE GRUTTOLA: I had one question on Dr. Van Damme's slide on CONRAD's approach to design of these studies. In that slide, you listed a one-year retention of 80 percent, and obviously,

that high loss-to-follow-up could be a concern regarding bias as well as attenuation of power. And I believe you mentioned that there was some evidence of problems of retention that would make this a plausible rate, so I was wondering if you could comment on that.

DR. VAN DAMME: This is based on the experience also within the COL 1492 trial, and in CONRAD's trial, we will again recruit women at high risk which can now be sex workers or general population women under the high risk criteria. And there is strong evidence in real life that these are very difficult populations to really keep in your trial all the time, for up to 98 percent. Those women are mobile; they often lack the motivation to stay in the trial. There are multiple reasons why, at one moment or another, they decide they may want to leave the trial.

So we try to have our sample size calculations based on real life experience.

DR. DE GRUTTOLA: I have a question there. If you expect your event rate to be considerably less than your loss-to-follow-up rate, do you have concerns about bias--Dr. Van Damme--or Dr. Karim--whoever would like to respond.

DR. GULICK: Victor, do you want to repeat the question?

DR. DE GRUTTOLA: Yes. I just wondered if the loss-to-follow-up rate is expected to be about 20 percent, but the event rate considerably less than that, I would think there might be a concern about bias as well as loss of power, since even a modest amount of differential loss-to-follow-up could impact on the study and impact on its validity.

So I just wondered if Dr. Van Damme or Karim or anyone else had any comment on this issue of bias and validity in the face of a loss-to-follow-up rate that may be higher than the event rate.

DR. NUNN: I'd like to make a comment which actually is picking up one of the points in my presentation, that we are concerned that that could well be the case.

In most populations in Africa, even in rural populations, not just in urban populations or populations with sex workers, there is migration, there is mobility. I was involved in a cohort study which has now being going on for 13 years in Uganda in which we saw 7 percent of the population

actually moving out of their address each year, some coming back again as time went on. And with this in mind, this is one of the reasons in fact that we are considering within the UK Microbicide Development Program looking at a shorter duration to overcome this problem--in other words, as short as possibly 6 months--because we believe that then we could actually considerably reduce the loss-to-follow-up rate and the biases associated with it and get a much closer estimate of true efficacy as distinct from perhaps effectiveness. We would be nearer efficacy than effectiveness. And we are actually considering that right now.

The other possibility is actually a site such as one of our sites which is a sugar plantation where people are much, much more constrained and not moving around. But in many other populations, we are already finding there is a great deal of mobility in populations.

DR. KARIM: I'll just make two points. I don't need to tell this group that it is really difficult to maintain follow-up in healthy subjects. It is a very different scenario from doing long-term follow-up on ill patients.

So in prevention trials, generally, it is

difficult for us to maintain very high levels of follow-up.

I will say that the big concern would be--and this is my second point--if the follow-up were differential in the arms, and if there might be some relationship between the outcome and the follow-up. I think in the one instance that we are dealing with, which is HIV seroconversion, fortunately or unfortunately, it is a silent condition, so it is unlikely to be the event that precipitates the loss-to-follow-up, I would hope. But it is a concern and it is a very deep concern in all the prevention trials that we are doing, and I share it with you.

DR. GULICK: Okay. We are going to need to begin to wrap up our question-and-answer period.

Dr. Fleming has one really important follow-up comment.

DR. FLEMING: And I think Dr. De Gruttola has just hit on a very key point, and just to reiterate what he was referring to--how problematic is it in settings where the number that are lost exceed the number that have events. And I would just like to reiterate to be careful not to assume that if you follow people longer, you are in

a worse situation.

Just to briefly use the actual data from 015 as an illustration, in the first 6 months, for every 100 people, we had 8 lost and one event. In the period from 6 months to 3-1/2 years in that same cohort of 100 people, we lost about 4 additional people and 4 additional events. We did much better by following over a long term to be able to be accumulating number of events versus number lost to follow-up.

So be very careful not to assume that just because longer-term follow-up means more people will be lost, you are actually going to be inducing more bias. That may not be the case.

DR. GULICK: Dr. Bhore, a follow-up?

DR. BHORE: Yes. I want to reiterate the same point as Dr. Fleming, which is that it is quite likely that most of the lost-to-follow-ups will happen early on, and those who stay long enough will likely stay longer. And there has been data in many clinical trials of longer-term follow-up in other disease areas.

Secondly, if you adjust the rate of lost-to-follow-up by time for shorter-term trials versus longer-term trials, the adjusted rate may

not necessarily be higher in the longer-term trials than in the shorter-term.

DR. GULICK: Dr. Brown, waiting patiently.

DR. BROWN: I think the discussions this morning have raised a lot of ethical questions, and I'll try to limit myself to one or two.

Obtaining informed consent has always been difficult for me. I have worked in populations where a chief of a tribe gave informed consent for the tribe. I think we are nowhere near that extreme in these studies, but I would like to ask the first two speakers how they are able to avoid investigator bias in the presentation of the study to the patient in the hopes of getting informed consent.

By the very nature of their work, these women have a person who has control over them because they are going to buy a service from them, telling them to do one thing; an investigator who, at least at a superficial level, is telling them the opposite thing--that is, to wear a condom--and yet down deep the investigator knows the more condoms that are worn, the harder the study will be, and it might wind up destroying the study if enough people do what they are supposed to do.

I am just wondering how you handle those issues, and do you really believe you get informed consent?

DR. GULICK: Dr. Karim or Dr. Van Damme?

DR. VAN DAMME: It is a very good point. Do we get really informed consent--I think we really do try to explain to the women as much as we can and is feasible and achievable what the study is about. One of the two that I used in COL 1492 trying to get an idea about whether or not they really understand is when I was in the centers, I would do random sampling of the women who were there and just ask them, "Can you explain to me what this is all about?"

But as you pointed out, there are different things. I think the staff working on the trials are trained enough not to bias and encourage not using condoms. But there are things which are very difficult to believe, like a doctor or a clinical staff who tells you that, yes, this is a trial going on, and there are definitely positive side effects for the women in the trial. So they assume that indeed it is good, and those women also hope it is good. And by being in the trial and having regular controls and STI treatment, indeed

they do feel better, and they may contribute to the gel.

So I think it is always kind of double-edged, where you trade off and try to do the best you can by explaining over and over. As I said, we also introduce some questions on the basic designed at the end of the informed consent session, which we repeat throughout the trial to be sure that women stay on track and try to have them forget as little as possible that this is a trial, and we do not know the effect; it may have no effect or a negative effect. We do the best we can, I think.

DR. KARIM: I'll just make two quick points, and I can refer you to a paper that we published in the American Journal of Public Health looking at this issue. In that study, we took women who were participating in a perinatal trial, and we assessed the voluntariness of their consent as well as the informed-ness of their consent.

What we found was that the women were very highly informed and were making the decision based on information. But what we found was that they were in a subtle way feeling coerced to participate because they felt that if they didn't participate

in the study, the quality of the antenatal care that they would get at this hospital would not be as good, that they would have to join the rest of the queue.

So there are subtle pressures, there are push and pull factors in the sort of setting that we are talking about. And it is true that the patients who are participating in our studies get a better standard of care. That is one of the incentives.

However, I think it is less of an issue in prevention trials, in a setting where the patients are not beholden on the health care service and the research is not linked to the health care service. So in prevention trials, the issues are slightly different. There, some of these pressures remain, but they are not as acute. And generally, from our experience in the COL trial and in several other studies, we have done quick assessments of the informed-ness of the patient, and what we find generally is that if you take the time and trouble, they do understand what is going on.

And lastly, I want to point out that no matter what I think about condoms undermining the studies, the people that we have, the community

educators that we hire and the nurses who are actually involved with the patients really care, they care deeply about these patients and these subjects, and they would go out on a limb to do what they can for these subjects.

These are not drug trials. These people are participating in these trials as people who are working from the community because they genuinely feel that they want to do something about this epidemic.

So I think it is less of a concern if I was doing the counseling. I am very confident when the community educators are doing it.

DR. GULICK: Thank you.

I have a few quick questions myself. Dr. Van Damme or Dr. Karim, when a woman is randomized to receive the microbicide, how much of a supply does she receive at each study visit?

DR. VAN DAMME: That depends on her own needs, so she would tell us how much she needed, and she could get as many as she wanted. The boxes contain 30, one for each day. Some sites put a limit, say, you can only get three boxes, and then you have to come back to the clinic, to avoid sharing of the product being on the market. That

was driven by the center itself.

But in principle, women could get what they thought they needed during that month, and some of the women are very active.

DR. GULICK: So essentially no limit.

DR. VAN DAMME: Essentially no limit.

DR. GULICK: Okay.

Dr. Wu, you mentioned "universal placebo." Could you say a little more about that? Is that something that is being driven by regulatory guidelines?

DR. WU: No. This is an idea which came from sponsors. The so-called universal placebo means it is the same placebo. It is unrelated to any of the known topical microbicides they wish to test. One company is willing to supply this to other companies, and therefore the data can be shared with other sponsors. This is the so-called universal placebo.

DR. GULICK: So this was developed by industry and is now being shared among--

DR. WU: At least so far, we know it is being used by at least the two sponsors.

DR. GULICK: And does the universal placebo need to fulfill some regulatory

requirements itself?

DR. WU: Yes. The highest burden is on the first sponsor who is going to test. First of all, they have to undergo a limited amount of a non-clinical study and also a Phase 1 study to make sure it is safe before they can be applied to humans. So there is some requirement for that.

DR. GULICK: Okay. My last question is for Dr. Bhore. If I understood correctly in thinking about the three-arm design, one of the goals is to show an incremental benefit of the microbicide above condom use, above baseline condom use.

DR. BHORE: It is not the baseline. Each arm is receiving condoms, and two of the arms are getting let's say the gel if it is a gel, and the third arm is not getting any such gel. So the third arm is getting the condom only. The goal at the end of the trial is to show that the infection rate in the microbicide-plus-condom arm is lower than that in the condom-alone arm, and the rates are lower than that in the placebo-plus-condom arm. So it is not what happens at baseline, at the end of the trial, whatever is planned.

DR. GULICK: And that's my point. So I

understand the design, but your assumption is that condom use remains the same in all three groups during the study.

DR. BHORE: Yes. That's why we would need to see the behavioral data. It is going to be a complex issue to analyze.

DR. GULICK: So this is something that we'll take up more in the afternoon, I believe.

Okay. We are really to the end of the hour, so are there any really burning important questions that must be asked right now?

DR. BHORE: I had a comment on the condom use raised by Dr. Brown.

DR. GULICK: Okay.

DR. BHORE: It is possible that the investigators and the study personnel could influence the counseling in terms of condom use. So for example, two of the arms would be blinded, and one is open-label, and if the study personnel were to influence the use of condoms by differential counseling in the blinded arm versus the open-label arm, this could create problems in interpreting the data.

However, if we had three arms, we would feel at least somewhat comfortable that the two

blinded arms would have the same kind of condom use patterns because they are blinded, and the investigators and study personnel hopefully cannot distinguish between a microbicide product and the placebo product.

Therefore, blinding is a very useful thing to do in clinical trials because it minimizes that kind of bias introduced by study personnel.

DR. GULICK: Dr. Wood, we will have one last question from you.

DR. WOOD: Since condom use clearly can change and is highly variable among populations geographically, the question I have goes to the studies that have already been done, and that is the incidence of STIs as a surrogate marker for condom use in clinical trials. We have heard about pregnancies, but has there been anything where people analyzed the incidence of STIs among arms as a surrogate marker for condom use?

DR. VAN DAMME: The secondary objective of the trial was to [inaudible] gonorrhea, chlamydia [inaudible], and we saw no effect.

DR. GULICK: So you saw no differences in the two arms.

DR. VAN DAMME: No differences between the

two arms.

DR. GULICK: Okay. That was very informative. Thanks to everybody.

It's 12:15. We'll reconvene at 1:05. Let me just say that we have a number of people signed up for the open public hearing, and we need to organize this in a way that we can get through as much as we can in an hour. So would people who signed up to speak please come back 10 minutes early and meet with Tara Turner to go over the podium and the speakers?

Thanks.

[Whereupon, at 12:15 p.m., the proceedings were recessed, to reconvene at 1:12 p.m. this same day.]

A F T E R N O O N S E S S I O N

[1:12 p.m.]

DR. GULICK: Welcome back from lunch.

We had one clarification that Dr. Van Damme wanted to make.

DR. VAN DAMME: Yes. I would like to clarify something about the retention rate, and I'm sorry I didn't grasp that correctly before the lunch break.

We do not plan to lose 20 percent of the women; we plan to have 80 percent retention after one year, so 80 percent of the women completing one year. The other 20 percent can leave the trial, but we will have endpoint definitions, but they can decide to leave the trial because they move, because they become pregnant, or whatever. So it's not that they are really lost to follow-up without any endpoint definition.

So we will have a majority of those women endpoints.

DR. GULICK: Thanks for that clarification.

We'll go now into the open public hearing part of the meeting, and we have a number of people who have signed up to speak. I will call people in

order, and it would probably be most convenient for you to use the podium--and we are going to be a little bit strict about time today.

Our first speaker is Dr. Richard Bax, from Biosyn, Incorporated.

Open Public Hearing

DR. BAX: Thank you.

I am Richard Bax, Chief Scientific Officer at Biosyn. Previously, I have been involved in the development of lots of antibiotics, such as kefluoroxin [phonetic], kefataxin [phonetic], marupenam [phonetic]. And I led the development of the eight indications and three formulations of famcyclovir [phonetic] and pencyclovir [phonetic] for Smith Kline Beecham, the new formulations of augmentin, and bactriban. I have been at Biosyn for 3-1/2 years.

[SLIDE]

Biosyn is the leading microbicide company. We have three compounds--one in Phase III, C31G, which is shortly to enter a Phase III in Ghana and Nigeria under FHI; also, under NICHD in the U.S. for a contraceptive gel claim. We also have just started under CONRAD a Phase I study of UC781, which is an NNRTI inhibitor for use as a

microbicide which has great promise. And we also have from the NCI a protein called synavarian [phonetic] which blocks GP120 in the preclinical situation.

[SLIDE]

What I am going to be talking about in the next 6 minutes is what Biosyn and others such as FHI--and they will talk for themselves--want to do. We want a Phase III trial design which prevents introduction of unknown biases because of the unblindedness.

We are using the HEC common or universal placebo in our studies both in C31G and later with UC781, which will provide a very useful frame of reference for other studies, and the HEC placebo that we are using promises to have the least effect of any placebo.

We believe that the 12-month maximum duration maximizes compliance and good clinical practice and reduces participant fatigue, and also will reduce significant changes in risk behavior of those at 24 months compared to 12 months.

[SLIDE]

We want to compare our active product, C31G, to a pretty inactive placebo to do a simple,

statistically correct study. We do not believe that the addition of a condom-only arm will actually provide the kinds of controls that are required--in fact, it will likely introduce bias.

[SLIDE]

Here are the choices for a three-arm study for no treatment, for placebo gel, active gel with condom controls. And as you can see, each of the three groups has different choices. Different choices lead to different behaviors. And we have no idea because of the uncertainty of compliance and of the sexual practice log whether or not those biases have been introduced post hoc of the randomization, and we will never know.

It seems to me that a statistician is a cynic in a world of uncertainty, and the addition of the condom-only arm will increase that uncertainty.

[SLIDE]

We want to produce the best, most effective, most credible clinical trial which will assess the effectiveness of this product against placebo. There are certain credibility machineries within clinical trials which include ethical statistical practices, which we will adhere to;

comprehensive protocol development and review with experts and the FDA and interim analysis; and the application of the baseline difference avoidance tools, and also, most importantly, replicate studies.

It appears to me that there are many more important issues for microbicide trials than we are discussing today. They include, clearly, study selection, site selection, how the study is conducted and, most of all, compliance.

I think the most important factor is that what will happen is that it will be easy to actually show that effective microbicides are not effective, rather than that not effective microbicides are effective, and that point is certainly endorsed by Dr. Andrew Nunn.

[SLIDE]

So I believe that the progress to date of the microbicide community into Phase III, which is the only possible way a microbicide will become available, has been at best regrettable and at worst appalling. I believe that now is the time to do a statistically correct, simple study which has a chance of showing an effective agent is effective rather than talking about a third arm with lots of

uncertainties, raising the hurdle unnecessarily and also talking about significantly long trials, which also are undoubtedly going to introduce biases.

The last point I would like to make--and it is an important point--is that there is a constant in medicine, and that is that the greater the likelihood of an adverse event like death due to HIV, the greater the benefit of the treatment or the medicine.

In the United States, I believe there are approximately 20,000 HIV transmissions a year estimated due to heterosexual sex. In the developing world, there are 16,000 per day. I believe that the risk-benefit of such a product is very important and very different in the developing world, but we should apply the right science, the right statistics, the right trial, and do it now.

Thank you.

[Applause.]

DR. GULICK: Thanks, Dr. Bax, and thanks for sticking to the time as well.

Our next speaker is Dr. Polly Harrison, Director of the Alliance for Microbicide Development.

DR. HARRISON: Thank you.

I want to preface what I am going to say with two observations. One, the origin of this presentation--it comes out of a n interactive process that has been going on over the last few months as these issues have come to a peak, shall we say, and this paper and the conclusions I am going to present represent the consensus among 17 participants from nine different entities. So it is a consensus document, and I want you to understand it as such.

Because time is limited, and a number of things have already been said, I will not focus on those; I will just proceed through the slides and pick out the high points or the points that have not been addressed.

[SLIDE]

There are some contextual issues that have not arisen in the course of the conversation today. One is that when we talk about HIV/AIDS, we are talking about one of a family of emerging and neglected diseases that are effectively orphaned by the pharmaceutical industry because the bottom line is not perceived as sufficiently rewarding.

This creates a set of issues for all of us that have commanded the interest of the world

community, so there is now a process that the European Medicines Authority and the WHO have engaged in, which is to examine how we can adjust for the different risk-benefit ratios we are seeing globally with the kinds of regulatory processes that we all engage in.

We urge--our recommendation is, if you will see the action item--CDER--the Center for Biologics is already involved in this activity--we would recommend or hope that CDER would become engaged as well.

[SLIDE]

The control arm--there has been a lot of conversation about that, and I'm not going to go into the pros and cons of the no-treatment arm. I'll just go to the bottom line.

It was the conclusion of the group that the contextual realities--and in the interest of full disclosure, I must identify myself as a medical anthropologist, so I am concerned with the behavioral realities, as I think many of us are--we believe that the contextual realities around the fields that we are trying to discover trump what would be nice to know. The closure that we have come to is that if the 035 trial goes ahead with a

no-treatment arm, that would be salubrious, perhaps, for the field in terms of satisfying a number of questions--in fact, whether indeed that is an interpretable addition to a trial design--but that the other trials that are approximately concurrent would go on in the same time frame. In other words, they will not be blocked by this enduring question.

[SLIDE]

Now, the duration issue. Again, I won't deal with the strengths; they have been discussed already today, and I won't repeat them. But I do want to point to one thing that I think has not been mentioned. One argument for a longer period of on-treatment evaluation and post-treatment follow-up is if the seroconversion rates are uneven over time.

The evidence that we have--and admittedly, it's not a lot--is that they are not uneven over time, and so that in effect disqualifies this criterion, perhaps, as an argument for a longer follow-up period.

[SLIDE]

I again won't deal with the limitations. The bottom line for us was that quality trumps

quantity for quantity's sake. In other words, we believe that the quality of the data that can be derived from a shorter period of follow-up will be superior to the actual number of datapoints gathered over a longer period.

The recommendation of the group was that there should be a maximum of 12 months on-treatment evaluation per participant.

[SLIDE]

Strength of evidence--I am not going to talk about p-values.

The bottom line here--and I think maybe we have sensed it in the course of the morning--is that in a way, we are in a data-free zone when it comes to how we put all the ingredients of the ultimate strength of a trial, the ultimate power, together, the action item that we perceive as desirable here is that you trade off the arm, the condom-only arm, the no-treatment arm, for a--"relaxed" is wrong there; it should be "a more stringent" p-value--in other words, you can ask more of your p-value of two arms, and you can perhaps add more subjects per control and placebo arms.

[SLIDE]

The final thing is the definition of a "win". Again, we have a double-standard, if you will, for 035 and other trials.

We urge that the criteria for defining a "win" with respect to 035 be that beating one control arm would be adequate. We have three. If you beat one control arm, that's adequate if the other goes in the right direction--and Dr. Fleming alluded to that earlier this morning.

With the other trials that are ongoing, we ask for flexibility with respect to dropping the no-treatment arm, and in that case, we would expect that the one arm would have to be beaten well.

[SLIDE]

Adherence--that is not something that the FDA has to do, but it is something to which the FDA is entitled in terms of quality of data. It is critical for interpreting results, for formulating claims, for labeling, for registration. It matters very much. And we don't have any true measure of adherence, so it is the job of the field to do better with the approaches that we have, to replace them with more rewarding techniques, and finally, to learn from others. And I would submit to you that we do have some learning on which to build.

The experience with the female condom is such that we can learn, and one of the most important lessons that perhaps we can learn is that if we engage the community and integrate it into the process of the trial, our chances of getting good data will be much enhanced.

Thank you very much.

[Applause.]

DR. GULICK: Thank you, Dr. Harrison.

Our next speaker is Dr. Ian McGowan from the David Geffen School of Medicine at UCLA.

DR. McGOWAN: Mr. Chairman, ladies and gentlemen, I'd like to begin by thanking the FDA for giving me the opportunity to briefly discuss the subject of rectal microbicide development during this session.

I would also like to acknowledge support from Ken Mayer [phonetic], Peter Anton [phonetic], and Michael Gross in preparing this very brief talk.

Oscar Wilde described a type of "love that dare not speak its name," and based on the proceedings so far today, I think we could add anal intercourse, rectal mucosal vulnerability to HIV, and rectal microbicide development as possible

other types of behavior that dare not speak its name.

However, the primary focus of this meeting is a discussion of the methodological challenges in designing vaginal microbicide efficacy studies, so perhaps to some, the topic of rectal microbicide development may seem irrelevant or at least a distraction.

I hope that in the remaining 6 minutes and 4 seconds, I can persuade the Committee and the audience that we really need to keep this issue of rectal microbicide development as an important component indeed of vaginal microbicide development as well as on its own basis of rectal microbicide development.

[SLIDE]

I would like to address three questions. First of all, why develop rectal microbicides; secondly, what are some of the challenges; and finally, what is the current status of rectal microbicide development?

[SLIDE]

Why develop them? I think it is self-evident to many in the audience that anal intercourse remains the primary risk factor for HIV

transmission amongst MSM. What is perhaps less appreciated and poorly-defined epidemiologically is that the prevalence of anal intercourse amongst the heterosexual population is underappreciated and indeed represents a significant risk for HIV transmission.

Much anal intercourse, particularly in the heterosexual population, is unprotected. The mucosa is incredibly vulnerable to transmission, and based on N-9 experience, vaginal products may just not be suitable for rectal administration.

[SLIDE]

These are some data, not comprehensive but I think illustrative, looking at prevalence of anal intercourse. The baseline data from the HPTN EXPLORE study demonstrated, perhaps not surprisingly, that approximately 50 percent of men who have sex with men practice anal intercourse.

Again, perhaps surprisingly, Michael Gross was able to define in his study of high-risk women a prevalence rate of 32 percent; in heterosexual college students, 20 percent; and in a California-based adult survey, 6 to 8 percent. I would argue that in the interpretation of microbicide studies, vaginal microbicide studies,

we will need to be cognizant of this fact.

[SLIDE]

What, then, are the challenges?

Well, I think the first challenge is just to create awareness that there is a need for this type of development and an awareness of this type of confounding variable in the interpretation of vaginal microbicide studies.

I don't think we're very clear yet about strategy. Are we going to have vaginal products, rectal, or combination products? And a very thorny issue is how do we begin the safety evaluation of this type of microbicide.

[SLIDE]

We know from previous speakers today that the pipeline is quite rich, particularly in the discovery and preclinical phase, less so in the more advanced phases. But I think when we look at this potential pipeline of rectal products, albeit labeled as vaginal at the moment, I think we need to think about how we are going to screen this pipeline for candidates to move into Phase 1, how we are going to actually design these Phase 1 studies, and perhaps more pertinent to today's meeting, are Phase 1 rectal studies needed perhaps

to support a vaginal microbicide indication.

[SLIDE]

Another issue which my group at UCLA is particularly interested in is are the conventional safety paradigms for looking at compounds in Phase 1 sufficient for rectal microbicides. We have all had lunch, so I hope you will bear with me--this is the appearance when we undertake a flexible sigmoidoscopy. Can we bring the lights down a bit, because I am going to show a histology slide.

This actually is a very normal-looking endoscopic appearance. And if I actually show you a histology slide from the same patient, that indeed is also very healthy-appearing. The fact of the matter is this patient actually has HIV infection. And when I undertake quantitative immunohistochemistry for CCR5, thus profound regulation, it is even greater than seen in inflammatory bowel disease and definitely more so than seen in control patients.

My point is not to talk about pathogenesis but to illustrate that you cannot just rely on macroscopic and perhaps histological appearances in this type of study. The more interesting question is what to replace or what to add to these

conventional ways of defining safety. I don't have an answer yet, but hopefully some of the studies that individuals, ourselves included, are undertaking might begin to address this issue.

[SLIDE]

What is the current status of rectal microbicide development? This is perhaps the briefest side in the presentation. I think the community now know that N-9 is not suitable for microbicide. Carraguard in a very small study appeared not to induce epithelial damage. But there are no Phase 1 microbicide studies planned at this point in time.

A recent development in the last month was the observation by Tsai [phonetic] and his colleagues at University of Washington that sinavirin [phonetic] was able to block rectal transmission of a SHIV [phonetic] 89.6 variant virus. That is very encouraging but I think suggests that we should be doing more to move this type of product into Phase 1 studies.

[SLIDE]

To summarize, I think there is an urgent need to develop rectal microbicides for the MSM population as well as to acknowledge that the

heterosexual population is at risk of transmission from anal intercourse, and that this is an underappreciated behavioral variable, particularly in Phase 2/3 studies of vaginal microbicides.

I would even go further to argue that I think it is very important that these compounds will be used both vaginally and rectally, whether it is labeled or not, and that the FDA should really include or ask for a Phase 1 safety evaluation of rectal toxicity to be included in the NDA filing package.

And finally, we still have a lot of work to do to define an appropriate preclinical and clinical development track for this type of product.

Thank you very much for your attention.

[Applause.]

DR. GULICK: Thank you, Dr. McGowan.

Our next speaker is Dr. Don Waldron, from the Population Council at Rockefeller University.

DR. WALDRON: Thank you, Mr. Chairman.

It is a pleasure to address you. I am Dr. Don Waldron. I am the Medical Director at the Population Council at Rockefeller University, and I want to share with you some of our experiences and

where we are going in the microbicide research conducted by the Population Council.

[SLIDE]

We started the process early in the eighties and identified a large molecular structure that would actually block HIV. We did in vitro studies in cell cultures, and we found it to be protective against HIV, and followed that up with in vivo mouse and monkey experiments and also demonstrated again blocking.

We knew that we were going to go into clinical trials, so we developed a placebo, methyl cellulose, which we found through in vivo studies was not protective against HIV.

[SLIDE]

We then conducted a series of Phase 1 trials in many countries, particularly in South Africa, which is the country that we are interested in at the current moment for Phase 3. The results showed that Carraguard was safe and acceptable.

We are currently doing a couples study for male tolerance and acceptability, and those results are under analysis, and I don't have anything to share with you on that.

We also have two studies underway in

HIV-positive cohorts, and those results will hopefully shed new light as to what does happen when people have HIV.

[SLIDE]

We then did Phase 2 experiences where we had some preliminary observations, and those data are still under analysis. There were two trials conducted, one in Thailand with 165 women, and in South Africa, where we had 400 women. They were two-arm, they were intent to treat trials, Carraguard against placebo.

They were shown to be safe, and acceptability was again confirmed. We didn't see any difference in adverse events, STIs, between those two arms.

Condom use was similar in both arms, although in Thailand, we noticed that the condom usage was significantly higher from baseline. I don't have those exact figures with me at this time, which we might have brought to bear in earlier conversations that we had.

Recruitment and retention was similar for both arms in both Thailand and in South Africa.

We had no seroconversions in Thailand, whereas in South Africa, we had an equal number of

seroconversions, eight in each arm. This was a 12-month trial.

[SLIDE]

I just want to share with the question of condom usage that we wanted to look at exactly at the end of the trial what was our overall usage for the gel-plus-condoms, and we see that it is relatively the same whether we were using placebo or whether we were using Carraguard, and that very few of the patients were using nothing, and condoms-only was equivalently the same as using nothing. So roughly 8 to 10 percent of the people were using just condoms only, and again, 8 to 10 percent were using nothing to protect themselves.

So that somewhere on the order of 60 percent of the people were using some form of protection whether it be gel with condoms or it was the actual gel only.

[SLIDE]

Now we are at the stage of doing a Phase 3 design, and we have several considerations that we are putting in place, and we are discussing those amongst ourselves and with other outside agencies.

It is going to be a classic placebo-controlled, two-arm, doubleblinded ITT

trial in roughly 4,500 noninfected women in South Africa. The active arm will be Carraguard with a methyl cellulose placebo. The maximum trial duration is 48 months with no patient being in any longer than 24 months. We are examining a design where we will have closing of the trial 12 months after the last patient's first visit, regardless of where we are into trial.

[SLIDE]

The trial criteria--these are very glossy--are that basically, we will exclude women who test positive for HIV--that is obvious--and pregnant women. Women who have STIs, unlike in the Phase 2 trial where they were not accepted, will be accepted in this trial. Primary endpoints will be HIV seroconversion, and the safety endpoints will be STIs and vaginal lesions.

[SLIDE]

Compliance is a big issue, and we have heard it throughout this meeting. Compliance is going to be tested using several methods. There will be visit questionnaires administered by clinical staff; applicator tracking using bar codes; compliance with visit schedule, which I haven't heard mentioned, but that's an important

compliance issue for us; and applicator usage tests are currently under evaluation in New York, and we are hoping to look at those further.

We are looking at using some of those criteria and whether or not we can more clearly define the ITT analysis and exclusion criteria and patient removal from the trial itself.

That's all I wished to share with you at this point.

[Applause.]

DR. GULICK: Thank you.

The next speaker is Dr. Tim Farley from the World Health Organization.

DR. FARLEY: Thank you, Mr. Chairman, and thank you to the FD for giving me the opportunity to address you.

I may say that I am the person responsible in WHO for the microbicide work, and we took over responsibility for the COL 1492 trial, seeing that to its conclusion when it was transferred from UNAIDS, so my experience in this field is to an extent influenced very much by the COL 1492 trial, as is all of ours.

[SLIDE]

I was going to talk about three key

things. The first one, which is measures of product effect--efficacy, effectiveness, and use effectiveness--I am going to skip, because the other issues I want to talk more on--however, if you want to ask me some questions about it afterward, then it won't count into my 7 minutes.

[SLIDE]

Moving straight to the issue which has been discussed quite considerably today, which refers to the issue of choice of control arm or control arms. Some of these points have been made before, but I think they are worth emphasizing.

The randomization ensures balance of factors which are related to individual risk and patterns of condom and product use. However, once the study group has been revealed to the participant, the randomization will no longer be able to balance changes in behavior which are induced by knowing which group the person is in. In order to be able to maintain the post-randomization balance, we need good masking and good blinding, and that is why we use a placebo-controlled doubleblind trial.

This is the gold standard of all evaluations whenever we can, and it is the

preferred design whenever it is feasible. The beauty of it is that the inferences from the study, particularly if you do an intention-to-treat analysis, are very compelling, and it also gives an unbiased estimate of the product effectiveness. This, for example, was seen in the COL trial.

It doesn't mean to say that you should not collect data on behavioral factors, compliance, and so on, but it must be recalled that those additional data are really there for exploratory or explanatory analysis, looking at internal consistency and so on. But the headline analysis of overall effect does not depend on those behavioral data.

[SLIDE]

If you have a no-product arm, it is absolutely essential when there is no placebo product available. That's absolutely clear. It is a no-brainer. If there is no placebo product, or it is not possible to make one that is going to preserve the blind, then you need to use a no-product arm.

The problem here is that you must collect very high-quality and extensive and reliable data on product and condom use, because you have to make

adjustments for this, and your primary analysis, your primary inference, must be based on these data where you are adjusting for rectal intercourse, you are adjusting for different patterns of condom use, condom non-use, and so on.

However hard we try, there will always be doubt as to the validity of these data. And I would suggest that in any trial, you are going to get some misclassification. You are going to get reported behaviors, but there is going to be a misclassification.

What is the effect of this misclassification? Well, effectively, you are going to dilute your estimated treatment effect.

So I can see a situation where we have a product where it has a certain effectiveness, you have a placebo which is totally inert, and you have a no-condom arm, but because of the effect dilution because of the misclassification, you may find that your product is significantly better than the placebo but is not significantly better than the no-product arm, simply because you need to do this adjustment. I think that the inferences from this are going to be very difficult, and it is going to be difficult to have these two inferences, as I

said, within the same study.

So if you have two control groups, fine. It is very, very costly; it adds cost to the trial, and I think we need to consider the costs of these trials. These trials don't come cheap, and at the moment, the majority of studies are mainly being funded by public sectors, and the funds are not unlimited.

I believe that you get no benefit for interpretation by adding a no-product arm when you have a placebo. I think it is potentially confusing. And I would like to cite the example of COL 1492. Had there been a no-product arm in that study, I don't believe that it would have helped any of the inferences which came out of the COL study, the headline being that N-9 had a higher incidence of HIV infection than the placebo. It may have helped to say something about the placebo, but it wouldn't have changed the overall inference about the study.

Now, the other issue I want to address is the issue of strength of evidence, which has come up a number of times today. Actually, I'd like to say just one thing back on the two control groups. I think it is an issue that sponsors might like to

consider. If somebody wants to do an active versus placebo versus a no-product arm, they should be allowed to do it. I wouldn't advise against it. I certainly don't think that the FDA should require it because it is going to have costs, it is going to cause a great deal of difficulties for other studies as well. So I think that the FDA may allow it, but to require it I think would be an extremely bad thing to do.

On the issue of strength of evidence, the discussion that we had this morning about how two independent studies at .05 is desirable, is the FDA's usual standard; however, there are difficulties with this, and of course, there are questions as to whether an ethical review committee is likely to approve going to a second trial once the first one has been done.

The statistics in going from two studies at P less than .05 to a single study at .0013 are impeccable. The problem is that the ethics are appalling.

If it is unethical after a first trial which is convincing at .05 to do a second study, then it is equally unethical to do a study of the size of .0013. Halfway through that trial, the

data which are available would be convincing as that first study.

So I submit to you that it is equally unethical to do a study requiring that level, that small P value.

I also think that ethical review committees--certainly mine in WHO--would not approve it. They would not allow us to do a trial where we are requiring significance at the .0013 level. And I suspect that the ethical review committees in the sites where such a trial would be done would also reject that.

I think we need to consider what are we protecting ourselves against here. Remember, this is the probability of a false-positive. This is falsely declaring a product which is not effective as effective. Normally, conventionally, we limit that at one in 20, possibly a bit less, but to limit that as to one in 1,000 I think is off-the-wall, quite frankly.

I am much more concerned about the false-negative here of not showing an effective product actually has an effect--not falsely showing at one in 1,000 that a product which is not effective actually is effective. And there is a

balance between power and size, and I would rather put it on power than on protecting against the false-positive.

Now, I fully agree that a single study at P less than .05 may not convince, and the COL 1492 trial came in just significant at P less than .05. Not everybody was convinced that N-9 was harmful by that, so I take the point that one study at P less than .05 is maybe not there.

What would I suggest? I don't know exactly what would be an appropriate P value to have. I certainly think that one in 1,000 is way off-the-mark. I also think that maybe one in 100, less than .01, is probably off-the-mark.

What I think you need to do is to discuss with ethicists, with regulators, with public health experts, with advocates, in a range of countries, particularly countries where the HIV epidemic is really raging and there is a need for this, and ask them the question very simply: Look, let's assume we had a trial that was significant at the .05 level, and it is internally consistent and so on. Would you think it is ethical to do a trial?

If they say yes, you ask the question again: What about at P less than .04--would it be

ethical--yes or no?

There is going to come a time--P less than .01, maybe P less than .02--when everybody says no, it is no longer ethical. So I suggest that you convene a consultation of that nature--I will convene it for you if you want--and then we can get an idea of where people feel very uncomfortable from an ethical point of view to do the second trial. And that is what I think you should aim at for your P value for a trial.

Thank you very much.

[Applause.]

DR. GULICK: Thank you, Dr. Waldron.

Our next speaker is Amy Allina, from the National Women's Health Network.

MS. ALLINA: Thank you.

My name is Amy Allina. I am from the National Women's Health Network which is a nonprofit organization that advocates for national policies that protect and promote all women's health. We don't accept financial support from pharmaceutical or medical device companies, and we are supported by a national membership of 8,000 individuals and about 300 organizations.

I want to start by thanking the FDA for

organizing and holding this meeting and for giving us the opportunity to speak about the importance of this topic to women.

The National Women's Health Network began working on HIV/AIDS as a women's health concern in 1987. Even before the advent of AIDS, the Network had articulated the need for sexually-transmitted disease prevention options for women, testifying before Congress as early as 1978 on the importance of research to develop these products. So we have been at this a long time.

In the 25 years that we have been working on these issues, particularly in the last 15 years with AIDS, the need for attention to women's prevention options has become increasingly urgent.

In a survey conducted just last year, our members identified microbicide development as a top priority on the Federal health research agenda. The Network is a participant in the Alliance for Microbicide Development and a partner in the Global Campaign for Microbicides, and we endorse the recommendations that you heard earlier from Polly Harrison, from the Alliance, and also that the panel at least received prior to the meeting from the Global Campaign.

Given the tight agenda today, I am not going to repeat those recommendations. You have all heard them and read them, I am sure. But I do want to address one in particular which is the recommendation that FDA shouldn't require as a matter of policy that sponsors include a condom-only arm in addition to the placebo control. There has been a lot of discussion about that already, and I'm going to try not to repeat too much of it, but there are a couple of things that I want to say about why we agree with that recommendation.

FDA staff certainly and possibly also some members of the Committee have heard the Network advocate in other settings for the agency to require new products seeking approval to be tested against existing products rather than just against a placebo. And in light of that, our endorsement of the recommendation that FDA should not require sponsors of candidate microbicides to compare their products to condoms alone in addition to a placebo control might seem contradictory. So I want to be clear about the differences that lead us to support the recommendation.

Our argument that some new products should

be tested before approval in trials which compare them to existing products has been based on our belief that FDA should demand more information and apply a stricter approval standard when there are already products approved and available for the same indication, when we are talking about the so-called "me-too" products. In that circumstance, consumers and health care providers who are considering using or prescribing the new product will need to know not just that it is safe and effective but whether it provides added benefit over existing and often less expensive options that are already available to them. But that argument is obviously not relevant in the current context of microbicide development.

There is no existing product to which a microbicide can appropriately and usefully be compared, and although condoms are an effective and important option for many individuals and couples, we all know that some women are not able to negotiate condom use with every encounter and with every partner.

We also share many of the concerns that have been articulated already today that the requirement that all microbicide clinical trials

include a condom-only arm may be an obstacle in some cases to producing interpretable data.

We agree with earlier speakers who have said that inclusion of a condom-only arm might provide useful information in some cases; in other situations, however, we believe it would further complicate interpretation of trial results.

So for those reasons and because of our concern that the requirement of all trials include two control arms might slow progress of this really urgent research, we urge FDA to maintain flexibility on this point and not to require all sponsors to include a condom-only arm.

I'll finish here and just say that I'd be glad to answer any questions from the panel about my statement.

Thanks.

[Applause.]

DR. GULICK: Thank you very much.

Our next speaker is Dr. Rosalie Dominik from Family Health International.

DR. DOMINIK: Thank you for the opportunity to present on behalf of FHI. FHI's decades of research and experience with contraceptive and microbicidal products has

provided us with valuable lessons regarding the conduct of trials in resource-poor settings.

Our experience with microbicide research in Cameroon encompassed three different study designs--an observational study in 1991 to 1992 of women choosing spermicidal suppositories versus those choosing other methods of contraception; a blinded randomized control trial in 1995 and 1996 of women using N-9 film versus placebo film; and an unblinded RCT in 1999 and 2000 of women using N-9 gel versus a no-gel condom-only control.

Comparisons of the first two trials demonstrated the strength of the randomized design in controlling for the intrinsic selection bias that can occur in observational studies. These studies also demonstrated the difficulties in interpreting self-reported data on sexual behavior. Analysis of the third trial demonstrated the limitation of interpretability of unblinded trials.

[SLIDE]

We believe it is useful to focus on the labeling claims that one hopes to make for an effective microbicide to guide the decisions about study design. We expect that the label for the first approved microbicide might include a summary

message that looks something like this: "Use of microbicide gel reduces a woman's risk of HIV infection during vaginal intercourse. To best protect against the risk of HIV infection during vaginal intercourse, use a condom during every act of intercourse. Use of microbicide gel provides additional or backup protection against HIV infection."

[SLIDE]

To obtain evidence to make such a claim, we need to design a study that can answer the primary research question of whether use of the microbicide reduces the risk of HIV acquisition compared to nonuse, holding all other risk factors constant. That is, the two groups of women being compared should have, for example, the same average frequency of intercourse and the same level of condom use.

A blinded RCT of the microbicide gel versus a truly inactive placebo would be of course the gold standard for answering this question. Unfortunately, we may never be able to definitively demonstrate that we have a truly inactive placebo, but the comparison of the active microbicide to the carefully-selected placebo, the best available

placebo, will provide the most useful data for answering our primary research question.

[SLIDE]

The other control arm that has been discussed is of course the condom-only arm, and we have talked about differences that the two groups will have in motivation, resulting in--also when you have the condom-only arm, you have a group that only has two options to choose from versus a group that has four options to choose from with each act of intercourse.

I mentioned earlier that in the unblinded N-9 trial that FHI carried out in Cameroon, women in the condom-only arm reported using condoms in about 87 percent of acts, while women in the gel arm reported using condoms about 6 percent less often.

[SLIDE]

Now I would like to walk through two examples showing the impact of a 10 percent difference of condom use on comparisons between the microbicide arm and the condom-only arm, assuming that when used, condoms reduced the risk of HIV acquisition by 95 percent.

First assume that we have a microbicide

that would reduce the risk by 50 percent compared to an absolutely inert placebo, and if we designed a study to have 90 percent power to detect this 50 percent reduction in risk of HIV acquisition, what would happen if the microbicide used condoms in 65 percent of acts, and the condom-only arm used condoms in 75 percent of acts.

In this case, the power would drop from 90 percent to about 50 percent.

[SLIDE]

If condom use instead were 80 percent in the microbicide arm and 10 percent higher, 90 percent, in the condom-only arm, the chance of finding a significantly lower risk of HIV acquisition in the microbicide arm would be only about 15 percent. And in this case, there would actually be about a 20 percent chance of observing a higher incidence of HIV in the microbicide arm than the condom-only arm.

[SLIDE]

So this example helps to illustrate why we are concerned that requiring that a microbicide arm be shown to be significantly better and have significantly less HIV infection compared to a condom-only arm could lead to failure to promptly

identify a product that truly protects against HIV.

[SLIDE]

The second example addresses another potential danger that can arise due to behavioral differences between the two arms. In this example, we assume that the microbicide truly has no effect on HIV risk compared to a true placebo, and we look at what can happen if the participants in the microbicide arm use condoms more often than those in the condom-only arm.

So if condom use is 90 percent in the microbicide arm and 80 percent in the condom-only arm, there would actually be a 65 percent chance of observing a significantly lower risk of HIV acquisition in the microbicide arm even though the microbicide is truly ineffective. This 65 percent chance of falsely concluding the microbicide is effective is far greater than the 2.5 percent chance of a Type 1 error in this direction that one would expect if risk-taking behaviors were truly balanced between groups.

[SLIDE]

Even though we don't believe a condom-only arm should be required, we do believe that a comparison between a placebo arm and a condom-only

arm may provide some useful information about the activity of the placebo. If we are willing to assume that the bias due to behavior changes will operate in only one direction--that is, that those in the condom-only group will use condoms at least as much as those in the placebo group--then the inclusion of a condom-only arm may provide some evidence that the best available placebo gel might actually provide some protective effect, but because of the unblinded nature of the trial, it may not be entirely convincing.

The HPTN 035 trial will help to define the role, if any, of a condom-only arm in subsequent microbicide trials, and FHI is supporting the 035 team in conducting this NIH-sponsored trial.

[SLIDE]

So in conclusion, what we most want to know is does use of the microbicide reduce the risk of HIV acquisition. Once we have a product that reduces the risk of HIV when used, public health researchers can turn to studying the best way to promote use of that product in combination with a host of other preventive measures. Showing the protective effect against a carefully-selected placebo should provide reasonable evidence that a

product protects against HIV if used. A blinded two-arm trial of a microbicide versus the best available placebo can provide sufficient evidence to support a claim that use of a new microbicide can reduce the risk of HIV acquisition.

Thank you.

[Applause.]

DR. GULICK: Thank you.

Our next speaker is Dr. Zena Stein from Columbia.

DR. STEIN: Thank you for giving me the opportunity to talk, and as I come at the end of many arguments, I just want to say two things.

One, we are talking about biological efficacy of the microbicides we are testing, and we have some biological information about inert substances, the placebo. And the purpose of the trials, I would say, is to look for human evidence that supports the biological evidence of efficacy, not to go beyond that.

Now, if we have done the classical approach, and then sexual factors lack useful microbicide, we have an enormous area for distortion of reports and diaries and statements.

So the wonderful idea of a blinded

microbicide, putative microbicide, which would feel the same and look the same and smell the same for a women and for the investigator, and to set it up in a little white introducer, it will make the difference between the putative microbicide and the putative inert substances invisible.

It allows you basically to cancel out all those factors in effectiveness and lead you to infer efficacy. You don't care how much adherence or how much frequency of use or any of those things, because it should be the same between the putative microbicide and the putative inert substances.

When you start bringing in a condom, another arm, you are asking another question, and maybe it is an important question that should be asked afterward. But now we ought to know do we have a microbicide which supports the biological difference between efficacy of the microbicide and efficacy in the inert substances.

The reason I entered this dialogue publicly is because my slide, which is basically the same options as Dr. Karim offered us--we tried to put down all the options we could think of, and we decided that A, B, and C in which the placebo

and the condom-only do the same thing, that that would give you confirmation that all the others--D, E, F, G, H, and I--would give you confusion, which is why we said stick to the placebo and the microbicide; otherwise, you'll get confusion.

I didn't like the idea of support where you don't get a difference between the microbicide and the placebo. You haven't supported your biological assumption of efficacy, so don't do it.

At the bottom here, "These interpretations assume a) that true levels of condom use do not vary across trial arms"--and this is a point that Dr. Farley and other people made, and the reason I came here to try to say something new is the point I mentioned earlier, that we have some evidence from the COL 1492 group that in fact women loved the microbicide or the placebo; they used it and they dropped the condom arm. I think they will do that. It is very good news for microbicide, but it will hopelessly contaminate any attempt to measures in this trial what condom-only does because again, it changes the risk behavior. If some of them are [inaudible] random and risk behavior, you put them into the trial, and they change their risk behavior, and you are just left reflecting with

what to do with that kind of mess.

Now, the other point--I am allowed a second point--it is only when you get a placebo, the microbicide versus placebo is only as good as what you know about the placebo. We've got this new universal placebo. If every trial would use a universal placebo, the same one, you could make comparisons across trials. If one trial uses this placebo and another trial uses this placebo, you will not be able to make comparisons across trials.

I would even suggest that, for instance, if Carragin [phonetic] wants its own special methyl sulfate arm, put another arm, put the universal placebo arm. You will learn more from that because the behavior is much the same, and you will be able to compare other trials. That kind of insert of an arm would make sense. But the insert of an arm which is open, which confuses the behavior, confuses the difference between efficacy and effectiveness, I consider a waste of time.

And I agree with everybody here saying that FDA should open its mind to whether it wants this or that behavior. If it wants to actually concentrate on biological efficacy versus effectiveness in one product and another, there is

no point in confusing the issue with a condom-only. That is asking another question and perhaps asking it in different ways, and this might not be the way to measure it.

I am also convinced by what Dr. Dominik said. A paper of Foss [phonetic] et al. which many of you might know, suggests that where condom use is only 15 percent or less in the population, and you have a reasonably effective microbicide, on the whole, you can't do wrong--put your microbicide in. If you get a microbicide that is as much as two or three times the placebo, you can use it happily, because so many populations use so few condoms that you can only win with that.

And remember that on the whole, the difference we get in effectiveness in protection against HIV only seems to work when people really use the condom at 100 or 90 percent in the various estimates we have based on discordant couples. You have really got to use that condom a lot to make a difference in the transmission.

So I think that condoms are in. It satisfies us ethically, but the real question is does the microbicide versus the inert substance make a difference for HIV infection. If it does,

we'll all put our flags up, and we'll have something to go with as soon as possible.

Thanks.

[Applause.]

DR. GULICK: Thank you.

Our final speaker to have signed up is Dr. Malcolm Potts, who is from the University of California at Berkeley.

DR. POTTS: I speak as a physician. I am from Berkeley, and as the former president and CEO of Family Health International, where we initiated the first-ever microbicide trials, I have been a strong advocate of microbicides for over two decades. In 1990, I triggered the UK MRC interest in microbicides.

Like many people, I initially accepted placebo control trials with condom counseling as licit. After a great deal of thought, I have slowly and painfully come to the conclusion that such trials may be flawed scientifically and ethically.

Ethically, I am deeply troubled by a basic contradiction. While the justification for recommending condom counseling is that we offer volunteers the highest possible standard of care,

the pivotal findings from any clinical trial are derived entirely from volunteer women who we know for certain are not using condoms.

I think we have misled ourselves into believe that if we recommend condom use, it is acceptable to use placebos. But the number of women not using condoms unless exposed to HIV infection in a placebo-controlled trial cannot be lower--cannot be lower--than it would be without counseling.

Further, a condom counseling design could actually increase the number of placebo users who will be infected and die, because counseling inflates the number of subjects needed.

Having had executive responsibility for a great many clinical trials, I am vividly aware that the more difficult the logistics, the higher the loss to follow-up, the more volunteers you need to recruit. We are talking about populations that are so different from those described by Dr. Fleming that they might as well live on another planet.

If we use placebos, then condom counseling complicates the study but does not solve the ethical problem for the women who provide the data on efficacy who are randomly allotted to exposure

to a lethal, incurable disease.

Condoms indeed are the best advice for those who use them, but those people dilute the results. I haven't heard a proposition for how to help most groups. I think that is our ethical dilemma.

In contraceptive trials, we do not use placebos presumably because an unintended pregnancy is an unacceptable burden. Can we use placebos when that is the outcome? Some women will not respond to condom counseling because their compliance with any instruction is low. This is exactly the group that we want to exclude from any clinical trial.

More likely, in my judgment, the non-condom users are simply unable to negotiate condom use with their partners. I feel deeply uncomfortable trying to shuffle my ethical responsibilities by relying on underprivileged volunteers to make mistakes.

Scientifically, as a possibility, we may reject an otherwise lifesaving microbicide which might have worked amongst those women who enjoy greater autonomy in their lives but which failed in this nonrepresentative subgroup of volunteers.

The Code of Federal Regulations under which the FDA operates is explicit. The test [inaudible] compared with known effective therapy and the administration of placebo or no treatment would be contrary to the interest of the patient. To ask a woman whose husband will beat her if she asks him to use a condom to accept a placebo is unambiguously contrary to her interest. The offer of a microbicide, even of unproven effectiveness, might be preferable.

The trouble, of course, is that we cannot predict in advance who is able to respond to condom counseling and who will not; and for those who will respond, condom counseling is indeed the highest possible standard. If we don't use placebos, we can't measure efficacy. But I suggest that ethics trumps any desire for statistical measures.

Perhaps we can obtain useful information by direct observation of women using a potential microbicide for another purpose. Professor Short in Australia and Conrad in the United States have shown that lemon juice is an effective microbicide. In some parts of the world, sex workers have a tradition of using lemon juice. Next month, a team from UC Berkeley will work with colleagues in

Nigeria to explore the consistency of use in one such group.

Whatever the study design, the outcome measure of interest will be use effectiveness, not biological effectiveness. Dr. Stein has just mentioned the very useful paper by Dr. Foss and colleagues that shows that while condoms are likely to be more effective than a microbicide, microbicides are more likely to be used consistently.

Personally, I think the overlapping use effectiveness might justify a straight Phase 3 comparison where a microbicide would be tested against condoms as a gold standard for protection.

I think we can demonstrate that a microbicide will not damage a woman's vagina by escalating dose studies in volunteers not exposed to infection, and we can make a plausible case that a microbicide has some degree of effectiveness based on in vitro studies.

Ultimately, we are called upon to make difficult judgments. Do we emphasize the needs of the women who we know will not use condoms or the needs of those swept up in a trial who will use condoms? As I said, I can't find a method that

will cover both.

Do we think it is possible to collect enough in vitro and collateral clinical data to judge the efficacy of microbicides will be in the same range as condoms? I think we can; others obviously will disagree with me.

Can we approve a method because it is comparable to condoms, but we do not know its true efficacy?

I am opposed to a condom-only arm, but with or without condoms, given the numbers and durations of trials suggested today, it is my judgment that non-FDA-approved trials probably in Africa and Asia will provide useful data before an FDA-approved trial is completed.

My plea to this Committee is to recognize that ethically-acceptable ways of designing clinical trials to test the efficacy of microbicides are not cut-and-dried, and sincere people can have a variety of views. I am confident the Committee will be cognizant of all possible alternatives.

Thank you.

[Applause.]

DR. GULICK: Thank you, Dr. Potts.

That concludes the people who signed up to speak at the open public hearing. Just to let people know, there were three written submissions submitted to the Committee. Those were emailed and faxed to Committee members, and they are in your packet as well. One is from Laurie Sylla, from the Yale University School of Nursing, one from Dr. Robert Munk from the New Mexico AIDS InfoNet, and one from Anna Forbes from the Global Campaign for Microbicides.

Is there anyone who didn't sign up for the open public hearing who would wish to make a statement at this time?

[No response.]

DR. GULICK: Okay. We will close the open public part of the meeting, and we'll turn to Dr. Birnkrant for the charge to the Committee.

Charge to the Committee

Questions to the Committee

DR. BIRNKRANT: Thank you.

I would like to begin by commenting on this morning's presentations. I know that I found them extremely interesting, and I know that my colleagues also found them interesting, and I know that they will lead to productive discussions this

afternoon.

I also want to thank the speakers during the open public hearing for their presentations as well.

There were a number of different views presented this morning and this afternoon, but that's good, because it makes us think about all types of possibilities, and we'll take some of these ideas back to the agency, mull them over and apply them to some of the advice that we'll be giving to sponsors.

So although there may not have been consensus with regard to particular issues, there was consensus, though, with regard to urgency. And as the speakers this morning and this afternoon pointed out, there is an extreme urgency to develop a topical microbicide rationally and get it on the market as soon as possible.

Another point I want to make is that what we are discussing today may apply only to the first generation of topical microbicides. That is, the need for a three-arm trial with two controls may be more appropriate for the first microbicide, but may be less appropriate as more microbicides reach the market. And we are well aware of that.

A couple of comments with regard to flexibility, standards, and risk-benefit. With regard to flexibility, the FDA has shown that it can be flexible in a number of areas, in a number of drug approvals that have taken place in the past. But with regard to microbicides, we can show flexibility in that we are willing to accept the one clinical trial as opposed to two adequate and well-controlled trials, we are entertaining the idea of having a P value between .01 and .001, et cetera.

With regard to standards, some people call our standards "hurdles," but I like to look at them as standards set for the world. And what are these standards? Well, our regulations in the Food, Drug, and Cosmetic Act that was amended in 1962 tell us that we need substantial evidence for a product to reach the market.

And what is the substantial evidence? Well, it has been interpreted as being not only safety but efficacy, and the efficacy should come, it has been interpreted, from adequate and well-controlled trials.

We have interpreted that traditionally as two, but we have a guidance document that does

allow for one large clinical trial that is multi-center, internally consistent, and highly statistically significant.

What does it mean, though, to have these standards? These are standards to allow us to approve a drug that is safe and effective in which we have a lot of confidence. And these standards, although they are U.S. regulatory standards, should apply to the whole world in that if it is a safe and effective drug for the United States, safety and efficacy should be the same whether you are in a developed country or a developing country.

So we feel as though the standards are absolutely the same.

With regard to risk-benefit, we look at risk-benefit on an indication basis, so we develop risk-benefit standards for various diseases. It may be different for cancer as opposed to sinusitis. But when it comes to HIV prevention, the risk-benefit is the same throughout the world. It doesn't matter if a drug is coming to the FDA for review and approval or coming to another regulatory body outside the United States.

The risk-benefit should be the same in that there should be greater benefit than risk to

the population.

Lastly, what are the risks of putting a less-than-effective microbicide on the market? Well, they are great. And why are they great? Because they may lead to high-risk behavior and thus increased transmission rates, and they may also lead to condom migration. And we wouldn't want people migrating from condoms to a much, much less effective and safe product.

With that, I'd like to turn to the questions.

The first question deals with trial design, which we have been wrestling with actually for a number of years. And as I said this morning, we are bringing it to the Committee today because we have received some proposals for Phase 3 and Phase 2 trial designs recently.

This morning, we and others presented the Phase 2/3 run-in design, which is somewhat different than traditional drug approval that proceeds from Phase 1 to Phase 2, where activity is shown, and then to Phase 3.

What we are looking for the Committee to discuss is the pluses and minuses of these different types of trial design and perhaps to

suggest alternatives to helping us provide sponsors with advice on Phase 3 clinical trial design.

DR. GULICK: So shall we take them question-by-question, or do you want to run through them all?

DR. BIRNKRANT: I think we can do it question-by-question, because they have multiple components, so it may get too complicated if we run through them all at this point.

DR. GULICK: Okay. And then, just one other point of information before we start. Could you or someone else review again the HPTN 035 study, the design of it and where it is in terms of development? We have heard a lot about that study over the course of the morning.

DR. BIRNKRANT: Maybe Dr. Karim can do that.

DR. GULICK: Thanks.

DR. KARIM: Thank you.

The HPTN 035 trial is an NIH-sponsored trial that is part of the Prevention Trials Network. It is a four-arm trial which involves two active products. One is Buffergel [phonetic] and the other is Pro 2000 [phonetic]. And it involves two control arms--a placebo control arm and a

no-treatment control arm.

The trial itself is being conducted--or, we plan to conduct it--in approximately--well, at this point, starting off with four countries and eventually expanding to seven sites throughout the world.

The current sample size and design that we have proposed is a Phase 2 leading into or running into a Phase 2B design, and we propose to study approximately 3,100 subjects in this study.

We are proposing that in conducting the study, each product would have to be shown to be effective either against the placebo arm or the condom-only arm in order to be regarded as efficacious.

Thank you.

DR. GULICK: And Dr. Karim, what is the status of the study? Has it begun?

DR. KARIM: No, the study has not begun. We are just preparing the final submission, what we hope to be the final submission, to the FDA, and it has gone to the NIH for regulatory approval. We anticipate enrolling the first patients early in the new year.

DR. GULICK: So the design is finalized,

and it has gone to the FDA and NIH for final approval.

DR. KARIM: That's right.

DR. FLEMING: I might just add to that, some of the more detailed statistical properties were those that I was presenting on the slide in the presentation in terms of the ability of this design to fairly reliably identify ineffective interventions and reliably identify effective interventions, at least in terms of either providing conclusive evidence of benefit or evidence of need for continuation of study. And NIH convened an external body in I think it was March to review this design, and it was endorsed by that body; that was one of the more recent actions.

DR. GULICK: Okay, thank you.

So let's turn to the question at hand, which is to comment on two different proposals, and then we'll take some suggestions. So let's as a Committee consider the first design--a Phase 2 run-in Phase 3 trial design.

Pros and cons? Dr. Paxton?

DR. PAXTON: Well, I think there are some significant pros to that approach. One is that for those of us who have done significant trials

abroad, logistically, it is much easier to not have to come to a complete full stop and let your patients go while you do your analyses and all that.

I think the advantage of doing a Phase 2 run-in the way this is, you don't stop, as was shown in one of the prior slides. You do manage to keep the women who were in the Phase 2, and they do continue to give you more information in your Phase 3. So I consider that to be a very significant pro.

Are we allowed to talk about the B part, too, or do we just want to talk about A right now?

DR. GULICK: Let's take one at a time, and then we'll come back to that.

DR. PAXTON: Okay.

DR. GULICK: Other comments on this design?

Yes, Dr. Fleming?

DR. FLEMINg: I think with the Phase 2 run-in to the Phase 3, one of the advantages of this design is we had mentioned the benefits of Phase 2 are multifold, one of which is to provide an extended experience in safety beyond what you would have in Phase 1, to be basically in a

position to justify the exposure of large numbers of participants in a Phase 3 setting.

So this Phase 2 run-in in essence allows one to restore that type of insight that you would have hoped to have gotten if you had had a separate Phase 2.

The limitations of the design are that in essence it is a Phase 3 trial, so you are basically, then, at this point jumping to a Phase 3 from a Phase 1. If in fact you believe that you understand what is necessary in order to design this trial and conduct it in a high-quality fashion, and you have a belief in plausibility of efficacy, it is a very appropriate next step.

So if you are confident that you have the right question, you have the right way to carry the study out, and you are adequately optimistic, you believe that you have established plausibility of efficacy, it makes sense to move into this step.

On the other hand, if there are key issues about quality of study conduct and implementation that are not fully understood that end up being better understood during the early phase of this trial, it can be very problematic in interpreting the result.

DR. GULICK: Dr. Fletcher and then Dr. Sherman.

DR. FLETCHER: In thinking about this Phase 2 to Phase 3, I think I need some help from my statistical colleagues to think about protection against proceeding when you shouldn't. Let me see if I can lay out a scenario.

Let's say you had done the traditional Phase 1 to Phase 2 to Phase 3, and in the Phase 2 study, you were left with, let's say, equal rates of seroconversion, which I think would be evidence, then, that the product has no evidence of effect, and therefore, why go on to Phase 3.

How would you have that same protection against going on to Phase 3 where now you expose a large number of individuals to a product that is not effective with a Phase 2 to Phase 3 lead-in? I don't quite see that.

DR. GULICK: Dr. De Gruttola, do you want to respond?

DR. DE GRUTTOLA: Yes. I think what Tom said--this is basically a Phase 3 study; you are just calling the first part of it a Phase 2--and like with any Phase 3 study, you can have stopping rules that allow you to stop for futility. So if

you have enough information to say that in this study, you are very unlikely to conclude efficacy, you could stop. That doesn't mean you have necessarily proved it doesn't work, because to prove it doesn't work may require the full information; but you may have enough information to say that in this study, you are not going to get an answer and to allow you to put an upper bound on what the efficacy is likely to be.

So I think if you just think of it as a Phase 3 study in which you are going to do kind of an extensive first interim review to make some decisions about whether to fully enroll or not, and the information you may use may, like in any Phase 3 study, include both toxicity and stopping for futility, that that is a way to think about it.

DR. FLEMING: I think this is a terrific question because it really gets at the essence of an issue that needs to be understood as you think about the appropriateness of launching this Phase 2/3. I fully agree with the explanation that Victor has given, and let me just try to add a little bit of specifics to make clear what the implications are of what he was saying.

Some people have said if, for example, you

do a Phase 2B trial as a separate trial, and it is based on one-quarter the number of events--and as you know, in an analysis such as this, information is number of events; if you have 100 events and 100 people, that is the same information as 100 events and 10,000 people in terms of statistical power to discern treatment effects--so if you are going to do a Phase 3 trial with 400 events, or a Phase 2B trial with 100 events, just do an interim analysis in the Phase 3 trial at 100 events, and don't you recover the same information.

The essence of the answer is not at all necessarily. I always say to sponsors that if you do a Phase 3--and this Phase 2/3 is a Phase 3, as Victor said--write the check for it, because in essence, if you want to preserve the power to the Phase 3 trial, you have to be very cautious about what you consider to be extreme results early on.

So very typical monitoring boundaries would stop a trial for lack of benefit when you have what--when you basically have an estimate of no effect when you are halfway through. Whereas you do get much earlier than that evidence about lack of benefit in a separate Phase 2B trial that would be based on just 100 events where, if you

recall, we were saying there a negative study would be an estimate of efficacy that, based on only 100 events, might be anything less than 15 percent.

So because of the need for conservativism in a Phase 3 trial to preserve the power and the preserve the false-positive error rate, you actually do end up going further into that trial, even if you are using interim monitoring, before you would, so to speak, shut off the faucet.

So again I come back to a Phase 2 run-in for a Phase 3 is a good idea in certain settings--when I am really confident I have the right question, I know how to design the trial in the right way, I know how to be able to achieve adherence, I know how to retain, I know how to enroll, and I believe plausibility of efficacy has been established.

So the question in this setting is can you do than when you have had a 100-person Phase 1 trial. If the sponsor thinks so, this is the right thing to do.

DR. GULICK: Dr. Fletcher, a response?

DR. FLETCHER: Actually, it was to almost that last point you made. So, if the development paradigm then becomes Phase 1 to the Phase 2/3, I

am wondering, then, how you establish proof of concept or plausibility of efficacy if Phase 1 is really to establish dose and bad adverse reactions and those types of things. Where does that plausibility come in in this paradigm to move from a 100-person study to a 4,000-or-so-person study?

DR. FLEMINg: Yes. That too is a critically important question, and as you know, the standard approach to this is to do a Phase 2 trial where we would be looking at biological markers. Those biological markers may not be valid surrogates that reliably tell us about clinical effects, but they give us clues, they establish proof of principle, they establish plausibility of efficacy.

We are at a substantial disadvantage in this setting without such information. We simply don't have those types of measures for plausibility of efficacy.

It then comes down to essentially how much risk is someone willing to take, and it is substantial risk, especially if you are going to deal with a study as the 035 study was planning to be, as a definitive trial looking at a 33 percent reduction in transmission with four arms that was

on the order of 10,000 people and a $100 million expenditure. That's a huge leap to make from a Phase 1 study without a proof of principle result.

It is not unlike what we have struggled with in the vaccine area for HIV for a long time. We have been awaiting having adequate evidence of efficacy. Now, at least there, we have immune base markers, although there is a lot of controversy about is it humoral or cell-mediated or what nature or whatever--we don't even have that in this particular setting.

It is--and now I am jumping ahead to 2B--but it is one of the reasons to say, then, proof of principle could in fact be based on the very endpoint. Doesn't that make sense to use HIV infection itself as the way to establish proof of principle in a somewhat measured intermediate step that is smaller in size?

DR. GULICK: Dr. Sherman and then Ms. Heise.

DR. SHERMAN: I am interested in the concept of this 2/3 run-in, and looking at the outline that was in Dr. Wu's presentation, you have two parallel arms running together. Do you plan to merge those arms in the data analysis? In other

words, will that Phase 2 run-in arm become part of the main dataset as a practical piece of data, and is that valid at the final endpoint of the study because there is going to be differential dropout and bias between those two groups?

DR. FLEMING: The answer is for those who advocate this design, their answer is yes. Is that valid? Yes. It is valid subject to the way it is being proposed here, which is that these interim data would be made available only to a data monitoring committee. That data monitoring committee would then assess whether various safety thresholds had been met, and if so, the study would continue, and all those participants would be included.

If, however, these data were released separately to the sponsor, then, many of the issues that we believe are important in monitoring trials would be violated if that same dataset were then used as part of the overall trial.

So the advantage of it being a separate run-in--if the sponsor wishes to have full access to the data, that's entirely possible, but then you would start over. But the way this was being proposed, which is an acceptable approach that some

of the sponsors were saying, is that this would only be viewed by a monitoring committee. Now, granted the sponsor doesn't have weigh-in in this now, except for the procedures and the criteria they set out in advance. The monitoring committee would then review this, ensure that the safety criteria were met, in which case then it would be acceptable to use all of the participants, including the two run-in participants, in the overall analysis.

DR. GULICK: Ms. Heise?

MS. HEISE: I have two points. One thing that I think is important in terms of evaluating the appropriateness, as Dr. Fleming said, of a Phase 2 run-in is whether the conditions apply that you actually know you can do the study. And I think that one of the things that is important for people to realize is that at every site where these trials are being mounted, there is a preparatory study called a feasibility study, a site preparation study, where in effect they are enrolling women, seeing whether or not they can follow them up, looking at retention, seeing what level of incidence is achieved with the condom counseling and the like.

So it is not like you are going from a Phase 1 study to this fullblown study without having field-tested any of it. I think that is an important thing. Frequently, that is at least a year-long feasibility or preparation study.

Then, the second thing--and someone should correct me if I am wrong--I think that it is not just a Phase 3 study with an interim analysis. I think what is being proposed is that there are certain types of safety tests, whether it be a colposcopy, cytokines, all kinds of things, which are done on the women in the Phase 2, on a subset, because it is very, very complicated in these settings to do 3-month colposcopies on 10,000 or 12,000 women.

So there are things that are being done to start to elaborate some of our safety concerns that are happening in this Phase 2 part of it, which is what we really think of as an expanded safety, as opposed to traditionally, in which you would be looking at sort of a pre-effectiveness.

So in our kind of development pathway in the field, I think you get a series of safety trials with women at very, very low risk, then women who are at slightly higher risk, then women

who have HIV as well and perhaps other STIs, and you keep trying to get closer and closer to the women who will be enrolled in the larger trial. So this Phase 2 is kind of your last step at trying to establish as best you can that you have all the safety information that we know how to get at this point prior to going on and look during an interim analysis.

DR. GULICK: A response, Dr. De Gruttola?

DR. DE GRUTTOLA: Yes, I would like to comment on that, because you can call it Phase 2, but it really is part of a Phase 3 study, and in fact, you can do intensive safety analyses on a subset in a Phase 3 study as well and then review that information before you continue to enroll.

I think the reason why the terminology is important is the reason that Dr. Fleming mentioned, that usually in Phase 2, you have time to evaluate the study, including the sponsor, and make decisions about how you are going to conduct a Phase 3 study. And in this case, if you do those safety analyses, and during the interim review, you find out that there is a problem, then you have a dilemma. Either you stop and start over again, which means now you have really a Phase 3 study

that stopped, even though you called that part a Phase 2; or you modify the study in order to deal with some of the safety issues that have arisen, but that is complicated in a setting where the sponsor is not supposed to be receiving that safety information, and it raises questions about whether you really should combine the Phase 2 part of the Phase 3 study with the rest of the Phase 3 study.

That's why I think that in certain ways--although I understand the point that is being made, that this run-in part is different, and there is a lot more safety analysis, and it is closer to a Phase 2--to think of the whole thing as a Phase 3 study may be helpful in terms of the kinds of commitments that need to be made. There is no reason not to do it if you believe you have all the information necessary to design the study, but if you are still worried about safety and doing a lot of intensive safety analyses in the Phase 2 portion, then you wonder, are you sure that the results of that information are not going to lead you to wish you had done another study or had designed things differently at the start.

DR. GULICK: Dr. Flores?

DR. FLORES: I would like to get some

clarification on whether the purpose of dragging Phase 2 into Phase 3, in addition to the safety evaluations that would be more intensive, also has an operational component that might actually allow some filtering in terms of the quality of the study, the ability to enroll and retain, and the potential that some sites actually may start early and others may take several months before they start. Is that also part of the purpose of this?

I noticed in one of the previous study--I believe it was the COL study--that one of the sites dropped out early on and had to be replaced. Is that a consideration in this design, or are we just talking about, as Dr. De Gruttola said, a Phase 3 with initial safety evaluation?

DR. GULICK: Dr. Fleming?

DR. FLEMING: Another great point. I think, Jorge, without question, as we continue in our clinical trials research, we learn. And as we learn, we try to implement what we have learned in our future studies to improve the quality and reliability of those studies. And when we do a separate Phase 2 trial, as I was trying to indicate in the presentation that I made earlier today, clearly what we are trying to do is look at safety

and look at plausibility of efficacy through effects on biological markers. But we are also trying to glean whatever insights we can from these types of studies and other preparedness studies to allow us to be in the most informed and best way possible to carry out the most reliable Phase 3 study, including issues that you mentioned, too--the ability to enroll in a timely way, the ability to retain participants at high levels, the ability to achieve high levels of adherence to the microbicide and high levels of adherence to other interventions.

If we launch a Phase 2/3 study without having adequate insights on each of these issues, we're taking a chance, because if we in fact learn these insights during the course of the study, we can make refinements; but if we are sufficiently far into it, some of the inadequacies that emerged early on are going to be there with us throughout the entire dataset.

And, as Victor pointed out correctly, if in the Phase 2 experience, we find substantial safety issues that lead us to make nontrivial changes to the regimen, it becomes very problematic to interpret the aggregate data.

So I keep saying the time to do this is when you do need to verify safety, and you may do so, as Lori was saying, by a more intensive monitoring of these participants. If you are quite optimistic this is going to be a favorable review, this 2/3 is an acceptable approach. If you are very uncertain, and there is a very realistic chance that revisions will need to be made, you are better-off for that to be a separate step that the sponsor can fully weigh in on and then make an informed judgment about how to better design this very expensive Phase 3 trial before it is initiated.

DR. GULICK: Dr. Barlett, then Dr. Haubrich.

DR. BARTLETT: I was going to comment that it seems that from an FDA standpoint, each of these trial designs could be viable within the limitations that have been articulated by Dr. De Gruttola and Dr. Fleming, and really, the risk is being borne by the sponsor, and the sponsor needs, with full transparency and understanding of this, to make decision. But from an FDA standpoint, these could all be viable.

DR. GULICK: Dr. Haubrich?

DR. HAUBRICH: It seems like the biggest thing you don't have from your Phase 2 study is an estimate of event rate which would help you plan how many people you need in your Phase 3. Is it legitimate during a DSMB review of the Phase 2 to adjust your sample size and still use all the patients that you've got?

DR. FLEMING: In fact I would argue in general that is one piece of information I would surely liked to have had up front but I can accommodate for more readily.

If you recollect some of these calculations, I think the CONRAD situation was saying they were targeting a 50 percent reduction with 80 percent power. That takes 65 events. If they had said 90 percent power, it would be 88 events. The example I gave was a 33 percent reduction with 90 percent power; that's 256 events, all of those to achieve an .025 traditional strength of evidence.

All of that is already known up front. What we don't know is the event rate, and that event rate requires us to then adjust either the sample size or the duration of follow-up. That is a totally legitimate thing to do except for the

fact that if it turns out the event rate is one-third of what you thought, the sponsor may not be happy when they get the message that your study is fine--you just have to triple the sample size.

So you are well-advised to get a decent estimate of that up front so that you don't end up hitting the sponsor with such a radical change during the course of the study. I would argue, though, that that is something I can live with as that refinement.

Something I am much less comfortable living with is changes in how to effectively carry out this study during the course of the study or to deliver the regimen to achieve maximum efficacy by getting maximum adherence, and reduce safety by getting a proper way of dosing this. That is the thing that is harder to correct midstream, because now you are changing fundamentals in the study design.

DR. GULICK: Okay. Let me try to summarize what we think so far.

The first thing we did was to remember why Phase 2 exists, and Phase 2 is here to expand our safety information and to gain preliminary efficacy information, typically with effects on biomarkers.

There are also other insights from Phase 2 which help inform the design of Phase 3.

There was a lot of enthusiasm around the table for this kind of design in this setting, realizing, as Dr. Fleming pointed out, that we need insights into the plausibility of efficacy in this stage, that you have to be confident of your design and what your plans are; and other details such as adherence and of course safety are paramount in importance for moving forward with this kind of design.

Other positives to this design mentioned are that it really extends and maximizes the safety information in terms of exposure, because it prolongs exposure in the set of individuals who enroll under Phase 2. As was pointed out, this could also be done in intensive subset analyses.

It also has benefits in terms of logistics and feasibility among the sites, and it is thought to be efficient and a timely way to do this. And the overriding sense of urgency in the field supports this kind of approach as well.

In terms of limitations, as Dr. De Gruttola summarized, this design is really a Phase 3 study, so you are jumping from Phase 1 to Phase

3, essentially. And the main limitation of that is risk itself, and there are several. There is risk in terms of condensing the time of development condenses your ability to make insights as to things that might turn out to be important for the design of Phase 3, but you are proceeding so rapidly that you actually didn't have time to make those observations and adjust accordingly.

As others pointed out, there is a potential risk to patients in that going from a few hundred patients to a few thousand patients potentially involves more risk.

And of course, there is risk to investment and to money here, going from a small study to a large one.

Also, if a safety problem is detected early on in Phase 2, that may actually sink the plans to go forward to Phase 3.

As was said, there are problems with details and uncertainties, but many of these, particularly the safety and early efficacy rules, could be addressed by writing in early appropriate stopping rules into the protocol, particularly for futility. And as was mentioned, it might be possible to adjust for event rates although other

significant changes would be problematic, such as differences in dosing schedules or adherence rates than what was initially planed.

All together, it was felt that if this kind of design were implemented, the first part of the study, it is critical to keep those data and information only accessible to a blinded interim review committee, that they should not be generally accessible by the sponsor or others, and then it would be appropriate to use that information in support of the Phase 3 endpoints as well.

Okay. Let's try another one.

Stand-alone Phase 2 targeted at high-risk groups, i.e., commercial sex workers, followed by a Phase 3 study. Please comment on the feasibility and, more generally, other design issues with this.

This is the more traditional development.

Dr. Haubrich?

DR. HAUBRICH: I think there are several advantages to looking at high-risk populations. Number one, I think some of the safety concerns might become evident earlier if there is a dose response as was seen in the 9 study [phonetic] that was presented, where I believe the people who used it the most had the worst outcome. So in that

sense, it could actually provide insight to safety.

At least my understanding from reading some of the material that was presented is that the Phase 1 studies are going to be fairly short in duration, and if appearance of lesions and stuff like that takes time and exposure to develop, you could be going into a Phase 2 study without having enough safety data; that may not appear until later.

So it seems that targeting high-risk populations could be advantageous from that standpoint. And jumping ahead a little bit to C, it seems to me that a Phase 2 lead-in might include some targeted populations to try to pick up early on some of these safety events as well, although it might confound the overall thing I talked about before, which is the rate of events, because it might be higher in that subgroup.

DR. GULICK: Other comments on the traditional?

Dr. Mathews.

DR. MATHEWS: I think this question raises some issues that we have not made as explicit as perhaps we should. I am referring to the concept of efficacy, effectiveness, and proof of principle,

which have been sort of thrown into most discussions today. It was only made explicit, I think, in Tom's presentation where you explicitly stated that effectiveness was the comparison between condom and microbicide, and efficacy the placebo versus microbicide. But I think those concepts really mean a lot more than that.

My understanding of an efficacy trial is one which you plan so that you have high adherence throughout the trial, and the trial is done under the conditions which are most likely to show an effective, and usually, it requires a homogeneous population that is studied, such as commercial sex workers, for example. Whereas effectiveness means a heterogeneous population who may be doing other co-interventions and so on throughout.

So I have wondered throughout the day exactly what an efficacy trial looks like in this way, and at the point the field is in right now, such an efficacy trial is really a proof of principle trial since there is nothing out there that has been shown to work yet.

So I think those have implications for who is studied, how long they are followed--for example, if people are followed for 24 or 48

months, and adherence wanes, which it probably does, at least it does in antiretroviral trials, those factors need to be taken into consideration--the intensity of the monitoring, and also another issue, for example, whether incentives should be provided to assure compliance with study visits and so on, which may not be part of a larger effectiveness trial.

So this question, should Phase 2 be done in a high-risk group, I would say whether it is Phase 2 or Phase 3, what is the purpose of the trial. If it is to establish efficacy, I think it should be done with the shortest duration of follow-up consistent with achieving high adherence, with very frequent follow-up consideration for incentives.

And the issue of homogeneity really raises issues about the characteristics of sites, because if, for example, in one site of commercial sex workers, condom use is very high, but in another very low, you haven't really achieved a homogeneous study population despite the fact you thought you were studying high-risk individuals.

DR. GULICK: Dr. Stanley.

DR. STANLEY: I think it's important to

target high-risk individuals in the Phase 2, because this is different from a drug to treat something. This is an agent that is used at the individual's discretion as often as they wish. And therefore, to prove safety, you have really got to expose folks to high levels of this that might reflect that end of the curve of folks who will be using it a lot in real life.

So I think that it is a little difficult to take somebody who is not at risk and expose them to high levels of this and cause damage, whereas if you have people who are going to be placing themselves at high risk and are going to be using a high level of this, they are the ones who are prime candidates for looking for safety.

DR. GULICK: Dr. Paxton?

DR. PAXTON: I guess I am going to just reiterate what Dr. Stanley said. I think this is a very efficient design to take the most high-risk women and study them first, because we learned from COL 1492 that there can be significant differences between high-frequency users and low-frequency users, and this way, we would find that out much more quickly.

Of course, a minor consideration is that

you might end up losing a product that might have worked well on somebody who has low-frequency use. However, I would argue that women who do use it very frequently are going to be using it, so therefore, maybe you still deserve to lose it.

So I think that that is a very efficient thing, and in a Phase 2 trial, again, since it is mainly safety and not so much efficacy; safety is the main thing we are looking at there.

DR. GULICK: Dr. Barlett, then Dr. Fletcher.

DR. BARTLETT: I'd like to ask Drs. Stanley and Paxton with regard to practical issues of doing this study in high-risk women, are these women--presumably, you might be recruiting them at international sites, and they would require more intensive follow-up with colposcopy and other issues. Does that affect this decision and make it any harder?

DR. STANLEY: Not to me, because if you are doing a time-limited Phase 2, you have got to apply those resources to it.

DR. PAXTON: Right. And we have experience with using sex workers and having them come in for colposcopy and the like, and yes, I

think it is feasible--if that's the question being asked as to if we could comment on feasibility and can it be done, yes. And should it be done, I would also say yes.

DR. GULICK: Dr. Fletcher.

DR. FLETCHER: I wonder if there might not be another advantage to this Phase 2 and high-risk commercial sex workers, and that could be an overwhelming demonstration of effectiveness. I have already gotten my certificate from Dr. Birnkrant, so maybe I can be a little bold here--

DR. GULICK: Let's be careful here.

DR. FLETCHER: --yes. Could you comprehend licensure after Phase 2? What if this were a 400-person Phase 2 study, and you had P less than .00--maybe even .01--in terms of seroconversion and excellent internal consistency, and everything just said this product works.

While you still may have a safety issue, in the past, FDA has certainly used Phase 4 to provide further evidence of safety. So what I am wondering is beyond just looking at safety as Dr. Stanley talked about because of frequency of use, might it also be an avenue that, for a product that showed overwhelming evidence of effectiveness--or,

I guess it would be efficacy--could it be approved at that stage with requirements for further demonstration of long-term safety?

DR. GULICK: Dr. Birnkrant, do you want to respond? He's asking you.

DR. BIRNKRANT: That's funny--we were asking you that question.

[Laughter.]

DR. BIRNKRANT: Because as it is written, not up on the screen but on the paper, we were concerned about the feasibility of conducting a Phase 3 trial after results were obtained from the Phase 2 that looked promising.

So do people feel as though a Phase 3 trial could then be conducted following promising results from a small Phase 2 study in a high-risk population?

DR. GULICK: Dr. Haubrich?

DR. HAUBRICH: I think we have all seen in the HIV field things that look very promising even with highly significant P values and very small numbers of patients that have turned out not to be true. The whole issue in antiretrovirals evolved to using surrogate markers and interim approval drugs based on fairly small Phase 2 studies when

other studies are planned and on the way.

I think it would be very dangerous to set that precedent here, although I think highly tempting to do so if a small study, even if it looked promising, would then preclude the use or any further Phase 3 study that compared to placebo. We have all argued the differences all morning of why there is so much confounding, and you need to do good placebo-controlled studies, and then to blanket, if you approved a product based on a 400-patient study, that would then make it unethical to carry out any other placebo-controlled studies then. So I think that would be a very dangerous thing to consider.

DR. GULICK: And I think the flip side of that is the safety issue. Clearly, judging safety based on a 400-patient study in a product that could be used literally by millions of people for years is difficult to do.

Dr. Sherman?

DR. SHERMAN: That said, a freestanding Phase 2 and a single Phase 3 is very appealing as an approval mechanism. If both of them separately--you do have two studies; one is a small Phase 2 in a high-risk group--it seems to meet

several of the needs that we discussed in this committee before.

DR. GULICK: Ms. Heise?

MS. HEISE: I think there is a concern that we need to consider safety issues among women who may have high frequencies of use. I think there is a separate issue, though. The assumption that people often make that there will be a higher event rate in a trial among sex workers has not been borne out in fact, because what we do know is that when we do our condom counseling with sex workers, these women are actually in a better position to be able to negotiate condom use with proper support.

So the working assumption that many people had in this field 10 years ago was that the obvious quote-unquote "population" to enroll in these trials was sex workers, there would be a higher incidence rate in the trial than if you had women in the quote "general population."

What we have found, and there is actually data to support this, is that frequently, because of the concomitant condom use that is achieved in these trials, you actually have incidence rates higher. It doesn't address the safety issue, but I

just wanted to point out that in this kind of design, you would actually have two separate populations probably. You wouldn't be able to have a population where you enrolled and used the same clinical and the same site and the same outreach workers and the same everything because you would be doing a safety trial among high-risk women, and you would most certainly probably want to do part of your large phase retrial among women recruited through family planning clinics or VCT clinics or whatever.

So I think there are real feasibility issues in the sense that with the run-in kind of scenario with safety, you are talking about a site infrastructure that you have developed over time that you are maintaining and that you are continuing, whereas with this, you may well be talking about two totally different sites, two different infrastructures, and two different teams.

DR. GULICK: Okay. So, as Dr. Bartlett pointed out, the consensus really is that we find both of these designs acceptable and that they each of pros and cons.

We were very accepting of the traditional approach with all the pluses, and people began to

gravitate right away to, well, how do you really prove proof-of-concept in Phase 2 if you did a stand-alone study, and the suggestion we leapt to was to look at an appropriate Phase 2 population that you could really study efficacy in. And the feeling was this should be somewhat of a homogeneous population.

Commercial sex workers were suggested although, as Ms. Heise just pointed out, that may be debatable in terms of risk of exposure. Certainly this would be a population who may use the product at a higher rate than others. And as others pointed out, you could counsel for adherence, make sure you had adequate follow-up, pick your sites to achieve a homogeneous population, traditional Phase 2, trying to prove the principle before you go into Phase 3.

All the negatives we mentioned before with the timely way of going from a Phase 2 run-in into Phase 3 become pros for the traditional approach. That is, now you do Phase 2, and you describe early insights that help you design your best Phase 3 studies. So those are obviously pros.

The two main cons that were cited for this design, number one--we didn't even state it because

it is so obvious--but this is slower. This clearly would take years longer than the previous approach. And as we heard from the beginning presentations today, the urgency of evaluating microbicides is great.

And then, as pointed out, feasibility of doing this, looking for this highly homogeneous population may be difficult to truly prove this proof-of-concept that an early candidate drug would work.

Then, you specifically asked us would a very convincing Phase 2 not allow us to go to Phase 3, and again, some discomfort with making the jump from a very convincing small Phase 2 study right into approval, both with efficacy and safety information.

And then, as Dr. Sherman suggested, possibly a convincing Phase 2 plus a Phase 3 might do the trick.

Shall we consider Point C--are there other alternative designs that people would like to suggest?

Dr. Fleming?

DR. FLEMING: I'll be brief, because I spoke about it at some length in my own

presentation this morning.

A variation, an alternative, would be the 2B intermediate trial which would be in philosophy more like Step B, because it would be a separate step. It would in fact be a study that typically would be one-third to one-fourth the size of your full-scale highly-powered Phase 3 trial. Its advantage is that it would provide for significant insights in quality of trial conduct issues for the ability to implement these insights in the design of any subsequent Phase 3 trials. It would provide extended safety experience in a controlled fashion. It would clearly provide very strong proof-of-concept insights for efficacy.

And there is a little bit of semantics here. If we look, for example, specifically at the implementation of this design in the 035 setting where, as Salim was talking about, it is a 3,100-person trial targeting the ability to get roughly 100 events for every pair-wise comparison, that actually is larger than some of the Phase 3 trials that we have heard about from others that are targeting bigger differences.

So in fact, it is semantics--it is a Phase 3 trial for a more aggressively assumed treatment

effect, but for a more conservative but nevertheless important treatment effect, it would in fact be more likely a Phase 2B trial.

DR. GULICK: I guess one issue that hasn't come up at all is a crossover design particularly for women who would be randomized to either the placebo after some period of time or, if we decided to proceed with that design, the no-treatment arm. That's a way to continue obviously people who randomize to, quote, "less attractive" arms in follow-up in the study if they are assured with being either re-randomized or getting something later on in the design. That is an effective way to address that.

DR. PAXTON: Is it really effective, though? It seems to me that since HIV is a definite endpoint, and once somebody has reached that, you can't get rid of it--there is no washout period.

DR. GULICK: No; I don't disagree with that. I guess I was referring to--let's say we recommended or a study was designed with the three arms, and there was the no-treatment arm, that part of the design of the study up front could be to offer that group the intervention later at some

point.

Dr. Wood?

DR. WOOD: In terms of alternative designs, I just wanted to throw out there the idea of possibly in terms of design scheme and randomization, rather than randomizing individuals, consider randomizing communities or populations. This could potentially be done during a Phase 2 study in which you have two centers of sex workers but one center is going to be randomized to receive the microbicide and the other will be randomized to receive the placebo control gel. That would allow you to look at safety issues in terms of intensity and frequency of use. Hopefully, the populations would be homogeneous in one sense in that they are commercial sex workers having intensive exposure. You might have a greater rate of events between communities if you have a community approach. And it might allow for a better assessment potentially of efficacy as well as an assessment of use effectiveness in a population that might allow generalizability when you went to a larger Phase 3 trial.

We haven't talked about that, but I just want to throw it out there. I don't know if it

makes it logistically harder or more difficult to do, but if it allows you to get a clearer answer by using populations and making things cleaner in terms of having the randomization at that level as opposed to the individual level, is that something to be considered.

DR. GULICK: Okay. So a brief consensus here--again, as John Bartlett pointed out, all of these designs may be appropriate. We identified pluses and minuses. As Dr. Fleming said, some of this is semantics. A Phase 2 study of 2,000 people is really more likely a Phase 3. And then we heard some suggestions about crossover and randomization of centers or countries as opposed to individuals.

Okay. Shall we move to Question 2?

DR. BIRNKRANT: That was helpful. We can move to Question 2.

DR. GULICK: As long as we are helpful.

DR. BIRNKRANT: Question 2 is a discussion of the debate between a three-arm design versus a two-arm design. And as I had mentioned, this may apply to the three-arm design, that is, more for first-generation microbicide than to subsequent ones that reach the market.

With regard to a two-arm design, though,

we do have a question as to whether or not the control should be placebo or a no-treatment arm.

DR. GULICK: So it is probably easier to discuss this as a group rather than take them one by one.

DR. BIRNKRANT: Yes.

DR. GULICK: Dr. Stek?

DR. STEK: I want to echo the comments that were made earlier about the inability to properly evaluate a no-treatment arm. I am a gynecologist, and I know how difficult it is to get accurate information about sexual activity, and I think we just make an uninterpretable result.

However, I would like to point out something that really hasn't been brought up about who is not going to be using condoms. It was pointed out that in the African experience, the women who are at the highest risk are those who are trying to get pregnant, so they will not be using condoms. And I know that some of these products are probably going to be designed to be contraceptive as well, but also, there are products that should be available for women who want to avoid HIV infection and are attempting to get pregnant.

I know that studying any kinds of medications or anything with HIV in pregnancy is very complicated. However, I think that we should not ignore this problem. I would urge this to be incorporated in the study design. As far as I know, the products that are under consideration have not undergone the more advanced reproductive toxicity evaluations, and I think that that probably should be done.

There are a number of reasons why this is really important. Women are going to become pregnant. They always become pregnant on any kind of HIV study that I have been involved in. And the risk, we think, is probably the highest for bad outcomes with exposure very early in pregnancy before women have had a chance to discontinue the treatment.

Also, there is the issue of perinatal transmission. We think that acquiring HIV during pregnancy greatly increases the risk of transmission to the fetus as opposed to someone who has already had HIV for a while.

So I know it is a difficult issue, but I think that it should be considered to not discontinue treatment in pregnant and do the

studies that would be required to assure safety in use in women who are attempting to get pregnant.

DR. GULICK: Dr. Stanley.

DR. STANLEY: Well, I have a real problem trying to compare a potential microbicide with just condom use only, because that is relying on behaviors, and behaviors are going to change depending on the options that they are given, as many of the speakers pointed out.

The reality is that once there is a microbicide on the market, there is a population that will probably stop using condoms as we heard from the African experience. So what are you gaining by comparing two options that in fact are not stand-alone options that are going to be out there in the real world once a microbicide is approved.

So I think you confound the issue. I think that you have the potential to rule out an effective product. Even the FDA said that if you have the three arms, you do have to know what condom use is. Well, you are not going to know because some of these patients are telling you what they think you want to hear, not necessarily what they are really doing on a day-to-day basis. So

you will never know what their condom use is, and I think that trying to include that arm is really a confounder.

DR. GULICK: Dr. Bartlett, and then Dr. Paxton.

DR. BARTLETT: I just want to echo what Dr. Stanley said. I was moved by Dr. Dominik's presentation about how the results could be affected by the lack of blinding and the differential condom use, even though in the small Cameroon study, it didn't appear that there was a big difference. But if there is a difference, it certainly could have a big impact on the result.

DR. GULICK: Dr. Paxton, and then Dr. Fleming.

DR. PAXTON: I think I am adding my voice to the chorus that we heard today. I think that we have heard significant and very plausible concerns about including a condom-only arm in that we will probably have unintended and, most importantly, unmeasurable effects of that arm.

Another thing that was alluded to but not specifically brought up but which we have in our packets is what the actual cost would be of these things in terms of money, but that also leads into

issues of time, and we realize that we don't have as much time as we would like to have.

So my personal belief is that what the FDA should require should probably just be the two-arm microbicide versus placebo trial. However, I echo what Tim Farley said. I think that the possibility of allowing for a three-arm trial--the scientific part of me would like to actually look at this to see what we can measure in a three-arm trial, but I don't think that that should be required by the FDA for these trials to go forward.

DR. GULICK: Dr. Fleming and then Dr. Flores.

DR. FLEMING: I guess I would say in conducting a Phase 3 definitive trial, it is really critical to answer the fundamental questions that are unknown. And as I think about ultimately, what do I want to know--I want to be able to do clinical trials that will assess what the real world role of an intervention would be. That is the traditional approach that I would always take. And a topical microbicide is really a regimen, and as regimen, there are I would say at least three areas of ways that it can affect a woman's risk of transmission.

One is the intended anti-microbial effect.

Another domain of ways that it can be affected is through other elements of the regimen, specifically, its physical barrier effect, its lubrication effect, and other effects as well. Those are other true protective effectives that the regimen can have. A third is that it may in fact have an intrinsic effect on the nature of risk-taking behaviors that an individual is embarking on. If in fact it has such an intrinsic effect, I would argue that that too is something that I eventually need to understand.

Now, what do I know from the comparison with the placebo? Somebody said it is an unbiased estimate of product effectiveness. And Chris, I'm going to come back to your earlier comments. We may use these terms in slightly different ways, so I'll just be precise in the way that I am using it.

I would think of efficacy as what is the effect of the microbicide in that hypothetical setting in which risk is identically controlled. To my way of thinking, that would mean that I want to include in that not only the antimicrobial effect, but if in fact the microbicide has other protective effects through lubrication, physical barrier, et cetera, I would want that in my

efficacy, and my concern is that that requires knowledge that the placebo is inert. I don't know that. So I don't know that a comparison with placebo is actually going to give me an unbiased estimate even of efficacy.

So I come at this saying I don't want to make assumptions about what I don't know. I would like to have the clinical trial be done in ways that can provide insights.

The other aspect is if in fact there is a true intrinsic change in risk-taking behavior, whether it is an increase or a decrease, it is something that I would want to know. Somebody had mentioned at the break that condoms are so effective that certainly we want to be sure that we aren't doing something that reduces adherence to condoms. Let's say that the adherence to condoms by virtue of being assigned to a microbicide, which is an intervention that you might think is protective, leads you to reduce your adherence to condoms from 90 percent to 80 percent, so you are doubling the number of people who aren't using condoms.

Somebody said that in the statistical calculation, that is going to decrease my power.

It should decrease my power, because if that's the truth, then the overall net benefit of this intervention is diminished.

We spent more than a decade talking about what is the standard for strength of evidence for an HIV vaccine. I was talking to one of my colleagues recently as I was defending what we are talking about as our standards for approval of microbicides, and I was saying we are targeting a 33 percent effectiveness ruling out no difference.

This person said, "What? For vaccines, we are talking about having point estimates high enough to rule out a 33 percent protection," because specifically, the point was that if you are on an HIV vaccine, and risk-taking behavior because of your sense of protection here is increased even by a modest amount, that would offset the overall benefit, and as a result, modestly protective vaccines may in fact not provide net benefit.

So with that as a backdrop, suppose you were in the setting which I described in my transparency, which was the middle setting on the left-hand side. Supposed you finish the study with only a placebo control. You have a 2 percent annual transmission rate in the microbicide and 2.5

percent in the placebo. That's a relative 20 percent reduction, just barely marginally on the area of statistical significance, that wouldn't in fact be evidence that would readily be judged to be conclusive. And if somebody says, wait a minute--you are estimating a 20 percent protection, when we actually think it is likely that there could be an associated reduction in implementation of condoms? How do I know that that in fact is adequately protective?"

And I come back and tell you, But we had a third arm. We had an arm that in fact compared directly to an open, unblinded experience. And I accept that the overall level of use of condoms can change. I want it to be real world. I am not trying to make that third arm the same level of use of condoms. I want to find out what happens when you are on an intervention that you think is protective against standard of care. And if in fact I have that third arm, and what in fact I found out is there is every bit as much protection--it is 2, 2.5 and 3--I am greatly reassured, first of all because I am getting a sense that the overall 20 percent reduction of efficacy might in fact be an underestimate of

efficacy because there is actually an additional level that the placebo blinded out.

Secondly, I can be reassured that I am not in fact losing this net benefit with condoms. I would think we should be very worried as we look at globally establishing efficacy of these interventions that we recognize that a microbicide regimen is a regimen that involves the anti-microbial effect, other protective effects, and a behavioral component, and if we aren't confident that we are able to maintain within a reasonable level adherence to condoms that we know are highly protective, then we don't have a regimen that is going to be effectively aiding the population, at least in the way it is being implemented. Shouldn't we know that?

The bottom line is I don't think there is a single right answer to this. I would accept, after all the discussion, that the agency should view there to be some flexibility in how these studies are designed. I don't consider that every study needs to have a placebo and an open label. But I do think that there is a need for a foundation of at least one or two early-generation studies that will provide us insights not only

about what the comparison is to placebo but what the overall more net benefit and effects would be that other studies can then build on and wouldn't necessarily also have to have the dual control.

DR. GULICK: Dr. Flores.

DR. FLORES: My basic problem with your concept, Tom, is that we don't know whether the trading in condom use is going to be similarly proportional in the three groups, and that is the big conundrum here. Because they are in a different arm altogether, there may be a totally different rate of lack of adherence to condoms.

The other problem I have with this concept of requesting or requiring it the first time around, I am not making it necessary later as if the trials are just going to keep rolling over in the same population and using exactly the same placebo, perhaps; I am just repeating the same thing. Either it is a concept that should apply to all the studies or to none.

DR. FLEMING: But Jorge, I think the very concern that you have is the essence for why I think there need to be foundation studies to address the point.

What you were saying is you are concerned

that there may be a different level of condom adherence in the two blinded arms from the open arm, and I am accepting--I share your concern. I don't know whether there is or not. I want to allow the real world to occur. And if, in fact, what we saw in the Cameroon study can be extrapolated so that there is an 87 percent adherence in the open, unblinded arm and an 81 percent adherence in the overall blinded arms, that true difference should be allowed to occur.

This is going to give me a sense in the real world whether or not the benefits that I get from my comparison with placebo from the antimicrobial effects of the microbicide will offset some unintended negative effects that would be associated with the reduction in adherence levels to condom use.

DR. GULICK: Okay--don't worry, I have a lot of people who want to speak, and we'll take them in order. So, everyone who is anxious, I got your names.

Ms. Heise.

MS. HEISE: I'm always the most anxious.

DR. GULICK: You are in good company.

MS. HEISE: Two things. I think that

exactly for the reason that you say, your solution to the problem is wrong, because what you are concerned about is what every public health official is concerned about, which is how will the combination of the biological effect of this product, whatever it may be, interact with behavioral and risk-taking behavior to influence protection or infection rates.

By adding a condom-only arm in this trial, you cannot answer that question because basically, what you are assuming is that that actually does give you a sense of the real world. But when we are counseling women in this trial, we can't tell them anything about the likely effects of this product. In fact, we are spending enormous amounts of time to convince them that they shouldn't have any faith in this gel. And therefore, trying to say that a trial where you are actively trying to dissuade people from relying on a microbicide will approximate people's adherence or risk-taking behavior once we have some evidence that we can counsel that this does reduce risk, I think is false.

I think the way to answer that question is you establish whether or not--you use straight,

placebo-controlled trial--is there some evidence of effectiveness. Then, you do, and I think we are going to have to do, a number of Phase 4, or whatever you want to call them, use effectiveness studies about how this microbicide interacts with all sorts of things in different settings to understand under what circumstances adding it to an existing package of interventions is helpful or not.

But adding on the extra cost, time, and so on of a condom-only arm that is not interpretive doesn't get you where you want to go.

The second thing I want to say is that I think it's actually a shame that the FDA did not invite someone to give data and background on some of the behavioral issues, because they are some of the most important issues. And I would suspect that there is probably not a single one of us around this table who may or may not be an expert in what is known or not known about some of these behavioral issues.

I do think that one thing we do know from the behavioral data--and this is from data from nine studies that have been done, which are reviewed in an article in the Global Campaign

testimony. There have been nine studies done to date that look at how people react when they are randomized to being offered condoms only versus condoms, female condom, diaphragm, or some other combination of multiple methods.

What you find in both those studies where the endpoint are STDs as well as from two decades of contraceptive research is that just the fact of offering choice increases adherence. And in fact in the studies where they were randomizing people between condoms-only and condoms, N-9, or female condom, condom use actually went up because people respond to having choice.

So I think that what we do know is that when you are offering one thing to one group of people and two things, or four options of how they might combine those things, to another group of people, we are likely to have large and probably more than 10 percent difference in behaviors.

So I think that the issues is real. We need a second generation of studies to answer that other question. We first have to convince ourselves, though, that what we can actually say to women that, "If you use this, there will be some reduction in risk."

DR. GULICK: Okay. I have Dr.--

DR. FLEMING: If I could very briefly respond, because she was--

DR. GULICK: Actually, let me stick to the list because a lot of people have been waiting to speak, so let me stick to the list.

DR. FLEMING: Okay.

DR. GULICK: Dr. Haubrich, Englund, Bhore, and Paxton.

Dr. Haubrich.

DR. HAUBRICH: I have to agree with the assessment that the use of microbicide could potentially have a deleterious effect on the overall burden of worldwide HIV cases.

I think that there is little evidence to suggest so far that the use of a microbicide is going to be as effective as condoms. So anything such as the availability of a microbicide in a trial or, even more so, once it is approved, could potentially lead to a reduction in the use of condoms which could have the untoward benefit or the untoward action of leading to a global increase in HIV transmission.

Therefore, I think trials that assess in whatever way we have, no matter whether they are

flawed or not, the impact of no treatment versus use of agents like this are critical.

That being said, I think that the regulatory perspective of showing that a particular agent is better than placebo is really a separate question than understanding the more global impact of the scientific question of how do these agents affect change of behavior, which is really a different question than the efficacy of a particular agent.

So in my view, the sort of two-pronged approach of ongoing studies like the 035 which are targeted to address the sort of clinical strategy, which is really a very different issue and has another whole set of confounders that we have all discussed today, and the regulatory issue of approving a drug should proceed.

I am very concerned--if the 035 and studies like it were not planned, I think that to simply charge ahead and say we need to find out whether microbicides work or not would be flawed, because once one is approved, the impetus and the funding to carry out these large studies like 035 would go away.

So the only way I would be comfortable

with the regulatory allowance of just a two-arm study is the ongoing study like the 035. We talk about allowing Phase 4 studies in this country to answer some of the unanswered questions about ongoing long-term safety and so on, and we talk about how hard these studies are. To blindly think that we are going to carry out Phase 4 studies to answer questions like this once something has been approved I think is a little bit naive.

DR. GULICK: Dr. Englund.

DR. ENGLUND: I just wanted to address two things. Number one, I think it is absolutely important, and some of my colleagues who have done studies--and I have not done studies, but I have worked over in these countries--have to absolutely emphasize is that this condom use is so population-dependent.

In the countries that I have worked on, the women will be killed, stoned, or thrown out of their house if they suggest the condom. When you are dealing with multiple wives with a single husband, these women are totally powerless to use a condom.

So for us to impose on all populations our ideas of what the control group should be is

actually a problematic. So I think first of all, the highest-risk people are the ones that many times are unable to use a condom in a clinical study, or they wouldn't be in a clinical study, and they probably won't be able to use one in practice.

Having said that, I think that makes us forced--and the one thing the FDA can help us do is to make sure that our Phase 2 safety is absolutely flawlessly done. And if that means that in Phase 2, we even have to have a placebo and a non-treatment arm so that we can absolutely assess the colposcopy and all these values before we go on to a Phase 2B or extended thing, that's where we really need to emphasize the safety, because I don't think we can do a 2B or a large study in some of the areas that need us most with condom usage.

I think South Africa might be a great place to do it, but Tanzania is not. It is just going to be very population-dependent.

DR. GULICK: Dr. Bhore.

DR. BHORE: Thank you for the opportunity, finally.

I just want to remind the panel, as I said in my presentation--and I am hearing a number of opinions from a number of people that dropping the

no-treatment or the condom-only arm would be the easiest approach to take--but I just want to remind people that this assumes all along that the placebo--which I put in quotes in my presentation "assuming this is inert"--and the biggest concern is if the placebo is a harmful placebo. And showing superiority of a product over a harmful placebo is not going to be sufficient in showing that it is effective, because at worst, if a placebo is harmful, a product that is superior to placebo can at worst be harmful itself.

So I would just like to remind you about that possibility.

Then, second, I have a question for maybe the statisticians or whoever wants to try to answer this. That is, we have heard a number of people say that one of the reasons for dropping the condom-only or no-treatment arm is the differential compliance of condom use in the three different arms.

My question is are there statistical methods out there that can address this issue of differential compliance rates and so still be able to analyze and interpret the data.

DR. GULICK: Dr. De Gruttola, do you want

to tackle that one?

DR. DE GRUTTOLA: Yes. I would say that there are two issues. One, Tom has made the point that in fact the difference in condom use is one of the things that is important to find out about and the impact of that on the endpoints.

Ms. Heise also made the point that behaviors are going to change once information is actually available regarding the efficacy of the product. But nonetheless I think the information about what happens in the trial with the current state of knowledge is of interest. That is the first point.

The second point is that if you want to ask the question what would have happened had compliance with condoms been the same across different arms even though it wasn't the same, I think that's a hard question to try to formulate because the use of condoms is associated with all sorts of other personal characteristics that may themselves have an impact.

So I think it is a little bit difficult for me to think even exactly about how you might formulate that question. Assuming that you can, there is a whole area of statistics, causal

inference, where people try to address questions like that to try to make adjustments for differences in behaviors in different groups, to try to make some inference about a kind of ideal setting that didn't actually exist, and I think that's an interesting research question, but I wouldn't put a lot of emphasis on it as something that is going to be useful for regulatory purposes right now.

DR. GULICK: Thanks.

Yes?

DR. SUN: This is Greg Sun [phonetic] from the FDA, Environmental Team Leader.

I echo what Victor just said. Essentially, the question of adjusting for compliance may not be relevant for the FDA in the sense that if the drug use is going to modify the behavior of the patients--and I think that is a reality--then there is no sense to look for adjustment, because if by introducing drugs on the market is going to reduce condom use, if that's a reality, then it doesn't matter--even if a drug is active, whatever the benefits may be offset by this less use of condom. Then we're not interested in answering the question if they have the same use of

condoms.

DR. GULICK: Thanks.

Dr. Fleming, to add?

DR. FLEMING: Yes. I would just add that I agree with both comments, that if one wanted to try to step back and make some kind of retrospective adjustments, of course, one of the real problems that we have heard many people state is that the self-reported risk-taking behavior is already just a surrogate for the true risk-taking behavior, and the true risk-taking behavior is in fact also a surrogate for what the actual true risk of transmission is.

So even if we had good statistical methods, it would be extraordinarily difficult to apply them. But I agree with what you are saying here. Ultimately, my interest in comparing to the open label is to look at what is the comparison against a standard of care where it is based on condoms alone, and if that's a different level of exposure to use of condoms, I don't want to adjust that out.

Lori, you make an important point, and your point was is there something a bit artificial about this trial, because we have just gone through

an informed consent process and told people that our best understanding here is that there is equipoise--we don't know for a fact that these interventions, and specifically the microbicide, will be protective.

That in a certain sense is artificial, because once the study is done, if it is proven efficacy, that could lead--you are correct--to a different level of commitment to implement that intervention.

The reality is that that same argument applies to the assessment of the comparison with the placebo as well. That issue that you are raising that could in fact cast some doubt into the generalizability of your conclusion when you are comparing the active microbicide against the open label, in fact I make that same argument all the time about our own placebo control trials.

My answer to that argument is that what we are hoping here is that what you have in a clinical trial setting is actually an artificial intensive oversight of participants to ensure adherence, so that level of oversight is going to offset what you correctly point out could be a level of intrinsic commitment to use an intervention once you have

already shown that it is effective.

But the fundamental bottom line to this is that if you are worried about this point, and hence you are as a result worried about the interpretation of the comparison with the open label, unblinded arm, I can make the same criticism about the interpretation of the comparison with placebo.

DR. GULICK: Dr. Paxton.

DR. PAXTON: Actually, one of the major points I was going to bring up was brought up by Lori. But one more minor thing is that I think your contention about whether we can say that the condom-only arm really does approximate the real world, because in no sense actually is this the real world in that these women will be getting intensive condom counseling repeatedly each time they come in, which doesn't happen in the real world. And then we have that other confounding thing about when you have somebody coming in and getting condom counseling and you ask them, "How are you using the condoms?" they might tell you what you want to hear, or they might be telling you the truth, and we have no way of knowing that given our present assessment measures for this.

So I just would not say that in any sense this approximates the real world. It might be of interest, and I do think it is of interest, to look at these things, but I don't in any way in my mind consider it a proxy for the real world.

DR. FLEMING: But Lynn, these issues are very parallel. The extent to which you are legitimately recognizing that our intent to do a real world comparison can't be fully achieved, you have got to look at the comparison against placebo in the same way. The blinding issue doesn't get rid of that particular concern--that is, what you can state is the generalizability of the efficacy that you get from a blinded comparison is also sensitive to issues of how well was there adherence in that specific setting to the condoms, how well was there adherence to the intervention.

DR. PAXTON: Can I respond?

DR. GULICK: Sure. Response.

DR. PAXTON: Just in response, I do think that when you are looking at two arms that are both using a gel, you are going to have less variability between those two arms in terms of behavior.

DR. FLEMING: But that's okay. The fact of the matter is that the adherence to the

microbicide gel in the placebo arm isn't my issue. I assume that is inert; I am hoping that is inert. My concern is what is the adherence. My biggest concern with microbicides, my biggest uncertainty of their efficacy is that unlike a vaccine that I can deliver once or on a periodic basis and be assured I have continuing, sustained adherence, I have got to use this microbicide on a regular basis to achieve the full essence of the benefit.

And Lori is right--if in fact I don't have the same commitment to that implementation when I haven't already been aware that it has proven efficacy, then, randomization hasn't protected me against that level of underestimation of efficacy as well, even in my comparison against placebo.

DR. GULICK: Okay. We are going to need to draw this important discussion to a close, but Dr. Stanley, you have the last word.

DR. STANLEY: Well, good, because that's about what I was going to say. That is, what we are really doing is dancing around the ethical conundrum that microbicides bring to us.

There are two populations of folks out there at a minimum. One is folks who are going to use condoms, who have the authority, if you will,

to mandate that their partners use condoms, and they don't necessarily need microbicides to the level that we are talking about.

But then you have the other disempowered population that cannot mandate condoms, and those are the ones we feel an urgency to have an effective microbicide out there--and it doesn't matter if it is only 20 or 30 percent effective, because they don't have another option.

The problem is that once you approve one of these and put it on the market, the group that has been able to use condoms will alter their behavior or some subset of that group will, and that's where you stand the risk of doing harm.

So, while you have done good for one population, you run the risk of doing harm to the other, and it is that ethical conundrum that then causes us--we are trying to design clinical trial designs that aren't going to answer that.

DR. GULICK: Okay.

Dr. Fletcher, you have the last-last word.

DR. FLETCHER: Mine is just a quick question for the FDA in terms of where we really are with a microbicide placebo. Is there a candidate product? Is there one in testing? Just

give me some sense of where that universal placebo development is at.

DR. WU: That so-called universal placebo is going to undertake a Phase 1 14-day trial as a safety assessment initially.

DR. GULICK: And is there a plan--this is a bit of a funny question--to go to a Phase 2-type design with the universal placebo versus no intervention?

DR. WU: Not at the present. At the present, once after that 14-day trial, the placebo will be used concurrently with a candidate microbicide into whatever the design, the next step will take them. If this is Phase 2 running to Phase 3, this placebo will be in use.

DR. GULICK: Okay. Let me try to summarize what we are thinking here.

Clearly, we have differences of opinion around the table. Dr. Fleming put it best to say there is no one right answer here as well. We are dealing, of course, with different cultures, different countries, where there is lots of different condom use, and that complicates our discussion of what the standard is even from population to population.

We recognize again the inherent issues about clinical trials and how they are different from life, and specifically here that making an intervention may change behavior, that a commitment to an intervention may also change behavior, and that intensive counseling which is critical for these studies actually is not often a part of what happens in the "real world."

These are all issues of generalizability and how you take one study and apply it to the whole world, but that's really what we are talking about here. Also, the recognition that sexual behavior is difficult to assess in a clinical trial or really in any setting at all.

We took some comfort in knowing that our recommendation for which design is optimal now may be the most appropriate for the initial studies, but then, when information is generated in these studies, other design could be considered, particularly simpler or, if some of the questions that we have been struggling with are answered, then a more complicated design would not necessarily need to be continued.

There was some debate about that, though, whether it is more appropriate to try to answer

these questions up front or limit the questions up front and then answer other questions in Phase 4 or down the road, and there were some differences of opinion on that.

And clearly, everything changes when one microbicide shows safety and efficacy, because then that would be the standard to compare all future microbicides to. So a lot of our discussion becomes less important when that event occurs.

As we heard earlier today, a requirement versus allowing a design--there was a lot of support for flexibility in both approaches, really.

So what did we say in all? The most attractive thing about the three-arm design is really that it gives you an overall net benefit. We are looking for benefit versus risk, antiviral effect versus the possibility that an intervention could actually change behavior or reduce condom use, and both of those are important in assessing the overall risk versus benefit.

As Dr. Fleming reminded us, the amount of effect that we are looking for here is quite different than we are looking for in, for instance, a vaccine study, so that small benefits in antiviral effect actually could be offset by

changes in behavior on the order of what we have been talking about. So that is a big concern, I think, around the table.

Using this three-arm study, the comparisons of the two arms actually give you different information, which was stated again and again. There are really two questions--how does a microbicide compare to the placebo asks a very different question than how does a microbicide compare to no intervention at all.

Safety was something that we had not talked a lot about, but Dr. Englund reminded us that safety is important here, both of the microbicide and the placebo itself, and we need to keep that in mind.

So people had concerns actually about all three of these designs. There were concerns voiced. On the two-arm versus the placebo, which you might think of as the efficacy comparison in that you are looking for antiviral effect above and beyond behaviors which we would like to think would be randomly distributed between two arms, is attractive; however, we are not convinced that the placebo is inert. It could have beneficial properties such as barrier or lubrication, or on

the other hand, it could actually be harmful, and we may not know enough about the placebo--I think that is what prompted Dr. Fletcher's late question--how much do we know about the placebo before we go into this.

Then, there is a big concern that just the use of any intervention here could decrease the use of condoms, and how do we evaluate that, and then, conversely, that's an important part of evaluating this kind of intervention in and of itself.

There were lots of concerns about the no-treatment arm. This is more of an effectiveness evaluation, in a sense. This is real world--or is it? There was a lot of debate about that, and I won't review that, but there is controversy about how real world this really is.

People noted again that it is difficult to evaluate behaviors or changes in behaviors. And there was a big concern that post-randomization, there would be different behaviors in the different arms, and condom use could go up or down and you really can't guess which might occur in each of the three arms, and that there might be a significant enough difference that it could actually affect the overall interpretation of the study. There were

lots of concerns about that.

So in summary, we're not sure.

[Laughter.]

DR. GULICK: But all approaches have value, and I guess--we talked about taking a vote on this before. I think that would go down in flames, so I don't think we'll do that. You heard our pros and cons, and I guess if I had to reach consensus from the vibes I am feeling right now, generally, I think that what people liked was a broader approach earlier on and then a quick answering of some of these questions and then focusing on a two-arm design may be more appropriate after some initial information. And I know there are differences of opinion about that.

Okay. How are we doing?

DR. BIRNKRANT: Okay. That was helpful.

Well, Question 3 is specific to the three-arm trial design, and even though not everyone favors that, perhaps we could get some opinions on FDA's definition of a "win"--that is, the microbicide arm has to show significantly better reduction in seroconversion rates compared to both placebo and the no-treatment arms. However, if Dr. Fleming could reiterate his

proposal from this morning, and that is having different P values for the various comparisons, that may also help the discussion here.

DR. FLEMING: As I mentioned this morning, I think the FDA has given a great amount of consideration in recent times to this concept of recognizing the importance for flexibility in certain settings to allow approvals on single trials. And as we were saying, this setting that we are in here certainly does seem to be within the mainstream of what the FDA has considered in the past to be such a setting--a setting where you have a compelling endpoint in settings where it is very resource-intensive to be able to do multiple trials.

What I have noted through numerous discussions across the wide array of situations with FDA is that there seems to be a very common aspect of how they characterize this. The results must be "robust and compelling."

I also respect why the FDA is reluctant to say what that P value is because any assessment of strength of evidence has to be a global assessment and has to factor in all issues that are relevant to understanding benefit to risk.

My general sense that I tried to characterize this morning, and I think it seems consistent with what I have heard from the FDA, is something that is basically a middle ground between the strength of evidence of one trial and the strength of evidence of two trials in such settings where you have such a compelling unmet need and very significant clinical endpoints would be an appropriate target, and that would be, then, something, as we have said, on the order of one-sided .0025 to .05 or a two-sided P value slightly lower than .01. But again, obviously, that will then depend on the nature of the totality of the data.

What I had mentioned this morning is in this two-arm trial, one strategy that I would think would be very consistent with that FDA philosophy would be to require that robust and compelling set of evidence against one of these two comparisons, so that one of them would have to be compelling, the other would have to be supportive, specifically being that if there were compelling evidence of the difference against the placebo, it wouldn't have to also be compelling. It would just have to be supportive that the comparison against the open

label was suggestive also of favorable effects--and vice versa, I would also think.

So essentially, my own sense is that that would incorporate basically what has been an FDA philosophy in other settings, I think, in a manner that would be consistently implemented in this setting.

DR. GULICK: Dr. Paxton.

DR. PAXTON: A question for clarification. Does the FDA's definition of "robust and compelling evidence" also include things like animal studies or a stand-alone Phase 2 that looked very promising?

DR. BIRNKRANT: It would be less likely to include the animal studies. We actually need the clinical data to make our decision in this setting.

DR. GULICK: Other comments on this point?

Dr. Mathews.

DR. MATHEWS: The rationale for requiring a more rigid P value for the single trial as I understand it is to minimize the chance in a single trial that the outcome would be observed by chance alone. But the problem that we have been dealing with all day has not a lot to do with random events or chance. It is differential effects of behavior

that could trump any statistical variation between the arms due to chance alone.

So in some ways, I don't understand the agency's rationale. It is almost as though you are saying that if the effect size is above a certain threshold, you think that any systematic biases that might be in that trial would be trumped by the higher precision of the estimate. And I think somebody earlier this morning, I think even Tom, made this point, that if you have a systematic bias, and you estimate it more precisely, you still have that bias. And if condoms are so much more effective than a microbicide which is actually being developed because people are not using condoms, then I'm not sure that requiring a smaller P value addresses that limitation, post-randomization changes.

DR. GULICK: Dr. Haubrich.

DR. HAUBRICH: Just to follow up on Chris' point, I think there may be a couple of issues here that we are combining. One is the need for one trial versus two, and the other is the statistical comparisons of the three-arm study. I am going to just comment on the three-arm study.

I would be a little afraid of requiring

rigorous comparisons of both the placebo arm and the no-treatment arm, and I would agree with something where if you were clearly better than the placebo arm and not worse than the no-treatment arm, that would be acceptable; but to require the hurdle of being highly statistically significantly better than both would be unreasonable.

To some extent, then, if you are not worse than the no-treatment arm, you have gotten rid of the problem of what is the effect of reducing condom use having on it, so if you are better than placebo and nor worse than the no-treatment arm, that in my mind would satisfy the requirements.

DR. GULICK: Ms. Heise.

MS. HEISE: I guess I just want to go on record and say that this is actually the most important decision that is being discussed today, and I fundamentally disagree with the concept of having to be better than both.

I think that that is a standard that, one, I think is uninterpretable, and I think that also again, this issue of how it is going to act--as a health advocate, I would give up the possibility of having a single trial to avoid this, because I am actually more concerned that we are never going to

be able to generalize to all of the settings. Behavior is so driving of how this is going to operate in different settings that if you showed me a trial with convincing evidence for sex workers, I would not be convinced of how that is going to operate in Tanzania with married women. I would want to see, if I were a regulator, even if it is a smaller trial, or it is an introduction study, or it is something--I think we cannot generalize to many of the settings that we want to generalize in, so I almost think we want more trials. And I think our hope that we are going to get it in a single answer is the chimera that is going to drive us crazy.

And I fundamentally think that the issue of how this operates and combines with behavior in real life settings, as well as underlying STD and HIV rates--you know, depending on whether or not this microbicide is also effective against certain STDs, will interact in different settings with the effectiveness achieved.

So I think that we are kidding ourselves in terms of thinking that adding this one arm in one study in one population is going to really address the use-effectiveness questions that are

very real and we need to deal with, but I think we are setting up a standard that stops us from being able to mount those next phase trials because we don't even have anything that we can say works to start to do the behavioral work and figure out how to introduce it so those things do not happen.

The last thing I want to say is that I think this issue of condom migration is very important. I suggest, though, that people look at some work that the London School of Hygiene has done that has been published in AIDS about modeling of these various different scenarios. What they have done is looked at the tradeoffs--because condoms are very, very efficacious; they reduce risk very well if they are used. But we have tons and tons and tons of studies around the world showing that inconsistent condom use confers very little protection in many populations, and we have tons and tons and tons of studies showing that most people use condoms inconsistently.

So this notion that the condom is so great--we also have to think about the number of people we are recruiting who are doing nothing to doing something, and when you look at those tradeoffs even on the individual risk level in

these models, what you see is that you don't even have to worry about migration unless you are at the level of 80 percent consistent condom use. Then, you have to worry about how good your microbicide is or whatever. But up to there, you could almost have total migration. If you could have something that is 30 percent efficacious used 60 percent of the time, it buys you more protection on an individual basis, not even on a population basis, than something that is 90 percent effective that is used 30 percent of the time.

So I think we have to be really careful when we make these judgments about tradeoffs even at an individual level.

DR. GULICK: Dr. Bhore.

DR. BHORE: I'd like to address the point about this win against both arms. In my presentation, I mentioned the alternative possibility of showing evidence of a single trial with evidence worth less than two trials--for example, evidence worth one-and-a-half trials. So that is an example where, as Dr. Fleming mentioned, one could have two different types of criteria for a win against the two control arm.

One arm, for example, could show

compelling evidence, and the other arm could show less than compelling evidence.

So in the example of the evidence worth one-and-a-half trials, a P value would be less than .008, which is slightly higher than what I mentioned, .001, but in that case, you could have two possibilities--both arms show an equal amount of evidence, or one arm shows more compelling than the other one.

So there are these kinds of alternative possibilities that one can look at.

And then, secondly, the topic of condom migration keeps coming back again and again, and if one were to have just two arms, the microbicide and the placebo, and here, supposedly, Lori mentioned that if such a trial is designed, then a participant would be strongly informed that we don't know anything about the activity of the gel right now, so condom use is very, very strongly encouraged.

If that kind of message is given to a participant, then that raises a question in my mind: Would that affect the enrollment? Would the participant just run away and say, "You just cannot tell me anything about the activity of whichever

product I am getting, so why should I be staying in this trial?"

So again, this issue also ties in with the three-arm design issue. I just wanted to bring that up.

DR. GULICK: Let me try to focus us because the hour is getting late, and these are important points, but I'd like to get us back to the question at hand.

So we have covered a lot of ground, and clearly there are differences of opinion around the table that we have not resolved, so they are going to continue to be. But the question that we are being specifically asked is if we accept the three-arm studies--and we have to take that as a given--how do we compare the two arms, and what kind of reductions are we looking at for both pair-wise comparisons.

And Dr. Fleming proposed "compelling" for one of the comparisons and "supportive" for the other comparison, and then Dr. Haubrich got more specific and said "compelling against the placebo," meaning a high degree of statistical significance, and "supportive" being defined as "not worse than the no-treatment arm at all."

Is that a consensus?

Dr. Sherman.

DR. SHERMAN: I just want to say that I don't think you can answer this question in a vacuum without taking into account is there going to be a separate and highly supportive Phase 2 trial and what are the P values that you accept. They are all tied to the same thing. If there was a very supportive Phase 2 trial, then you could be more generous in your P values and be more allowing in terms of the comparisons in your groups here.

If you are going with a single trial, then you might go with higher P values and be stricter in the requirements that are going to be used here. And on the front end, a sponsor might discuss this and negotiate what set of conditions would be acceptable to the agency, because this question really cannot be separated from those other things.

DR. GULICK: I think that's a good point, but we are not asked to come up with specific P values in this question--and you are right, it could be different at different times, but we use the word "compelling" to say some high degree of statistical significance, Richard's suggestion, versus the placebo arm versus not worse than the

no-treatment arm. That seems to be what we are migrating toward.

Dr. Flores.

DR. FLORES: I think in addition to this [inaudible] P value that has been discussed, the other worry that I'm sure is in the minds of everyone and that hasn't been mentioned is the issue of compliance, because it is truly going to be much harder to ascertain compliance in that third arm.

Therefore, perhaps not just because of the comparison level that we are trying to establish here, but because of the potential for that arm not to have the same level of compliance, that might sink the entire study.

Now, if you determine at the end of the day that, yes, the two active arms, meaning two placebo or other two study arms, versus the non-intervention arm, they might be okay in terms of compliance, because women may be more enticed, if they think they are receiving some benefit, to continue on, but that third arm where they are getting nothing is going to be a challenge to maintain at the same level.

DR. GULICK: Well, again, I would say a

priori you cannot predict which way adherence would go in that arm. It could go down or it could go up because women are not receiving something and they know they are not receiving something. But let's not revisit that at this point.

Have we addressed this question to your satisfaction?

DR. BIRNKRANT: I think so, but I also think that we have rolled in Question 5 with regard to discussion of the P value--

DR. GULICK: We have.

DR. BIRNKRANT: --so that's good; I don't think we have to spent more time on that.

But what I'd like to spend more time on and get the Committee's input is in the area of what other supportive evidence should we have. It is part of Question 5--but if we go with the approach where we have compelling evidence against one arm, that is, against the placebo, and it is not worse than no treatment, what other data should we have along with this approach?

DR. GULICK: Okay. So essentially, we have lumped Questions 3 and 5 together in our discussion.

DR. BIRNKRANT: Right.

DR. GULICK: And you would like us to focus on the last part of Question 5.

DR. BIRNKRANT: Right, and specifically but not limited to are there other STIs that could serve--that is, reduction of transmission of other STIs that could serve as supportive evidence, because we are frequently asked this question.

DR. GULICK: Dr. Paxton.

DR. PAXTON: It seems that that would be highly dependent on what product you are testing. For example, if you are looking at a highly specific product like an NNRTI, you wouldn't expect it would have any efficacy against STIs; whereas if you are looking at something that is more broad-based, yes, again, I think this is going to be a highly product-dependent decision.

DR. GULICK: Other suggestions about other supportive evidence in this case?

DR. HAUBRICH: I guess it does raise the conundrum that if you have a product that theoretically has broad activity, and it shows reduction in HIV but fails to show reduction of other STIs, that might fall in the category of being negative supportive evidence, because theoretically, if the combination of biologic plus

behavioral things leads to a reduction in HIV, you would suppose that you would have reductions in others as well. So that might be a bit of a conundrum.

DR. GULICK: Although I suppose it depends on the mechanism of action, if it is a physical barrier, or is this something specific to viruses?

Dr. Paxton.

DR. PAXTON: I just wanted to respond. I think, yes, it wouldn't be as desirable to have something that is useful against both, but frankly, if you offered me something that was effective against HIV and said, "but it's not going to be effective against gonorrhea," I would say fine, give me penicillin.

DR. HAUBRICH: No. What I meant was if the agent theoretically had activity against the STD, so it was broadly in the test tube active against all of the agents or several agents yet failed to protect against the some but did protect against HIV, I think that would make me scratch my head.

DR. GULICK: Well, and interesting--the COL 1492 study, as you mentioned earlier, showed no differences among secondary endpoints which were

STI occurrences.

One thing that seems obvious for supportive evidence is behavioral information, although fraught with peril, and how do you collect this most effectively, and those conversations came up earlier today. But I would suppose that some data is better than nothing, at least to try to get a handle on what condom use is doing on the three arms, for example.

Other supportive information that we would suggest?

[Pause.]

DR. GULICK: Okay. So we'll turn to our last--yes, Dr. Fleming.

DR. FLEMING: I wanted to wait to make sure there weren't any more comments on that. Since I didn't realize we were actually fully addressing Question 5 when we answered Question 3, I would at least like to make a brief comment about the second-to-last sentence in Question 5, which was specifically asking us about a strength of evidence issue.

I think it is worth at least pointing out that since in the open session, there was a comment made about the ethics are appalling, that we could

consider a necessary strength of evidence on the order of two adequate and well-controlled trials or .025 squared is to say that the FDA has enormous experience, and through that experience, there have been a plethora of examples where an initial trial that might provide evidence at roughly a one-sided .025 level in fact has not been confirmed, i.e., the concept of the value of replication in clinical trials science I think has strongly been established by the experiences that FDA has seen, and as a result, that does need to be considered seriously if we are going to go with a single trial; what is that strength of evidence.

It is worthy of at least just reiterating why this is important, and that is it surely is true we want to get timely access to promising interventions, but it is also important to avoid an unacceptable level of false-positive conclusions. It was once said it isn't so much what we don't know that can get us into trouble; it's what we think we know that isn't so.

I just gave an example this morning of the 5-FU levamisole and levamisole alone experiences in a trial in a very compelling situation, a life-threatening disease situation, that talked

about reducing mortality by 33 percent. Was it proper to do a confirmatory trial there? If there hadn't been 5-FU levamisole and levamisole would be out there, levamisole might be a very attractive regimen because it is much less toxic. Yet it provided no benefit, a false-positive conclusion. If we have multiple microbicides out there, we want to protect women. It is important for us not to put a microbicide out on the market if one that is out there is highly effective and another one is not effective.

Furthermore, to in fact be using an ineffective microbicide that might in fact even lead to or be associated with reduced condom adherence would also be very negative.

So I think the balancing issue that has to be kept in mind here is that it is a serious problem to be in fact judging something to have been established when in fact it hasn't been reliably established. And I won't go into a lot of examples that I have written down here, but there are many examples where a single positive trial at just the strength of evidence of one-sided .025 hasn't been validated.

So I think there is real wisdom in the FDA

asking for "robust and compelling" evidence if it is based on a single trial.

DR. GULICK: Yes, Dr. Mathews.

DR. MATHEWS: I think this discussion is framed around an assumption that Dr. Birnkrant made in her opening remarks that the risk-benefit ratio should be the same across the world, if I understood you correctly, and I don't think I agree with that, because we are dealing with vastly different incidence rates of disease in this country compared to the countries where the need is greatest.

If I were a decisionmaker in a country where one out of three people had HIV infection, I would be willing to take more risk in terms of the levels of confidence and the effectiveness of a particular intervention than I might be in a country like this one, where the risks are lower.

I mean, ideally, what you are saying is true, but the urgency of the threat is very different. In some ways, it is kind of odd that we are even talking about this in an American setting, because this is not where most of the need is, and whatever guidelines we set up for this country surely--I mean, the people from WHO who have been

dealing in these other countries have a very different perspective on what the needs are.

What does it really mean if a product gets licensed in the United States for this indication in terms of what will be done in Sub-Saharan Africa?

DR. BIRNKRANT: That's the sponsor's choice, though, whether or not to submit the marketing application to the United States or not. Once it is submitted to the United States, it has to meet the Code of Federal Regulations, and if it is approved, then clearly, American women will be entitled to use the product, so therefore the standards are what the standards are.

DR. MATHEWS: Right, but we are making recommendations based on our conditions in this country, and I'm not sure that they necessarily apply, particularly if the standard set by the FDA is expected to be implemented in the developing world, as you implied it should be.

DR. GULICK: Isn't it true, though, that many countries around the world actually look to the FDA and their decisions in evaluating this and then take those recommendations back to their own countries in terms of accessing drugs, so the

standards and the approval of the FDA really does carry a lot of weight all over the world.

DR. BIRNKRANT: And we are also told that some countries rely solely on an FDA approval, that they don't have the regulatory bodies to do the type of work that we do.

But I understand what you are saying. We are having data come in that are generated outside the United States and may likely have a greater benefit there, but nonetheless it is subject to U.S. regulations.

DR. GULICK: Dr. Haubrich.

DR. HAUBRICH: Although I agree completely with Chris' assessment that the risk-benefit ratio is very different from country to country and that some countries might be willing to accept a greater risk potentially, I think that here, we are talking about the likelihood of finding a false-positive. And if we are willing to accept a single study versus two studies, I think the bar has to be higher, because if we accepted something as being efficacious, and it was truly a false-positive, we wouldn't be helping anybody.

DR. GULICK: Okay. Other comments?

Yes, Dr. Stek.

DR. STEK: In the discussion, there was mention made of what if we find that there is some efficacy of the microbicide intervention, but it is actually less than regular condom use. I don't think that is an appropriate thing to actually discuss. Our goal here is to determine whether the microbicides are safe and effective, not which is the most effective intervention. That is information that should be made available to everybody here and internationally to make the decisions that are appropriate for the local setting.

So I was a little disturbed by the thought of the ethical issue of perhaps approving something that might not be as effective as something else that is already available. I think the goal should be to increase the options.

DR. GULICK: Well, and as we have been reminded by the agency before, it is to demonstrate safety and activity, not necessarily better than something else in many cases, but in this case, the standard of care is condoms are the best you can do essentially, isn't it.

Okay. Let's move to Question 4, back to Question 4, which talks about duration of

follow-up.

DR. BIRNKRANT: Both on-treatment and off-treatment.

We have received proposals that either call for 12 months for every participant, and that's it, they are finished; or the other approach is 12 to 24 months when the last participant is enrolled.

Now, for treatment trials, although we look at 24-week data, we also look at longer-term data, that is, 24 weeks from when the last patient is enrolled in the clinical trial, and that is for treatment of HIV.

This is prevention, and we were wondering what the committee thought about having a fixed period of time--for example, 12 months--versus 12 to 24 months when the last patient is enrolled. That would give us extensive data with the product, both for efficacy, safety, and durability of effect.

DR. GULICK: So let's consider duration of follow-up generally, and then we can see if we like one of these choices more than the other.

Dr. Stanley.

DR. STANLEY: Well, I think we have to be

realistic with where these studies are going to be done, and the comparison with the American MSM population having intensive behavior modification intervention is not a valid comparison in any way, shape, or form in my view. These populations are migratory, transient to some extent. Getting them to stick for a year is going to be a challenge and probably a target that we ought to target, and if you can keep them longer, that's great, but I think it is unrealistic to set the bar that high.

We have to look at the examples that we have had, and in the N-9 experience, it was 48 weeks, and there was 81 percent follow-up, which means that you had almost 20 percent who dropped out. That is real life experience, and that is with experienced researchers who know what they are doing in this setting.

So I think you can be permissive in trying to get longer follow-up and trying to allow for that, but if you set the bar too high, you are not going to be able to enroll folks.

I think the other issue is that even if people have enrolled, and as the comment has been made, they tend to drop off in the first few months, and then they stick it out--if you came to

me and said, "I want you to enroll in this study, and I'm going to follow you for 2 or 3 years," that's a disincentive to me to even enroll to start with, and I think it would be a real disincentive for some of the populations we are looking at in Sub-Saharan Africa because they can't guarantee to you that--if they enroll, they may be doing it with the knowledge that they are misleading you, because they probably won't be there in the same place possibly.

So I think you can't set the bar too high here, or you are going to hurt yourself and hurt the ability of sponsors to conduct the studies.

DR. GULICK: Can I just clarify this point again? The follow-up 81 percent actually refers to the COL-1492 study, but it's not that those patients were lost to follow-up; it's that they came off drug. Is that right? They discontinued--well, let me not guess. Can you tell us again, Dr. Van Damme?

DR. VAN DAMME: In COL-1492, the study done and finished, the retention after one year was indeed 81 percent, and that's indeed people who were still in the study after one year. That is not all the other people who were not lost. But

that was also open, so people could come into the study and stay as long as they wanted. But based on experience, I would recommend a short follow-up.

DR. GULICK: Okay. Again, I'm not sure I understood. Let me ask you one more time. So those are people who were lost to follow-up after one year, 19 percent, or simply discontinued study treatment and were continuing in follow-up?

DR. VAN DAMME: Yes, that could also be.

DR. GULICK: Which one?

DR. VAN DAMME: Well, both. I mean, there were people that we really lost, and there were people that discontinued.

DR. GULICK: Do you know the exact figure of people who were actually completely lost to follow-up after one year? I think that would be helpful for us, because someone who reaches an endpoint or becomes pregnant but is still being actively followed doesn't--they are not really lost to follow-up. They are still in study follow-up even though they have completed their participation.

Dr. Haubrich.

DR. HAUBRICH: We have heard proposals today to have studies that have very short

follow-up because of retention issues, and there is always a balance in clinical trial between what is going to happen with differential dropout and how that affects the interpretation of your results and wanting to have more data, to collect more endpoints, and to have better safety data.

I would argue that if we approve something based on 6 months' follow-up which we are asking hundreds of millions of people to take for the rest of their lives, I would be a little concerned about that.

In fact, if people are dropping out, it may be telling you something. Of course, you may not be able to discern what it is telling you, but I think we should have trials that are of adequate duration to evaluate safety concerns as well as efficacy. So anything shorter than 12 months I think would be problematic, and the longer, the better, as far as I'm concerned, although arguably, that then increases the risk of having dropouts--although I think we did hear that in other trials that looked at these, many of the dropouts did occur early, and that once you had crossed a certain threshold, the follow-up was good. So that would actually argue for at least 12

months if not longer follow-up which I would advocate.

DR. GULICK: Dr. Paxton and then Dr. Wood.

DR. PAXTON: I Just wanted to follow up again. I can't remember--I think, Tom, you were talking about that study--was that again an MSM? I think it is an important thing that we are talking about an African population of heterosexual women, and I don't know that we can say that the ones who stay are going to have the same number of events. I think it has been our experience in some of the trials that we have done at CDC that the people who stay tend to be the ones who are more compliant, are more interested in their health, they use the condoms and all that.

That would factor into my recommendation because I think we are trying to balance the effort that goes into keeping somebody, because it is an enormous thing to follow someone for 2 or 3 years, and if they aren't going to really contribute much in terms of events, then I think that might argue for having a shorter follow-up time.

DR. GULICK: Dr. Fleming.

DR. FLEMING: Yes. I was giving that .015 example just specifically to refer to a prevention

trial setting where it came up in the context of we don't have data on whether or not an open arm could be followed, and it is just intriguing to see in that 4,000-person prevention study for HIV prevention that we actually had a higher retention rate in the open-label control arm.

I would certainly agree that what I'd like to do most specifically is look at settings as close as possible to the settings that we have here. And we were given data that was reported in JAMA '02 for the Cameroon study of N-9 gel against a no-treatment condom where there were very high levels of retention.

The 012 trial conducted in Uganda I think is another very relevant experience. In Uganda, before we launched that trial for prevention of transmission from mother to child, we were told things that you might hear now--it's just not realistic to think you're going to retain 80 percent. We were told it's just not going to be possible. Women go up-country; they are just not going to be able to be tracked with their infants.

Efforts were made to have high levels of retention in that trial, and at the primary endpoint of 3 months, there was 98 percent

retention. At 18 months, at the final analysis, there was 95 percent retention. We were told that we couldn't do better than 80 percent; there was 95 percent retention.

Somebody said earlier that "quality trumps quantity," and I would agree with that. I think it comes back to a question that Victor had asked earlier about what is the risk of bias when you have more people missing than you specifically have events.

I would actually rather have a study that was somewhat smaller where intensive efforts were made to obtain reliable, interpretable results because we have high levels of retention.

It is possible--it is possible--to do better than one might think by putting specific energies and efforts into achieving high levels of retention, and it is extremely important to do so.

I would like to jump on, though, and reinforce something that I think Richard had said, and that is what drives me to think more than anything else about what is the right duration of follow-up is what is the clinical question; is this an acute setting or is it a chronic setting?

To my knowledge, this is a chronic

setting. This is not a situation where we have to identify an intervention that is going to get a woman through a 2- or 3-month at-risk period, and then she's going to be risk-free.

When in fact you envision delivering an intervention in a chronic setting, it becomes even more important to obtain results that in fact are as relevant as possible in a practical fashion for that overall time period.

My own view is that participation in clinical trials, whether you are in a control arm or the active arm, generally provides benefit to people, not just because of the altruistic aspect of contributing to understanding benefit to risk, but because overall level of care generally is at the highest level of what would be achievable.

People are getting very high levels of attention compared to normal care. So if somebody in fact is followed for an extra period, let's say, 24 months, is that a burden or is that in fact a privilege that this person is in fact in a circumstance where they are going to be getting just that much more attention to their care and to their needs over a longer period of time?

And as Richard pointed out, I do want to

know about safety issues, I do want to know about efficacy issues. Some people have said maybe adherence wanes. If adherence wanes, isn't that relevant to understanding what the actual protection of the intervention is going to be over a chronic risk period?

There has to be a practical tradeoff here, but surely I would strongly support the point that some people have made that quality trumps quantity. I would rather see a high-quality study that achieves interpretable, unbiased results, minimize loss to follow-up. At a minimum, I would like to see 12 months of follow-up, although I would be delighted to see trials run to 24 months of follow-up if in fact we could achieve that.

DR. GULICK: Dr. Mathews.

DR. MATHEWS: On the issue of whether there should be a fixed follow-up or until the trial ends, my sentiment is that it should be fixed, because the people who are continuing until the trial ends are probably going to be different than the ones that have dropped out, and the sample size in that group is going to be smaller, and if adherence wanes, that effect alone will just attenuate whatever the effect of the intervention

is.

And the point that Tom just made about wanting to know about whether adherence wanes, what's the long-term impact of this intervention, again, I think that's an effectiveness question, and if the purpose of the trial is to establish that you have an active intervention and to precisely estimate it, then I think the population study should be as similar as possible throughout the trial, and that implies that their duration of follow-up should be similar.

DR. GULICK: Dr. De Gruttola.

DR. DE GRUTTOLA: Yes, just to respond to that point, I think that there can be value in continuing to follow patients. Let's say you look at the options of doing a 12-month follow-up on each patient versus following them to the end of the study, where you have at least 12 months on each patient. Those studies are going to take the same amount of time. But if you follow all of the patients longer, you'll get additional information, and it can be safety information as well.

I think the point you raise, that as time goes on, you are going to have more dropouts, so your population is in a sense increasingly

self-selected, is true, but I think finding out about that dropout and about the acceptability is important as well.

So I think if you are going to be taking the same amount of time to do two studies, you can only gain by having the additional information about safety, tolerability, and about efficacy, taking into account your point that you do have to be concerned about the dropout and selection issues.

But I think that the implication of that is that a lot of effort has to be put into retaining patients for the longer haul, and whatever creative strategies can be developed would be important to avoid selection bias.

DR. GULICK: Dr. Wood.

DR. WOOD: In examining studies, the issue of duration, one of the reasons is not only to look for safety but efficacy as well. So your ability to detect event rates is either going to be determined not only by the duration of follow-up but also by the sample size.

So on the one hand, I understand the need because of issues of retention and concerns about dropouts, the desire to have shorter-duration

studies. I would maintain that if there were going to be shorter-duration studies that were less than 12 months that there would be an appropriate requirement for a larger sample size to allow you to have an adequate detection of events that you would lose since you are observing the population for a shorter period of time.

On the other hand, I have got to agree with the fact that we are talking about potentially approving a product that would be used by women potentially by the rest of their lives. So the issue of longer-term safety and adverse events diminishing efficacy over time, whether that is behavioral, whether or not depending on the product it is related to the development resistance, but say with the NNRTI microbicide candidates, would be very critical to ascertain.

The other point that I would like to raise in terms of Phase 4 follow-up that is done post-marketing is that for the most part, what happens is that we always hear about what goes wrong and when something is bad in terms of safety. What we really hear from Phase 4 marketing studies is that people's livers are being killed, they are dying from the drug, there are unanticipated

toxicities.

So anticipating to get additional information from that type of Phase 4 mechanism I think is unlikely.

DR. GULICK: Dr. Stek.

DR. STEK: Just to point out that continuing follow-up for a long time to assess adherence doesn't seem to make a lot of sense, because adherence to an experimental regimen that you don't know if it is efficacious or not, you wouldn't expect that to be comparable to adherence in real life to a product that was shown to be efficacious. So that would argue against following for a long time for that purpose.

DR. GULICK: Do we want to entertain some of the specific choices that we have here? There was a proposal that anything less than 12 months would not be acceptable. Is there general agreement about that? We heard earlier suggestions about 6 and 9 months.

Dr. Fletcher.

DR. FLETCHER: On that, I think one of the themes that we have heard today is flexibility and what a sponsor may approach the agency with. And I guess on the issue, then, of duration, I wonder if

there is not an opportunity for flexibility.

Let me try floating this and see how it goes. What if a sponsor came to the agency and said, "We are willing to do two pivotal studies," two traditional Phase 3 studies, "but we would like the first one to be of 6 months' duration to try to get an early answer of efficacy, and then, if that is present, we'll do a long-term Phase 3 study. Might that be an acceptable approach?"

In my mind, I could think about buying something like that, so therefore, walking in, everything has to be 12 months in some settings might be inflexible.

DR. GULICK: Dr. Stanley.

DR. STANLEY: I want to echo the flexibility issue, because I think again, we are trying to balance the sense of urgency to get something out to women who have nothing else, and every day, 16,000 people are getting infected. But we also have to balance that with a responsibility to first do no harm.

So I think, as I said earlier, that 12 months is probably a good length, but I think there may be circumstances where a 6-month or 9-month trial might in fact be justifiable, particularly if

the sponsor is willing to commit to a Phase 4 to look at longer-term use.

Again, we talk about adherence, but this is going to be a product that clients can use at their own volition and their own choice and their own discretion and not like taking a drug regimen where they have to make sure they get their TID dose in.

So I think there are some different considerations here, and I think "flexibility" is a key word.

DR. GULICK: Dr. Fleming.

DR. FLEMING: I agree with what one of my colleagues said earlier, and that is the Phase 4 post-marketing study is really not the venue or the approach that is going to give us reliable efficacy and safety data usually. I doubt we are going to do a proper no-treatment or placebo control in a Phase 4 environment.

If the issue is urgency--and it is certainly one of the key issues--I would say it is urgency to get a reliable answer, not urgency to get a study done, but urgency to get a study done that will provide robust and compelling results, then actually, you do yourself a disservice by

doing a 6-month study rather than a 12- or a 24-month study. And to be specific, let's even say you're doing just the intermediate-size trial with 100 events, and let's say you are targeting a population that has a 5 percent event rate. That's going to take 2,000 person-years of follow-up. If you follow those people for a year, that's the size of 2,000. If you follow them for 6 months, that's 4,000.

There's no way I am going to finish that 4,000-person enrolled trial until the last person is followed 6 months anywhere close to the time frame I can finish the 2,000-person enrolled trial where I follow that person for 12 months.

So if you are going to drive this issue based on finishing the study sooner, you are clearly going to be doing a disservice by just doing a shorter-term follow-up study--and that's just an approximation. But the bottom line here--I guess I would go back to what you were saying earlier, Dr. Gulick--is that I like the concept of flexibility, too, and if in fact you were saying that some experience could come from a trial with shorter duration as long as there was essential experience coming from at least a 12-month. But I

like the guideline principle as you stated, that A, B, and C are fine in principle, that in essence there ought to be substantial data within this overall application that allows us to at least look over a 12-month period, and that will actually get us answers sooner in calendar time in almost all cases.

DR. GULICK: Okay. I'll summarize what we thought of here.

There were differences of opinion once again. Balancing length of time and sample size came up. I forget who said it, but we would like to follow patients, quote, "as long as practical," which takes a lot of things into account--the urgency of the question, the feasibility of doing long-term follow-up in these particular populations, the fact that safety is a big issue, a really big issue, and obviously efficacy as well, which is why we are doing this study in the first place.

There was an assumption around the table that adherence would decrease over time, but we were challenged by Dr. Fleming over that. The HIVNET 012 and the results of the Cameroon study earlier suggested that there was actually pretty

good follow-up there.

As Dr. Stek reminded us, adherence to a microbicide that is shown to be effective later may actually change over time, so future studies may actually have less of a problem with any kind of adherence issues than earlier studies.

The basic principle, dropouts, missing data is hard to account for statistically, so we heard the phrase "quality trumps quantity," but as Dr. Haubrich pointed out, dropouts can actually give you information if you are able to assess why they dropped out and may speak to the acceptability question as well.

Good retention on a clinical study takes effort, and with limited resources, resources aimed toward that question or that issue are paramount in importance, so planning up front to have specific efforts that will allow people to continue follow-up on the study are key--and it has to be culturally and setting-specific. Whether it is money or food or whatever it is that will keep people coming, those interventions are extremely important and may ultimately save the study and make it interpretable.

On the issue of fixed versus rolling

enrollment, we had some disagreement. In general, it was felt that you gain by following people longer, so that perhaps the rolling idea that you continue to follow people who are already enrolled rather than discontinuing after a fixed amount of time would increase the amount of safety, acceptability and efficacy data you get. But as Dr. Mathews pointed out, it makes the population somewhat less homogeneous when you do do that given differing lengths of follow-up; and selection bias for those people who don't drop out and continue.

In terms of the length of time, flexibility, flexibility, flexibility is what people said, and feasibility as well.

There was a general consensus that 12 months of data is necessary. Whether that could be coupled with some studies that went shorter period of time was something that should be entertained, and longer follow-up data again was felt to be really important, whether it is in the context of a Phase 2 run-in Phase 3, or a Phase 4 where less formal data is generated, but some data can be generated, was a subject of disagreement as well.

How did we do?

DR. BIRNKRANT: I think we have some

ideas.

Then, the other follow-up issue has to do with follow-up once the trial has stopped or once a participant has discontinued. We want to be able to capture seroconversions within the time frame when a patient stops the trial.

So what is a feasible and scientifically sound time frame? Is it one month, or is it longer than one month?

DR. GULICK: Is that clear to everybody? We want to capture late events--the day the study participation stops is not the day you want to stop seeing the patient. So is 4 weeks a reasonable amount of time? Eight weeks?

Dr. Fleming.

DR. FLEMING: Could I seek clarification from Debra. There might be two different ways of interpreting this question.

Let me be real specific. Let's suppose a sponsor plans to do a 12-month, fixed follow-up period on participants. If someone stops treatment at 6 months, it is imperative that that person be followed out to 12 months for an intention to treat for an unbiased assessment.

So I think you are not referring to that

issue, are you, or if you are, I would say that once you stop treatment, clearly you should continue to follow that person for the uniform period of follow-up that the study is designed to obtain so that you get an unbiased assessment of overall treatment effect, i.e., in the spirit of intention to treat.

Now, a separate question that you might have been referring to is let's say you do say 12 months, and you are saying if the trial in fact specifically then indicates that treatment is stopped or that treatment can be continued and stopped at the participants' discretion. Then, are you saying in that context beyond the time period of the formal analysis, should you continue to follow--is that the context of your question?

DR. GULICK: And also, you want to pick up--it depends on how you assess seroconversion.

DR. BIRNKRANT: Right. We were interested in the late seroconverters. However, if the trials are long enough--let's say they are 24 months--we are not as concerned as if they are shorter, perhaps.

DR. GULICK: So--and I'm sorry I don't know this--but on most of the studies, it's true

seroconversion that is the endpoint, so antibody-positive rather than using viral load levels, for instance, which probably are prohibitively expensive--or are both being used in some of the trials?

DR. BIRNKRANT: I don't know.

DR. WU: So far, all the trials have been using seroconversion as the endpoint.

DR. GULICK: So standard antibody testing.

DR. WU: Correct.

DR. BIRNKRANT: Right.

DR. GULICK: So to avoid a window period, you would really want a three-month follow-up to capture most people who--worst case scenario is that they seroconvert on the last day of the study, so 90 percent would be positive by three months later.

Am I getting that right? Dr. Mathews?

DR. BIRNKRANT: But suppose we use a different type of diagnostic test so that we wouldn't have to go that long.

DR. MATHEWS: Right. I think, like if you were going to use viral load, a month would probably be fine. But if you stretch it out too long, and you are doing either modality, then you

may be picking up endpoints that aren't attributable to the--

DR. GULICK: That's right.

DR. MATHEWS: So we would need to know what the medium time to seroconversion is probably in the country or the region. I don't know if that's uniform.

DR. GULICK: Does anybody know that information? So we all carry around 90 percent within three months in this country. Is that the same worldwide? Anybody?

[No response.]

DR. GULICK: Okay. We don't know.

Dr. Bhore?

DR. BHORE: Yes. We do want to know what should be the off-treatment follow-up of those participants who are not lost to follow-up but have discontinued the study drug. So this off-treatment question would apply to those who prematurely discontinue the study drug but not the study, as well as those who have completed the study.

DR. GULICK: I think that's the point Dr. Fleming was addressing before. Strict intent to treat approach, they should be followed for the duration of the study.

Okay. Dr. Birnkrant, did we do everything we needed to do today?

DR. BIRNKRANT: Almost. I have one more question since we have an expert panel here, and that has to do with the population. Do you think we should be enrolling homogeneous subjects, or should we look at a more heterogeneous population given we may only be able to do one trial. Should that one trial be one particular type of subject--for example, high-risk commercial sex workers--or should we get a broad view of the population who will be exposed subsequent to marketing? In other words, once it's on the market, everyone is using it, so should we try to get some of that information ahead of time?

DR. GULICK: I'll make a suggestion here and let others chime in. If we have one study that is our Phase 2/3 study for this compound, it should look at much like the world at-large as possible in order to be able to generalize the results to everyone.

If you were going the traditional Phase 2 and then Phase 3, then I would choose a every homogeneous population for Phase 2 to get the proof of concept and then a much larger population in

Phase 3.

Dr. Haubrich, to add to that.

DR. HAUBRICH: I partially agree with my colleague from the Democratic State of New York but would like to add that if you were going to do the 2A/3 lead-in type of study, you could accomplish both by picking the homogeneous population for your lead-in phase and then widening it out in the Phase 3.

DR. GULICK: That's a good point from the Schwarzenegger State of California.

[Laughter.]

DR. GULICK: Okay.

DR. BIRNKRANT: Now we have accomplished everything.

DR. GULICK: Yes, including making it political right at the end.

[Laughter.]

DR. GULICK: I'd like to thank everyone.

I would like to thank our speakers from the morning for being available all day, for their excellent presentations and really setting the stage for the discussion.

I would especially like to thank the people who presented at the open public hearing.

We had a lot of you, and people were very nice to keep to the time limits, but also some very important points came out both in the oral and the written presentations that people gave. So thanks for doing that. That was extremely helpful to the Committee.

Thanks to the agency, and thanks to the Committee, especially our retiring members; we are sad to see you go.

Thanks.

[Whereupon, at 5 o'clock p.m., the proceedings were concluded.]

- - -