DEPARTMENT OF HEALTH AND HUMAN SERVICES

FOOD AND DRUG ADMINISTRATION

CENTER FOR DRUG EVALUATION AND RESEARCH

ADVISORY COMMITTEE FOR

REPRODUCTIVE HEALTH DRUGS

Volume I

Tuesday, January 23, 2007

8:30 a.m.

5630 Fishers Lane
Rockville, Maryland

Paper Mill Reporting
atoigo1@verizon.net
(301) 495-5831

PARTICIPANTS

Charles Lockwood, M.D., Acting Chair
Teresa Watkins, PharmD, Executive Secretary

MEMBERS:

Maria Bustillo, M.D.
Ronald Gibbs, M.D.
Daniel Gillen. Ph.D.
Julia V. Johnson, M.D.
James R. Scott, M.D.
Jonathan Tobert, Ph.D., Industry Representative
Lorraine J. Tulman, DNSc, RN, FAAN,
Consumer Representative
O. Lenaine Westney, M.D.

TEMPORARY VOTING MEMBERS:

Paula J. Adams Hillard, M.D.
Abbey Berenson, M.D.
Paul Blumenthal, M.D.
Eve Espey, M.D., MPH
Melissa Gilliam, M.D.
Johanna Perlmutter, M.D.
Herbert Peterson, M.D.
Diana Petitti, Ph.D.
Bruce Stadel, M.D., MPH (retired)
James Trussell, M.D.
Elizabeth Shanklin-Selby, Patient Representative

FDA-CDER (NON-VOTING):

Scott Monroe, M.D
Lisa Soule, M.D.
Shelley Slaughter, M.D.
Phill Price, M.D.

C O N T E N T S

P R O C E E D I N G S

Call to Order and Introductions

DR. LOCKWOOD: I want to welcome everybody here. In a minute we will ask everybody to introduce themselves, but I wanted to make a few general comments, which I am sure will be repeated several times today and again tomorrow, and that is that, in the United States, 50 percent of pregnancies are unintended.

That doesn't necessarily mean that they are unwanted, but 50 percent of pregnancies are unintended and about 50 percent of women in the United States will have unintended pregnancies.

In addition, about 50 percent of those pregnancies will result in abortion, and about 50 percent of the other 50 percent can result in late entry into prenatal care. There is an increased rate of low birth weight and possibly prematurity. There are higher rates of child abuse and, in the children, as they grow into adulthood, higher rates of behavioral abnormalities. And for the women that are affected, there are lower socioeconomic

status indices.

Even the Government's Healthy Person 2010 goal has been to have 70 percent of pregnancies be intended and, clearly, safe and effective contraception is going to be an important way to achieve that goal.

In the United States today, about one million pregnancies occur each year as a result of method failure, so the purpose of this meeting and the purpose of this panel is to help the FDA develop a guidance document for clinical investigations of hormonal contraceptives.

Just to sort of set the general rules of engagement, we want to have as much free discussion as possible, so the presentations will be hopefully brief, and I will attempt to restrain the Committee only if we are going on, I think, in more detail perhaps than we need to with regard to a specific question that we have been asked to address.

However, I don't want to stifle good discussions. So we will try to keep to a time frame that is outlined in the agenda. But I am

going to exercise considerable latitude to be able, in the spirit of the Super Bowl, to call audibles and make adjustments as necessary.

I think that, with that, we will go to Teresa.

DR. WATKINS: If we could go around the table and have the Committee members introduce themselves, let's start with Dr. Tobert.

DR. TOBERT: I am Jonathan Tobert. I am an independent consultant now, but I worked for 27 years for Merck, mainly in the cholesterol-lowering field. So I claim no particular expertise in reproductive health except that I have spent a lot of time doing clinical trials.

DR. JOHNSON: I am Julia Johnson and I am a member of the Advisory Committee. I am from the University of Vermont where I am the Vice Chair of Gynecology.

DR. STADEL: I am Bruce Stadel. I a retired FDA medical officer serving as a consultant to the FDA.

DR. PETITTI: I am Diana Petitti, recently

moved for a sabbatical to the University of Southern California, and most recently before that, Kaiser Permanente, Southern California.

DR. GILLIAM: Melissa Gilliam from the University of Chicago.

DR. HILLARD: Paula Hillard. I am Professor of OB-GYN and Pediatrics at the University of Cincinnati where I practice pediatric and adolescent gynecology at Cincinnati Children's Hospital Medical Center. I sometimes say I practice preventive obstetrics.

MS. SHANKLIN-SELBY: I am Elizabeth Shanklin-Selby. I am a Patient Rep.

DR. GILLEN: My name is Daniel Gillen. I am a member of the Advisory Committee and I am on the faculty in the Department of Statistics, University of California at Irvine.

DR. BLUMENTHAL: Paul Blumenthal, Professor of Obstetrics and Gynecology at Stanford University and consultant to the Committee.

DR. GIBBS: Ronald Gibbs, Department of Obstetrics and Gynecology, University of Colorado

School of Medicine.

DR. TRUSSELL: James Trussell, Professor of Economics and Public Affairs at Princeton University.

DR. WATKINS: I am Teresa Watkins, the Designated Federal Official for this committee.

DR. LOCKWOOD: Charles Lockwood, the Chair of the Committee and Professor of OB-GYN at Yale University.

DR. WESTNEY: Lenaine Westney. I am a Committee member and Associate Professor at the University of Texas Health Science Center, Houston, from the Division of Urology.

DR. PETERSON: Bert Peterson, Departments of Maternal and Child Health, and Obstetrics and Gynecology at the University of North Carolina at Chapel Hill.

DR. BERENSON: Abbey Berenson, Professor of Obstetrics and Gynecology, University of Texas Medical Branch, Galveston, Texas.

DR. TULMAN: Lorraine Tulman, University of Pennsylvania School of Nursing, and Consumer

Representative to the Committee.

DR. SCOTT: Jim Scott, Professor of OB-GYN at University of Utah, and Editor of Obstetrics and Gynecology.

DR. MONROE: Scott Monroe, FDA.

DR. SOULE: Lisa Soule, Clinical Team Leader, Division of Reproductive and Urologic Drugs, FDA.

DR. SLAUGHTER: Good morning. I am Shelley Slaughter. I am also a Reproductive Medical Officer/Team Leader in Division of Reproductive and Urologic Products, FDA.

DR. PRICE: Good morning. I am Phill Price, a Medical Officer in the Division of Reproductive Drug Products.

DR. LOCKWOOD: Thank you.

Dr. Monroe is going to inform us about what our job is.

DR. WATKINS: If the Committee members could turn your microphones off so that we don't get any backfeed or interference.

Welcome and Comments

DR. MONROE: Good morning. I am Scott Monroe, the Acting Director of the Division of Reproductive and Urologic Products of the FDA. On behalf of the Division, I would like to welcome all of you to this two-day meeting of the Advisory Committee for Reproductive Health Drugs.

I also want to convey the Division's appreciation to the members of the Advisory Committee who have found time from their very busy schedules to participate in this meeting, and I particularly want to thank Dr. Lockwood for serving as Chair.

[Slide.]

This two-day general meeting will focus on oral, transdermal, and intravaginal hormonal contraceptive products. Although many of the issues that we will be discussing today also apply to injectable products and hormonal implants, such products will not be the focus of this meeting.

There are two major objectives for this meeting. One objective is for the Division to obtain advice on issues that need to be

satisfactorily addressed during the regulatory review of contraceptive products prior to their approval for marketing.

The other objective is for the Division to obtain advice that will assist the Division in creating a Clinical Development Guidance Document for hormonal contraceptives. Currently, there is no FDA Clinical Development Guidance Document for these products.

[Slide.]

To facilitate the Division's obtaining the guidance and advice that it is seeking, the Committee will be asked to discuss seven general topics that are listed on this and the following slide.

The discussion for each of the topics will be guided by specific questions that the Committee will be asked to address.

Topic 1 primarily concerns clinical-trial design issues.

Topic 2 includes two components. The first component concerns clinical-trial design

issues and data analyses to best assess the efficacy of a new hormonal contraceptive product.

The second component concerns the issues that need to be considered in assessing the acceptability of the risk-benefit profile for new hormonal contraceptive product prior to approval for marketing.

Topic 3 will focus on the translation of clinical-trial findings of efficacy and safety into real-world effectiveness and safety.

Topic 4 concerns cycle control, namely, scheduled and unscheduled bleeding or spotting and other measures of product acceptability to the user.

[Slide.]

Other topics to be discussed include extended dosing regimens and post-approval or Phase 4 commitments. The type of post-approval commitment that the Committee will be asked to discuss is that which is requested by the FDA generally to investigate further an uncommon but potentially serious issue that cannot be adequately

investigated in Phase 3 pre-approval clinical trials.

The final topic to be discussed is the role and impact of labeling for communication of clinical-trial findings. Such findings include those related to product efficacy, risk, and other potential benefits.

[Slide.]

The general format of the meeting will be to have each of the seven major discussion topics introduced by one or two brief presentations. Each presentation will be made by a member of either the Division or the Advisory Committee and will be followed by committee discussion.

The agenda for the remainder of today is listed on this slide. Today, the Division would like the Committee to address four of the seven major discussion topics. These topics are issues related to clinical-trial design, assessment of product efficacy and risk/benefit profile, translation of clinical-trial findings to the real world, and lastly, bleeding and spotting or cycle

control.

[Slide.]

Tomorrow, the Committee will be asked to address the remaining three topics: extended dosing regimens, Phase 4 commitments, and the role and impact of labeling. Oral presentations by interested organizations and individuals are also scheduled for Day 2.

I would now like to introduce the first speaker of the day, Dr. Phill Price of the FDA.

DR. WATKINS: Before we turn it over to Dr. Price, I would like to read into the record the Conflict of Interest Statement.

DR. MONROE: Fine.

Conflict of Interest Statement

DR. WATKINS: The Food and Drug Administration is convening today's meeting of the Reproductive Health Drugs Advisory Committee under the authority of the Federal Advisory Committee Act of 1972.

The Committee will discuss current issues that influence the consideration for approval of

oral and non-oral; i.e., transdermal and intravaginal hormonal contraceptive drug products.

Issues for discussion will include clinical-trial design, expectation for efficacy and safety outcomes and measures of acceptability of the product to the user, including cycle control.

This topic is a particular matter of general applicability. Unlike issues in which a particular firm's product is discussed, the topic of today's meeting may affect all hormonal contraceptive drugs currently on the market and in development with the exception of implantable and injectable hormone products and their sponsors.

The participants have been screened for potential financial conflicts of interest with respect to the products and firms that could be affected by today's discussion.

In accordance with 18 U.S.C. 208(b)(3), full waivers have been granted to the following participants:

Dr. Melissa Gilliam, Paula Adams Hillard, and Johanna Perlmutter.

Waiver documents are available at FDA's docket web page. Specific instructions as to how to access the web page are available outside today's meeting room at the FDA information table.

In addition, copies of all of the waivers can be obtained by submitting a written request to the Agency's Freedom of Information Office, Room 12A-30 of the Parklawn Building.

FDA acknowledges that there may be potential conflicts of interest, but because of the general nature of the discussions before the Committee, these potential conflicts are mitigated.

Further, with respect to FDA's invited industry representative, we would like to disclose that Dr. Jonathan Tobert is participating in this meeting as a non-voting industry representative acting on behalf of regulated industry. Dr. Tobert's role on this committee is to represent industry interests in general, and not any one particular company. Dr. Tobert owns Tobert Medical Consulting and is a retired employee of Merck.

In the event the discussions involve any

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

In the interest of fairness, FDA encourages all other participants to advise the Committee of financial relationships that they may have with any firm whose product upon which they wish to comment.

Topic 1 - Clinical Trial Design Issues

DR. PRICE: Good morning. My name is Phill Price and I am a Medical Officer in the Division of Reproductive and Urologic Drug Products. I will be presenting clinical design issues that have emerged in the Division of Reproductive and Urologic Products over several decades of our review of hormonal contraceptive products.

[Slide.]

Hormonal contraceptive products, and products in general, are normally revised or

normally developed in four phases, 1, 2, 3, and 4. I am sure everyone is familiar with those.

In general, hormonal contraceptive trials are usually conducted in Phases 1 through 4 after initial animal testing and in suitable animal species. In the development of newer contraceptives, these phases may not follow in sequences 1 through 4 but may be abbreviated.

For example, for a new molecular entity, a complete Phase 1 through 4 developmental program is necessary while a previously developed estrogen and progestin might skip Phase 1 and accelerate Phases 2 and/or 3.

[Slide.]

Phase 1: Phase 1 safety of hormonal drug development is usually limited to initial safety issues especially tolerability although initial pharmacokinetic and drug interaction data may be accrued.

[Slide.]

Phase 2: Phase 2 hormonal contraceptives is focused on ovulation suppression, studies with

several dosages that have been identified in prior animal studies. These dose-finding studies attempt to suppress ovulation in one or more dosages at 90 to 100 percent of subjects.

Preliminary data to predict efficacy is obtained in a small number of subjects, as well as some additional data.

[Slide.]

Phase 3: Safety and Efficacy. This slide outlines Phase 3 hormonal contraceptive trial development by sections that are presented in a typical Phase 3 protocol. Sections in Phase 3 are type of trial, trial design, entry criteria, study procedures, efficacy, safety evaluation, cycle control, and discontinuations.

[Slide.]

I will now review each section in Phase 3 clinical-trial development.

Type of Trial: Phase 3 hormonal contraceptive trials are usually open label and compare expected pregnancy rates in sexually active women not using a contraceptive method.

Active controlled trials compare the proposed product against a product in its class--for example, a 20-microgram product against another 20-marketed microgram product. It should be noted that for approval in the U.S., data accrued in an active controlled trial is not required for approval.

[Slide.]

Trial Size: Trial size is based upon whether the product is a new molecular entity or a non-new molecular entity. For new molecular entities, it is recommended that 20,000 28-day cycles, or equivalent, within the first year is completed.

By "equivalent," we mean that other regimens, such as extended dosing regimens, such as 28-day cycles, are further extended, that efficacy and safety comparisons will be compared to a 28-day regimen. In addition, 400 women should complete 13, 28-day cycles or equivalent.

Importantly, the number of these trials have been variable. All the products may enroll

20- to 30,000 treatment cycles or more. Of note, primary modification in the past 15 to 20 years have been to the progestin component of the oral contraceptive--for example, norgestimate, desogestrel, and drospirenone.

For a new molecular entity, two trials are generally recommended because one trial serves to validate the findings that we have seen in the previous trial.

Secondly, the Division would consider one robust clinical-trial if, indeed, the safety and efficacy of the trial documented that.

[Slide.]

For a non-new molecular entity, 10,000 28-day cycles, or equivalent, within the first year of treatment is recommended, as well as 200 women completing 13, 28-day cycles or equivalent.

[Slide.]

Entry Criteria: Protocols generally specify the following parameters. The subjects should be sexually active and not using any other form of contraceptive on a regular basis. The age

is usually identified. Body Mass Index is identified, smoking, percentage of the switchers and fresh starts, labeled contraindications and exclusions, as well as other exclusions. Except for No. 1, sexually active, there have been significant variability in these criteria.

[Slide.]

Variability in Entry Criteria Phase 3: Modifications that are most relevant to the variability in Phase 3 are:

Age; since an advisory committee meeting in 1994, the age range for entry was increased to allow women greater than age 35 to be entered into clinical trials if they were healthy and had no serious risk factors.

BMI; generally, sponsors have sought to limit subjects with a BMI of less than 30 to 35. The Division would encourage no limit on the BMI if the subject has no other risk factors.

Smoking; some trials would limit the number of smokers who entered the trial.

Switchers Versus Fresh Starts; some trials

do not identify switchers versus fresh starts or have very few fresh starts in the trial. The Division encourages identifying the number of fresh starts, as well as increasing the number, so that it represents the general population.

Some exclusions have included eliminating subjects who have family members with a history of thromboembolic disease. Some trials have also not enrolled subjects who have had adverse bleeding patterns while taking another similar oral contraceptive.

[Slide.]

Standard Procedures: Standard entry procedures in Phase 3 protocols include baseline history and physical, baseline vital signs, physical examination including pap smear, baseline laboratory tests, chlamydial screening tests, hemostatic profile, and possible mammography for age greater than 35, and a serum HCG.

Pregnancy testing has also been variable except for baseline HCG. Some trials propose only serum testing while others propose only urine

testing. Some trials also perform monthly urine testing while other trials missed perform urine testing if a period is missed. Other studies propose all pregnancy tests be sent to a central laboratory while others accept a home urine test.

[Slide.]

Study Procedures continued: Historically, paper diaries have been collected for over 40 years. Recently, two trials have used electronic diary data exclusively to collect data. Diary data captures pill use, bleeding and spotting collection, and, since the 1980s and 1990s, diary data also documents the use of back-up contraception and whether there has been any sexual activity in the monthly cycle.

Most studies include a section on treatment compliance that state how many consecutive pills a subject may miss--for example, two days, three days, or five days. The investigator then informs the medical monitor to discuss possible withdrawal from the study.

The Division would encourage more

uniformity with subject withdrawal and specific reasons why a subject withdrew from the clinical trial should be documented in the diary.

In addition, diary data has also sought recently to outline subject-satisfaction data, and even more recently, there has been use of a patient report outcome instrument, which some companies are seeking to use.

[Slide.]

Efficacy: All Phase 3 protocols identify the following; efficacy, open cycle control, discontinuation rates, as well as safety.

[Slide.]

Efficacy: Efficacy is based upon on-treatment pregnancies. On-treatment pregnancies are calculated from the start of pill intake to taking the last pill and extends up to 14 days after the last pill intake.

In the past, this has been the primary analysis method used by the Division in evaluating hormonal contraceptives in the primary analysis. Secondary analysis has also been assessed by

sponsor for pregnancies that occur within two to five days.

Failure Rate Assessment: Failure rate assessment has historically been by the Pearl Index. The Pearl Index outlines a specific point estimate plus 95 percent confidence interval, and the Division looks at both the upper and the lower confidence interval.

The life-table analysis method is also utilized. The Division normally looks at consistency between the Pearl Index, and well as the life-table analysis.

Dr. James Trussell and Dr. Daniel Gillen will discuss efficacy and failure-rate assessment in much more detail later today.

[Slide.]

Failure types are primarily method failure and user failure. In method failure, the subject has recorded that she has taken the medication perfectly, while in user failure, the subject records that she has missed one or more dosages.

In general, for primary efficacy analysis,

cycles in which subjects have used no back-up contraception, cycles in which subjects are not sexually active, and subjects who are greater than 35 years of age are excluded from the primary analysis. Secondary analysis may be performed on other populations, such as combining all subjects in a trial who are above and below age 35.

[Slide.]

Cycle Control: There is presently no standardized way of addressing cycle control in clinical trials. Bleeding, spotting, bleeding and spotting definitions are plentiful and variable in various trials. There is no uniformity between sponsors in this section of the protocol. This topic will also be discussed in detail later in the meeting.

[Slide.]

Discontinuations: Discontinuations are usually driven by protocol-termination criteria. Patient withdrawals may be high, in the range of 50 to 60 percent. The range may be as low as 10 to 15 percent and, typically, the range has varied

between 17 and 35 percent. However, variability has been noted in a number of clinical trials and this can be seen by the number of subjects who may have been excluded for missing either two, three, or five pills.

Importantly, the Division feels that the evaluation of patient withdrawal rates can provide an assessment of how acceptable a method is likely to be in the general population of potential users and should be well documented.

[Slide.]

Limitations of Phase 3 Trials for Assessment of Product Safety. Phase 3 trials have limitations in their adverse events; for example, thrombotic events occur infrequently and their frequency cannot be well defined in the Phase 3 trial. They have evaluated somewhere between 10- and 20,000, 28-day cycles or equivalent.

To better define these risks, Phase 4 studies may be requested.

I will now turn the meeting over to Dr. Lockwood and our assembled experts.

Thank you.

DR. LOCKWOOD: Let me just start by thanking Phill and asking the panel if there are any questions that are specific to Phill's presentation.

[No response.]

DR. LOCKWOOD: Is it possible to have the questions that we are going to--amazing. Okay. What I would like to do, I will just reiterate what the question is and then invite comments, arguments, debates, discussions, anecdotes, et cetera, from the group.

The first question that we have been asked to address is: Should entry criteria be more reflective of typical or actual clinical prescribing and particularly regarding variation in the progressively increasing BMI in the United States, smoking, and family history of thrombosis and thromboembolism?

Why don't we start with that. Well, let me start with a comment about family history of thrombosis and thromboembolism. This is an area, a

moving target, I think, in general, and there is certainly a growing body of evidence to suggest, for example, that inherited thrombophilias represent a significant proportion of people that have thromboembolic phenomena and that the presence of an inherited thrombophilia is important primarily in the context of family and personal histories of thrombosis and thromboembolism.

So, I do think that a family history of thrombosis and thromboembolism ought to be a red flag to contraceptive use. This is not necessarily saying that it ought to be prescribed, but I think that I would recommend that maybe the one setting where it would be appropriate to do a limited screen for thrombophilias is before allowing patients to be enrolled in the study, and I would certainly encourage physicians in the real world to think about pre-testing patients of European extraction for Factor V Leiden prothrombin gene mutation, at least those two--those are the most common, the most prevalent--and, if they are negative, then, with a family history, I think that

probably a legitimate case could be made that they still could be enrolled.

Dr. Scott.

DR. SCOTT: Charlie, of course the risk is even higher than it is with--oh; sorry. The risk of pregnancy, of course, is much higher than it is with contraception with thrombophilias, and I think most of the studies that looked at screening for pills, for pregnancy, and so on, have found that it is not really cost effective.

I just wonder. You say the only two that you would recommend would be Leiden factor and what else?

DR. LOCKWOOD: And the prothrombin gene mutation.

DR. SCOTT: What does it cost?

DR. LOCKWOOD: Probably around 800 bucks.

DR. JOHNSON: Now, you are talking about testing before the studies. Not every patient that is going to start contraceptives, or am I misunderstanding?

DR. LOCKWOOD: That's correct.

DR. JOHNSON: Just for the studies.

DR. LOCKWOOD: I think that this argument about cost effectiveness depends on exactly how high the risk really is and, in some studies, the risks on oral contraceptives with Factor V Leiden are as high as 35-fold increases. If that is borne out in larger studies, cost-benefit analysis may actually favor test.

DR. SCOTT: When you say 35-fold increase, though, that depends on the denominator. How many is that in 100 patients or 1,000 patients, and so on, compared to non-thrombophilic patients?

DR. LOCKWOOD: We will get back to Dr. Johnson, but the prevalence effect of Factor Leiden varies in the United States--it actually varies in Europe--but in general, it's about 5 percent.

The risk of thrombosis, when there is a personal or family history, is increased probably on the order of, well, certainly a minimum of 10-fold, perhaps even higher than that, maybe 50-fold. So, again, context is critical and it is in those patients I am talking about, with the

personal or a first-degree relative with a history of thrombosis or thromboembolism that I am advocating testing.

DR. JOHNSON: I wanted to agree with you that I think it is reasonable for all new hormonal contraceptives to get the family history and potentially exclude patients with a family history of VTE.

Having said that, though, there is very limited data on these patients. You could put them into the trials in a method to learn if indeed this puts these patients at greater risk.

I would disagree, however, in testing all patients in these trials for these disorders. I agree that it is a high cost. I think excluding family history is reasonable, but the testing, I would say, would have to be up to the manufacturer and whether they thought that testing was useful information. But I agree that family history is important and I think that is going to add to the knowledge that we know that these patients are potentially at lower risk to start with and could

be used to advise physicians to counsel patients appropriately when the products come to market.

DR. LOCKWOOD: Paula.

DR. HILLARD: One of the issues that I would suggest that at least be brought out in the open is the difficulty of obtaining that family history and the reliability of that family history.

I think it is important you stated first-degree relatives. So I think that is important to note. But even with a first-degree relative, asking an individual about their positive family history for blood clots and describing what a clot is and talking about that, indicating that the individual with the clot would have been hospitalized and placed on a blood thinner, it is difficult to obtain that history.

So, I think we have to acknowledge that that is the case and many individuals are unable to give that accurate family history, and we are left wondering was it really an episode of VTE.

DR. PETERSON: Looking at it, I think there is an even bigger issue than family history.

Looking at, I think, in the big picture, one of the fundamental tradeoffs we have got to grapple with is the issue of protecting the study subjects and the generalizability of study findings.

Clearly, for the reasons already mentioned, there are going to be a lot of women who are obese, who are under 35, who smoke, and who have a family history, who are going to be using these products once they are approved.

So, the question is how do you balance those tradeoffs.

DR. LOCKWOOD: I don't want to be an advocate for the tobacco industry, but I would say this, that smoking prevents preeclampsia, and it also is not associated in studies with venothrombosis. I am not advocating--

DR. PETITTI: I am going to follow up on Bert's comment. I think we have to decide fundamentally, what we are attempting to accomplish in a trial, in this trial for new products.

If indeed we are attempting to estimate something that can be generalized to the population

and is representative of use in the women who will use, then we should take every single person who a physician would put on oral contraceptives or on hormonal contraceptions absent the trial.

As Bert points out, when we believe that the product that is being tested might have a different risk of venous thromboembolism, then there may be a compelling reason to exclude those women who have a higher risk, underlying background risk of venous thromboembolism.

But in this day and age, with hormonal contraception being what it is, I don't think we should be testing products that we think, a priori, have a risk of venous thromboembolism different from those of the existing products, since those products are not acceptable in any marketplace given the alternatives.

So, I am going to make a very radical suggestion that I personally think that there should be no exclusions except those exclusions that are on the label.

DR. LOCKWOOD: Now that I have stirred the

pot, which was my intent, why don't we broaden the responses to all these issues; BMI, smoking and venous thrombosis. It is creeping into the discussion anyway.

DR. LOCKWOOD: Who is next? Dr. Stadel.

DR. STADEL: Thank you. I basically agree with Dr. Petitti. I think that, in a trial to license a drug, we need to think about how the information will be used, how it relates to the marketing of the drug, the advertising, that the entry criteria for a licensure trial should correspond to the intended marketplace population.

There are actually ways using various kinds of data sets to actually examine who uses oral contraceptives, what is their mix by BMI, and so forth, and perhaps such data should be looked at by people who are planning studies, so that insofar as possible, they test what they are proposing to market in the people they are proposing to market it to. Thank you.

DR. LOCKWOOD: Before I respond to--this

really does get at the issue of our obligations to do no harm in these clinical studies and balancing that with the applications that we know will occur in the real world, and it really is an ethical and a public health question. But I would point out an important factor, which is if you are doing a study that involves 400 women, it is very unlikely that you are going to see the kinds of adverse events we are talking about.

If you really were to do a study to assess the true risk of thromboembolic disease, for example, in obese, non-smoking, people with a family history, you would have to purposely select them and then compare the two agents.

So, one of the themes we are going to come back to again and again is how much of a burden do we put on the sponsor in terms of the size of a study. If the number will be 20,000 cycles and 400 women years, we are very unlikely to discover real risks of venous thrombotic events in obese or non-obese patients, et cetera, in those kinds of studies, and this will also be discussed when we

discuss Phase 4 studies, as well.

We are way behind. Dr. Gillen.

DR. GILLEN: Thank you. One thing I want to say is that I agree with the previous comments in terms of having entry criteria best try and reflect at least the target population for generalizability.

I think one thing that is going to come up, and it is going to come up in a few minutes when we talk about trial-design issues, is, as you are starting to change entry criteria, and the precedence has been in the past to use historical controlled trials, you now start to set a moving benchmark in some sense because of the entry criteria and different confounders may be coming into these trials. So. how do you compare with the current trial to the past trials with respect to entry criteria and what is going on?

I think it is going to start setting--we are either going to be moving this benchmark in some hopefully non-arbitrary manner, or going to something more along the lines of active controlled

trials where we do have randomization.

So, it is just a comment in terms of as these things progress, we may need to change the standards by which we are evaluating efficacy.

DR. LOCKWOOD: Thank you.

Dr. Espey.

DR. ESPEY: I don't have a comment.

DR. LOCKWOOD: Dr. Trussell.

DR. TRUSSELL: Just in counting, I strongly agree with Dr. Petitti's recommendation. If companies want to protect themselves against an adverse trial outcome, then they put an active control.

DR. LOCKWOOD: Dr. Berenson.

DR. BERENSON: In follow up to what Dr. Lockwood said, I think there are two questions that are being raised here. One is should we exclude women that represent many women in the general population from these studies, and number two is are we going to actually be examining the efficacy or the side effects in certain populations.

The first one seems fairly easy to

implement. But the second one could be a major problem in terms of study size, because, if we get into BMI and say we want to prove that these products are safe in women over 300 pounds, and then, as some of the literature that was sent to us in advance proposed, we have to have 20,000 cycles in women over 300 pounds, I think this is going to place an undue burden on the manufacturers and make it difficult to label these methods for many women that need them.

DR. LOCKWOOD: Dr. Blumenthal.

DR. BLUMENTHAL: I think that the overall theme of this meeting is really twofold; generalizability and relating research results to the real world. The problem that I sometimes see in what happens in the research world is, if you take that statement, should entry criteria be more reflective of actual clinical prescribing, what actually happens is just the reverse.

The clinical prescribing is reflective of the research entry criteria and what we want to do is, I think as Dr. Petitti said, open things up so

that the actual research criteria actually lead people in the direction of what is going to happen clinically, because when research is restrictive, clinical practice becomes restrictive, and we can't say much about what happens in clinical practice because of the limitations of research.

So, if you are really attempting to get generalizability, then you have to open things up and you have to be sure that you either include some of these subgroups in specific substudies, or you plan things like case-controlled studies in advance.

You don't wait for things to happen post-marketing, but you plan them more or less in advance knowing that you want to look at these groups later.

DR. LOCKWOOD: Dr. Gilliam.

DR. GILLIAM: My original point was about the study size that you would need to adequately explore some of these issues, but I think the other issue is that what we initially want to do with these drugs is to prove that they work, which is an

issue of efficacy.

As we start to combine these other issues about safety and effectiveness, then you can understand why it might not be advantageous to a company or to someone trying to fund a study if, for example, using a woman with a BMI of 35 somehow inhibits the efficacy of a drug.

So, I think that is the other balance that we are trying to make, do we somehow undermine how effective a drug appears biologically if we start to add these diverse populations.

DR. TRUSSELL: I don't understand, Melissa--but we want to know how effective the drug is, not how effective the drug appears. If it is going to be used by women--I mean, look what has happened to BMI in the United States. It would not be reflective of the country to not put any people in the trial with a BMI over 30.

DR. GILLIAM: I agree. I think it is what we are balancing, and I think it represents sort of a frame shift. We have really thought about how to prove that a drug actually works, and now we are

trying to say, well, how does it work in the reality of today's situation. It is sort of playing devil's advocate.

DR. LOCKWOOD: This really gets at yet another issue, which is that, as doses of the ethinyl-estradiol component of the pill drop, the potential for forgiveness of the agent is likely to drop, too, and then what do we do about a woman that has a BMI of 35, who misses three pills, and doesn't start her pill again for 10 days instead of at 7 days? Just how much data are we demanding from a clinical trial to be able to model the impact of that agent in subpopulations.

I think that that really does frame some of the discussion we need to have.

I think the other point I want to raise is sort of beyond the ethics of this debate. There are sort of the political implications, just how much do we demand that the government do to ensure the safety and efficacy of a drug versus how much is the individual prescriber and patient's responsibility to obtain information and make

educated, intelligent decisions, the caveat emptor, the libertarian argument.

Dr. Gibbs. Sorry; I didn't see you there.

DR. GIBBS: No.

DR. LOCKWOOD: Dr. Johnson.

DR. JOHNSON: I just wanted to agree with Dr. Trussell. I think there actually is a difference in these three things, the BMI, smoking, and VTE family history.

The second two really have to do with safety issues primarily. The BMI really is efficacy, and I do think we need to know, so we can counsel patients effectively. I would argue that that is one area that we really do need to include patients with higher BMI, so we can counsel our patients appropriately, because we know these women, just like all women, need contraceptives. And I don't think it would be difficult to recruit a reasonable number of women with higher BMIs.

DR. LOCKWOOD: Dr. Petitti.

DR. PETITTI: I think, when we talk about the design of clinical trials, Phase 2 and Phase 3

studies, that we should stop pretending that we are gaining any information whatever about the safety when safety is defined in terms of major adverse events like venous thromboembolism, stroke, and myocardial infarction.

Any event that occurs in a trial is certainly a random event in trials of these size. So, in the clinical-trial design, I would like to suggest that we focus on how we can better estimate efficacy or effectiveness, whichever we decide we want to estimate, and put safety into the post-marketing realm.

DR. LOCKWOOD: Dr. Stadel.

DR. STADEL: I agree very much with what Dr. Petitti just said about efficacy being sort of the primary guiding thing in choosing a study population that is representative of the intended marketplace population in the evaluation of efficacy.

There are always some difficult decisions with regard to the extreme "n's" of safety, such as before the family history. I don't know what we

will come up with an 100 percent answer to something like that.

I do think there is one other issue here, and that is, a company--it's a company that markets a drug, and it develops it, and it does have to deal with its liabilities when it markets it, so, to some degree, there is a dialogue about what a company chooses to define as its marketplace population.

I, myself, see some little room for positioning there provided the marketplace population is defined clearly in advance and the trial populations are defined with regard to that intended marketplace population, so that one develops a good, clear tracking for who that company will be pushing its marketing of the drug to.

I think there has been some disconnect in those areas in the past, at least based upon my experience over the years. There has been some historical development in this kind of thinking. So I would very much encourage that concept that a

company choose who am my intending to market to and to come up with a realistic plan to test the drug for efficacy in the intended marketplace population, and there could be a little variation between companies and who they say the drug wasn't studied in, or something like that.

DR. LOCKWOOD: Maybe I am going to at this point try to move on to the next question by summarizing, I think, the sense of the panel.

I think there seems to be a consensus that it is virtually impossible to obtain adequate and accurate safety information given the enormous size of a trial that will be required. It would be impracticable and, in fact, it would restrict the access of new drugs to the market because it would be so impracticable.

I think that there seems to be consensus, as well, that more real-world testing is necessary and that the inclusion criteria for clinical trials ought to be expanded to include women that smoke, women that have a much wider range of BMIs.

I don't know that there is consensus on

whether or not women with a first-degree relative or a personal history of thrombosis ought to be included. I think that may be at the discretion of the drug company, because they do incur substantial liability if that were actually included, I suspect.

However, the implications of what I just said are that the clinical trials would have to be larger. It seems to me that if you are including a much wider range of women who are likely to have higher failure rates presumably, particularly with lower dose drugs, that 20,000 cycles in 400 women may not be adequate.

I just want to finish this question up by general comments about is there a size that would limit the real-world application of your trial. What if including women, an adequate number of women with a BMI of greater than 30, would require a trial of 70,000 cycles and 1,200 women.

DR. TRUSSELL: It wouldn't. All it needs is an active control.

DR. LOCKWOOD: Can you elaborate a little

bit more on that?

DR. TRUSSELL: Well, let's suppose that you--I will just make this up--let's suppose that you have a not-new molecular entity, but you have another 20-microgram pill. You let everybody into the trial. You randomize against a product that is already approved, and if it looks as good as that one, fine, even if the Pearl Index is 3.

DR. LOCKWOOD: And just let people know how efficacious it is, and providers have to then counsel their patients accordingly?

DR. TRUSSELL: Well, it makes no sense to say that a pill has a certain effectiveness if the population to which you are speaking is not the population on which the drug was tested.

DR. ESPEY: I think the whole idea of this question looking at study entry criteria really can't be separated from the question of study design.

I think that if we are going to be, as we should be, more open about who is entered into the studies, then we have to talk about an active study

design as opposed to the historical controls.

DR. TOBERT: Yes, I think this is a question which lies behind a lot of these other questions, perhaps the primary question. I must say I was quite surprised when I started preparing for this meeting to find out that FDA and, indeed, the European agencies are quite happy with uncontrolled trials.

When I got into it, I sort of started to see some of the reasons why. But really, I mean, if you think about it, if you are going to study, say, 2,000 patients, do you get more information by putting all 2,000 onto your test product, or do you get more information by, as I think Dr. Trussell is suggesting, dividing them either equally 1,000 and 1,000, or perhaps a 2 to 1 randomization ratio?

I think that is more informative because then you get information relative to a standard and you are not so dependent on the kind of choices that you make in selecting the population to be studied.

DR. LOCKWOOD: In a sense, that strategy

is sort of like the law. You are building on precedent, and you have sort of said, well, this is an accepted agent, we are using this 20-microgram dose, and we are going to test this new agent against it, just so we understand it.

Dr. Gibbs.

DR. GIBBS: I wanted to go back to the issue of sample size. Women of high BMI have increased problems in pregnancy, increased Caesarian-section rate complications, Caesarian diabetes, hypertension, so they need contraceptives, too.

I think what the idea should be is to encourage development of contraceptives for these women. So, increasing sample size from 20,000 cycles to 70,000 cycles would be one way to do it. But I think that would be an awfully expensive way, and maybe what we could do is say, well, of those 20,000 cycles, maybe a dedicated percent should be enriched by women of high BMI.

DR. LOCKWOOD: Any comments about that idea?

DR. TOBERT: Yes. I mean, this is certainly relevant. I was looking at, more or less at random, the Ortho Evra label coming down here yesterday, and, of 15 pregnancies that occurred, five were in woman who weighed more than 90 kilograms. A third of the pregnancies were in these heavier women, but they only accounted for 3 percent of the study population.

So, there may well be quite a large effect here, although I think they picked it up with open label trials, as well as active controlled trials. I still think active is better for this sort of thing.

DR. LOCKWOOD: Any other questions?

[No response.]

DR. LOCKWOOD: We have given the FDA a lot to think about. Let's move on to the next question, which is: The Division has seen different efficacy results in foreign studies compared to U.S. studies, often better efficacy results in Europe. Should a certain minimum percentage of the subjects in Phase 3 studies be studied at U.S.

sites?

Dr. Johnson.

DR. JOHNSON: I would say that I agree with this philosophy that it is important to look at the population that we have in the U.S. with these studies. It is a mixed population, we can get a good variety of individuals.

I think it is reasonable to ask that the Phase 3 trials, a certain minimal percentage be done in the United States. Now, what exact percentage that should be, that is maybe for the research individuals to comment.

DR. STADEL: I would just offer whether it should consider the possibility of a minimum number as opposed to a minimum percentage depending on a company's Phase 3 program and the degree to which they are adapting a common development program to the needs of different countries, and so forth.

Establishing a minimum number as opposed to a minimum percentage might work out more practically, but I certainly agree that some minimum part of the Phase 3 development program

should be done in the population to which marketing is intended in this country.

If they are asking this country who are marketing here, then, some basic portion of the data should come from here.

DR. LOCKWOOD: It is actually evolving into the third question, which I guess they ought to be considered together. The argument that is being made is not in any way meant to reflect poorly on study designs in Europe, but rather the population heterogeneity in the United States.

I don't want to get into any diplomatic issues here. Dr. Gillen.

DR. GILLEN: I view this as no different than any other international trial that I have worked on in any other setting. Contraception is no different. We have heterogeneity across nations, and we know that.

I mean, there could be differences in baseline sexual activity rates across nations in terms of study participants that are participating in trials with respect to these confounders that we

have mentioned, such as BMI, smoking, and age.

I mean, anytime we go into any other trials where we have potential differences in study populations that are entering our trials, and we are doing comparisons, we need to consider the populations that are going to ultimately be marketed to, and this is the exact same concept to me that we have been talking about.

So, yes, I mean the bottom line is definitely--I mean, if we are going to be marketing these contraceptives in the United States, then we need to assess their efficacy within the United States, as well.

DR. LOCKWOOD: Dr. Blumenthal.

DR. BLUMENTHAL: I think you are right that all of these questions are interrelated, even the second two with the first, and what we are really trying to figure out is, well, if we see better efficacy results in foreign locations, what is the reason, and does that actually relate to some difference in physiology, which seems unlikely, or does it really relate to the cultural

and more social differences.

I think that, when we see data coming in from foreign sites, that, if a company wants to market in the United States, then things done here have to make up the difference.

In other words, we look at the different populations and look at some of the aspects of those studies that cannot be replicated here--I mean, that must be replicated here--that is the focus that we should take in looking at U.S. studies, make up the difference between what we feel is relevant to the U.S. population and what is provided from abroad.

DR. PETITTI: I would say that there is one caveat on this. If, indeed, a study done in a foreign country had an active comparison group which was a product which is widely used in the United States, and it were shown to be equivalent, I would be willing to think that that would be a situation in which we might not require a U.S. site.

On the other hand, I think the thing that

we keep tripping on is what I think is fairly recent, which is the recognition of the degree to which BMI is a modifier of the effectiveness of hormonal contraception, coupled with the epidemic of obesity in the United States.

DR. BLUMENTHAL: I agree with you and I think that BMI is probably the most glaring difference sometimes between foreign studies and U.S. studies, or the foreign population used in studies, that have participated in studies, and the U.S. population likely to use the product. As we see what may be happening and see more countries being involved in these studies, particularly in Asia, we are going to see even a different group of patients who may lead much more ordered lives than women in the U.S.

That may be one reason why efficacy--you know, the banks close at 5 o'clock in many countries. So I think that obesity or BMI is one issue that just needs to be specifically addressed even if there is an active control, but issues of literacy and fear and ambivalence about methods

play into efficacy, as well, and those are issues that I think are often very particular to the U.S.

DR. STADEL: I agree very strongly with the comments about active comparator, but would note that one does have to label the product with a pregnancy rate, so the active comparison alone doesn't give you all the information that you have to have, which I think is no substitute to studying the drug in the population you intend to sell it to.

I do think that probably some work could be done in the proposed foreign data by looking at the baseline characteristics of the proposed foreign population, what the birth rate is, and so on, to establish whether a proposed foreign population is a suitable population to include in a marketplace, in a license or application to the U.S.

DR. LOCKWOOD: Dr. Gillen.

DR. GILLEN: I absolutely agree that I think effect modification is truly what we are worried about here. I think another potential

effect modifier that we could run into, though, is also compliance, and maybe this is potentially what you are getting into.

When you are going into compliance, okay, if you have equal method versus user failure rates and find you don't have an issue--but if those things are differential and you are doing even your comparison, active controlled trial and you have differences across the two study populations, then you could be seeing different results in that respect.

Again, that is going to dictate the efficacy with which we would observe in the United States.

DR. LOCKWOOD: Dr. Scott.

DR. SCOTT: Just a question for my own benefit because I don't know about this, we are talking about comparison studies.

If 20 different companies wanted to compare their products with another 20-microgram oral contraceptive, and they are already all these on the market, can they just go ahead and do it? I

mean, is there any limitation?

I am talking about from a patient and a physician standpoint where there are already plenty, can anybody just keep adding more and more?

DR. LOCKWOOD: Phill?

DR. PRICE: Yes.

DR. LOCKWOOD: Free market, right.

Dr. Espey.

DR. ESPEY: Just to play the devil's advocate here. I would have some concern that making an actual requirement for a percentage could have the negative effect of potentially excluding some drugs from being tested.

If, for example, it cost more money to do it here, or there were other barriers that companies found for doing testing in the United States as opposed to elsewhere. I mean, I think overall we are really not all that different.

There are particular things like BMI that are a concern, but to create an actual quota, I would just have the concern that that could have a chilling effect.

DR. LOCKWOOD: So, we are advocating outsourcing. Just kidding.

Dr. Tobert.

DR. TOBERT: Yes, I have a somewhat similar point. I think, well, firstly, I think Dr. Stadel said this, too, that it should be a number, not a percentage. I mean, if different companies market their drugs, well, if each region of the world wanted 50 percent, the math wouldn't quite work out, would it.

But, I mean, you could have X hundred patients, although I think we should recognize that you are really statistically not likely to see any real differences. I think it is more of a sort of "feel good" issue than a real solid statistical issue.

DR. LOCKWOOD: If I can sort of summarize what I think is the sentiment of the group, studies from Europe and other areas of the world are potentially very valid and useful, and that a careful analysis of those studies may indicate areas where their applicability to the real world,

to typical use in the United States may not have been adequately assessed.

So, that would allow for sponsors to do a more defined study, for example, in a larger BMI group in the United States, to buttress that European data, and it could be used collectively for approval status.

Is that pretty much what we are saying? Great.

Okay. The next question is: Should a certain percentage of the study population represent "fresh starts" as opposed to "switchers"?

DR. TRUSSELL: Can I say something.

DR. LOCKWOOD: Well, I think we included three.

DR. TRUSSELL: Again, I would say what is it likely to be in the real world, and let that be it. There is no point in requiring 100 percent fresh starts if, in the real world, 50 percent are going to be fresh starts and 50 percent are going to be switchers.

DR. LOCKWOOD: Dr. Berenson.

DR. BERENSON: I can definitely speak to this topic having a protocol currently that I am doing where everyone has to be a fresh start. And the studies are going to have to include women under 18 if we are going to have a large percentage be fresh starts, because the mean onset of sexual activity in this country is now 16.

So, if you are trying to do studies on women 18 through 50, how are you going to recruit fresh starts is my first question. It will take the studies a long time to recruit.

Number two is why. Are we saying that efficacy is different if you have taken another birth-control pill previously when you are using this birth control pill, that that pill had a lingering effect, that it is going to help you? Are we saying that the side effects are different?

So, what issue is it that is so important that we would put this burden on the companies?

DR. LOCKWOOD: I want to come back to age for one second, and we will come to you in a minute, but just to make a point that I think the

concept here is that switchers have experience. They may actually be more potentially compliant, more knowledgeable about sort of the rules and regs of contraceptive use, and so forth. Correct me if I am wrong on that.

You do raise I think a really interesting question about age at both extremes that needs to be addressed by this panel, and I would like to hear people's comments about that, as well.

DR. TRUSSELL: I think a lot depends upon whether you mean immediate switchers or some other definition of switchers, because the discontinuation rates for all of these hormonal contraceptives are extremely high.

So, it is not as if people are getting on them. Many people are getting on them and staying on them for years. There is a tremendous amount of switching.

Prior use of OC's, I think, is completely different from currently using one brand and switching now, today, to another one, and I am not sure what is the question intended to mean, direct

switchers or prior users.

DR. LOCKWOOD: Phill.

DR. PRICE: I would say it would be more focused at prior users and just what you outlined, the fact that they are more experienced using the method.

DR. TRUSSELL: But I think the real issue is on direct switchers because if you have direct switchers, you know that they have used the product for a certain amount of time without getting pregnant. So they are selected for direct switchers.

For prior users, there is a huge population of users out there who got pregnant previously on birth control pills. I don't think prior use has much to do with it. Current use does.

DR. LOCKWOOD: There may be shades. It may be that fresh starts are the least experienced. Prior users, not current users, certainly have experience and may have better compliance and, obviously, current users aren't pregnant, you are

right, and are very experienced.

DR. ESPEY: But I don't think we have any great data that women that have used birth control pills in the past do any better than women who are just starting. In fact, some of the electronic pill data would suggest just the opposite, that the longer you use it, the more pills you miss.

So, I mean, I think to make these kind of arbitrary distinctions particularly given the difficulty of trying to enroll women in trials is maybe arbitrary.

DR. LOCKWOOD: Again, age plays a factor there.

Paul.

DR. HILLARD: I did want to speak to the issue of age, because very clearly, oral contraceptives are widely used in young women under the age of 18. So, if we really are wanting the clinical trials to reflect the population in which they are used, then, absolutely, individuals younger than 18 should be included in the trials, which does add more complexity to the trials.

There are many issues that I will speak to a little later related to adolescents and compliance or effective use, but I think that opening trials up to those under the age of 18 is really important.

DR. TOBERT: I was a little surprised to see hear the question about fresh starts versus switchers, because, I mean, in other areas of medicine, this doesn't seem to be an issue at all. I have spent most of my career with the statin drugs, and I don't think FDA ever cared very much whether a patient had taken a statin before entering a trial or not. So I am not quite sure why this is a particular issue here.

So, I agree with most of the panel that it should be however it works out in the trial, not prespecified.

DR. LOCKWOOD: Getting back to the age issue again, taking greater extremes, so 40-year-olds to 14-year-olds--I may get in trouble with Congress or others--the problem I again foresee is power, that if you have a significant

number at the extremes, both safety in terms of the older age group and efficacy in terms of the younger age group may be lost in the overall analysis.

Even doing subanalysis, I will leave that to the statisticians to comment, but that again gets at the issue of should we have an absolute number, 20,000 cycles, 400 women, should we require substudies in which a focused enrollment with an active group are used to assess these different extremes, extreme BMI, a very young group, and so forth.

DR. LOCKWOOD: Dr. Gillen.

DR. GILLEN: One of the points that was raised earlier was that switchers may be in some sense more experienced or have a better efficacy effect because of that, or more compliance, which obviously is up for debate down there, but again I think a lot of this speaks to basic study design.

If you are talking about a historical controlled trial, then you are worried about fresh starts versus switchers being a confounder; i.e.,

have populations changed as time has progressed. Whereas, if you go to an active controlled trial where you have randomization and you can compare these two groups moving across, then everything should be fine.

Then you only have to worry about the effect modification. So then the question is does your new therapy work better in fresh starts versus switchers. And that is going to be an issue of power and doing the subgroups effect, and I don't even know if it's a clinical plausibility. I mean, some people have raised objections, or, you know, questioned that I would say.

But that opens up another can of worms. So again I think a lot of this can be taken care of in terms of study design if you are just worried about this factor being a confounder in the relationship rather than being an effect modifier.

DR. LOCKWOOD: Apparently, my Boston accent is disturbing my ability to distinguish Gilliam from Gillen, so I am going to point at that person from now on.

DR. GILLIAM: I am very sensitive to the issues of recruitment because it is difficult for clinical trials. But if you do have the information on fresh start versus on continued users, it would be very interesting to see whether there are differences in compliance or adherence, bleeding, and a side-effects profile. So, while I am not sure that you have to require it, I would be very interested in seeing what the data looks like.

DR. LOCKWOOD: Dr. Berenson.

DR. BERENSON: It would be helpful to me if someone from the FDA would speak to what the requirements are for non-contraceptive medications with regards to approval and labeling.

Is it required for an anti-hypertensive to demonstrate that it is beneficial for women and men of all different BMIs? Is it necessary that they prove that it is useful if you miss a bunch of days of the medication, because I feel that we are mixing up theoretical effect in this and actual effect in this.

It seems to me that the burden on the

pharmaceutical company needs to be that they must demonstrate that these products are effective if taken every day, and that doesn't get into compliance, and the compliance issues should not be considered in the clinical designs unless they are in the post-market studies.

DR. LOCKWOOD: Dr. Blumenthal.

DR. BLUMENTHAL: I think Dr. Berenson, I agree with her in that efficacy and the actual effect of the drug, or let's say method failure shouldn't be affected by whether you are a fresh start or a switcher. So this really has do much more with effectiveness or user failure or typical use.

That is much more about behavior, and, you know, the comment about more experience with switchers or with people with previous experience, you know, you can get into bad habits just as easily as you can get into good habits with respect to pill taking, and I think that is what some studies have found.

But with respect to efficacy and actual

method failure, I don't think there is any reason to differentiate fresh starts from switchers. It is useful to get the information about behavior, but I think that relates to user failure and the real world as opposed to just efficacy itself.

DR. LOCKWOOD: One of the themes that I keep hearing is that efficacy ought to be the target and goal of clinical trials, not effectiveness, because we probably just can't fully assess effectiveness across the full range of possible subpopulations and it would be an undue burden to expect the sponsor to be able to do that.

Dr. Johnson.

DR. JOHNSON: Actually, we have already covered it.

DR. LOCKWOOD: Dr. Stadel.

DR. STADEL: Just a little afterthought on we are talking about a representative study population, and people have touched on the fact that you won't get an estimate of effectiveness for different BMI groups, and so forth, unless you went to an extraordinarily large study.

If the study population is representative, it will give you a number that has some real meaning for the population as a whole, and I would raise later the possibility that that effectiveness in some subgoups, like by BMI, possibly could be studied using surrogate outcome of follicle suppression, because I think one is going to have to be realistic about the extent to which you can study subgroups and the overall trial size or the "n" will just escalate.

DR. LOCKWOOD: Dr. Petitti.

DR. PETITTI: I want to return to Dr. Gillen's point, and you will be happy to know as the Chair that this brings us to Question 5.

All of these problems of fresh starts versus switchers and some of the problems that we have on the prior set of questions are really solved with active controlled trials. I think it is actually--here, I am going to be very radical once again--I think it is silly in this day and age to do a trial, a study that we call a trial, and make claims about anything based on historical

controls.

I mean, I just--I don't get it. It is not as if you are going to randomize people to placebo. I mean, you are randomizing them to green contraceptives versus orange contraceptives and, from the point of view of the subject, the randomization becomes an easy recruit.

It is not the same as some of the problems of recruiting people to randomized trials where you are asking them to forego the possible benefits of the drug. So I would say that we should stop talking about approval of products based on historical controls.

I think we are doing an enormous disservice to women by letting products onto the market based on some theoretical number that came from some studies done in, you know, the 1960s on some group of women who nobody can even figure out who they are.

DR. LOCKWOOD: I think that is fair to say that is the consensus of the group. Correct.

Dr. Peterson.

DR. PETERSON: I think the way forward is the active controlled trial. I think the fresh start versus switcher issue is a legitimate question, and we will see it again later with the Pearl Index versus the life-table analysis.

James has made the point I think well in the past that, if you look at somebody who is continued on a method for 9 months or 12 months, their risk of pregnancy is different from somebody that has been on the method 2 or 3 months. But, unless there is a need to develop a stratum-specific estimate for that group than pre-market, it would be the active controlled trial and then. if you need it, post-market surveillance would be the way to go.

DR. LOCKWOOD: I think we have covered that. So, that actually answers the second question and the fifth question, is there a role for active controlled trials, and it looks like under all circumstances is the answer.

The sixth question: Should electronic diaries be recommended for pivotal actively

controlled contraceptive clinical trials?

DR. JOHNSON: Just going back to No. 5 briefly, my concern about the people organizing these trials, is this prohibitive to have active controlled trials? Is it going to make it so much more expensive that fewer new contraceptives are going to be studied?

DR. LOCKWOOD: Comments about that?

DR. TOBERT: Well, I actually gave that some thought. Clearly, you don't want to raise the bar so high it is going to discourage manufacturers from getting in this game and trying to make better contraceptives. I mean, many of these products actually don't have very--they are not blockbusters, the sales are not that big, so you don't want to raise the bar too high.

On the other hand, as I said before, I think you get more information from dividing X thousand patients into active and test than you do by putting them on open. So, I think you could actually achieve this without increasing the burden upon the sponsor.

The point I was going to make--you said, Mr. Chairman, that the panel was in favor of active controlled trials, and that, of course, is quite true. But are we saying that there is no role for a non-controlled trial.

I think I would say that, I think, but I am wondering if the panel would say that, and what would FDA say.

DR. LOCKWOOD: Dr. Stadel.

DR. STADEL: I think a shift to active controlled trials, the time has come for that. I think, historically, it is very easy to understand why uncontrolled trials were used initially with oral contraceptives when there weren't any on the market, and they were coming in. But I think the time for that shift--I do want to say that I think there is a very important issue that the FDA has to deal with if they move to active controlled trials, and that is what is allowed as the comparator. There are a range of products that are approved on the market.

Now, an argument can be made that any

approved product should be usable as an active comparator and a general problem that has emerged over the years with active controlled trials can be the stepdown in efficacy that comes from always using, making the obvious choice if you are in a competitive business and using the comparator to which your products are most likely to look advantageous.

So, there is a difficult issue of establishing a band of acceptability for active comparators or the possibility even of saying that there are categories of OCs, one that has been tested against products with the following established level and the other which has been tested against the lower level.

I am not saying what the answer is, because I don't know the answer. I do want to note here that there is a very important issue that the agency would have to work with the industry to establish a fair playing field for how comparators are chosen.

Thank you.

DR. LOCKWOOD: So, this actually raises the question of what if there isn't an adequate comparison group. Now, we have just introduced a new agent with 10 micrograms of ethinyl estradiol, but, you know, a gallon of norgestimate.

What is your comparison group, how are you going to do an active controlled trial in that context?

Dr. Blumenthal, answer that question.

DR. BLUMENTHAL: Well, I will answer the question I know the answer to. I originally wanted to answer Question No. 6, and that is a one-word answer, which is yes.

I would like to say one thing, again something about the active controls. I think there can be a role for uncontrolled trials. It depends on what you want to know. If all you want to know is a number and you can categorize the group that was exposed to the drug.

Okay. You have got a number and you know who your study population was. But as soon as you start asking questions about what about this and

what about that, what if, then what, then you need comparative groups and it may not necessarily be--and as we have discussed before--it may not necessarily be two different agents, but it may be two different groups, such as a larger BMI group and a normal BMI group.

So, there are different reasons to have, I think, active controls. And the question about the 10 microgram and 10 microgram gallon comparator group, I think that you may have to find the closest substitute or perhaps a pill that is recognized as a generic standard, if you will, and compare from there, sort of the closest generic competitor, and then the new product, but recognizing there is still going to be differences that you can't account for. This was brought up in even some of the materials that we were given in preparation, but you should only alter one thing.

So, either you have a gallon of norgestimate and 20 micrograms, or, you know, 10 micrograms and 2 gallons of norgestimate.

DR. LOCKWOOD: Dr. Slaughter.

DR. SLAUGHTER: Thank you. Actually, the question that I had, or the request that I had, has already been partially introduced, and that I would like the panel to discuss a little bit more about active controls, specifically, the comparator, how would we select a comparator, things like blinding, other ways of conducting the active controlled trial, so if we could just spend a few moments on that.

DR. LOCKWOOD: For housekeeping purposes, I am going to anticipate that electronic diaries, we have a very quick conversation, so we can keep going on this line for about 20 more minutes. I think we probably should.

Dr. Monroe.

DR. MONROE: I think both Dr. Stadel and Dr. Slaughter have introduced the complexity of it, mandating an active controlled trial at least as part of the normal development program for a new agent.

Obviously, in the past at least, the Agency did not consider that a necessary component

for approval, and it is sort of the global position, I believe at the moment, although in the European regulations, they do require some comparison against an active control, but it is more in a small subset to look at certain parameters, more like bleeding profiles, and so on.

Dr. Peterson may want to address it, because I think he is probably more familiar with requirements outside the U.S. than I, but it would require a great deal of thought.

I think Dr. Stadel raised one of the key issues is that there are a myriad of products out there, and to talk about changing just one variable or another is not that easy to do because, not only do we have changes in dosages, we have changes in progestins. We have changes now in dosing regimens going from 21 to 24 to 84, whatever they may be. So the numbers of variations are myriad.

Also, the issue of what would be comparable to a previous product, and you get into issues of are we asking for non-inferiority, and that poses a whole different gamut of challenges in

clinical design.

These are all things that if we can't answer this morning, we do have additional discussion time, and I think it's a very important issue, because, as I said, it would represent a very different way we are developing contraceptive products for approval in the U.S.

But just because a product is on the market, who would be the judge if this is a good product to compare against versus a non-good product. The way a product perhaps performed 30 years ago, at least in clinical trials as they were done then with perhaps a different BMI mean or median that was in that trial, might mean that it would perform today in a manner that we would not be particularly pleased with. Yet, if the new product only performed to that standard, would that be acceptable?

So, this is going to require a lot of thought, a lot of consideration, and I think it should be a topic of discussion. But if we don't reach closure now, we do have time tomorrow, we may

want to readdress this again.

Thank you.

DR. LOCKWOOD: This gets again at the issue of sort of the precedent that just because a particular agent has been approved doesn't necessarily make it a great agent, and you can pick and choose and cherry-pick your control in a way that would potentially make your new agent look quite effective and potentially safe.

Dr. Trussell.

DR. TRUSSELL: I would strongly favor doing active controls, but I would say that if a company wants to just do a non-controlled trial and comes in with a high pregnancy rate, then you say, fine, label it saying that it has this pregnancy rate, and that is going to be a powerful disincentive to taking a gamble. But if you are really damn sure that you have got a great product and you are going to come in with a Pearl of 1 or so, fine, go out and do it.

You can get bit in the butt by taking that risk, but if you want to do it, fine. I mean, I

personally haven't seen, I mean, the randomized trial of the ring against the pill and the patch against the pill. I thought those were perfectly fine comparators. I don't have any problem with it.

I don't understand what the great difficulty is. I mean, if you are worried about pills that were approved 20 years ago, then say you can't use a pill if it was approved 20 years ago, just use one approved within the last X years which you think has a reasonable trial design. There are plenty of them.

DR. LOCKWOOD: Dr. Gillen, with an n.

DR. GILLEN: I think that, you know, the blanket statement had come up earlier that yeah, under all circumstances, maybe we should be doing this, and, I mean, we have guidelines for this; right?

If we go to the ICH, the guidelines for active controlled trials say that, hey, you have to have a comparable active controlled treatment that is truly active within the study population. That

is number one, I mean, you have to have that in order to be doing active controlled trials.

Number two is it is not all roses once you decide to do an active controlled trial. I mean, typically, an active controlled trial is going to be a non-inferiority trial, which means you need to come up with a non-inferiority margin, which is not trivial. That is not a trivial task to decide what is appreciably worse than the active control that you are starting to deal with.

So, I don't think that, once we just jump to this, you know, setting up, saying, okay, we should do all active controlled trials, that everything is going to be taken care of. There is a lot of thought that needs to go into the points have been made, what is the active control, but also what is the non-inferiority margin that we are willing to deal with, and that has to obviously weigh against possible safety and side effects, and things of that nature.

DR. LOCKWOOD: Dr. Tobert.

DR. TOBERT: Actually, just to take up

from where Dr. Gillen left off, I mean, certainly that is something that we are implicitly accepting if we recommend active controlled trials.

I think unless you accept also a fairly wide non-inferiority margin, then you are saying you have got to do huge, huge trials, and that would raise the bar excessively--and maybe Dr. Stadel wants to comment on this--but I think you are going to have to allow perhaps three percentage points in the Pearl Index, although I think that's obsolete, but anyway, or the life-table equivalent. Otherwise, you are demanding huge, huge studies.

DR. STADEL: I think that there is an important issue here about sample size and non-inferiority. I have done some of these, worked with these kind of trials when I was with the FDA, and they do pose some problems.

A couple of thoughts that occurred to me, that I mentioned earlier. I think the possibility of using the surrogate outcome of follicle suppression for some randomized trial work would greatly reduce the sample size requirements.

Also, I think there is a distinction that needs to be drawn between whether you do a blinded trial or an open label trial and what they measure. In one case, you want to measure the inherent difference between the two drugs and you do a blinded trial.

If you want to know the real-world efficacy, that includes how the drug is marketed, how it is packeted, how women are taught to use it, and so forth. So there is a case that could be made for the large open-label Phase 4 trial which establishes the comparative value against another product, and which an open label Phase 4 trial at a large "n" is a much easier issue than a Phase 3 trial that is blinded at a large "n."

So, I would like to suggest consideration of the blinded Phase 3 trial might at least include consideration of surrogate outcome use and that some thought be given to the open-label Phase 4 trial for measuring the bottom-line impact of this product and how it is sold.

DR. LOCKWOOD: I have one question about

that. You would have a control group with the Phase 4 open label? What would your control group be?

DR. STADEL: You randomize to two different contraceptive products including how they were marketed, or you deliver the products approximating how they are sold, yours, here is how you would sell it, and the others, approximately how it is sold.

Now, there are some difficulties to be overcome in that kind of area, but I think that at least thought should be given to it because what effect you get depends not only on what you are giving a person, but how you give it to them.

DR. LOCKWOOD: Dr. Perlmutter.

DR. PERLMUTTER: Most of what I was going to say has been already covered, but one of the difficulties I always have in evaluating products is when you look at the comparisons, you will see that the bleeding effects, the bleeding side effects, are the same with two products. And yet when you look at what they have done, it's a

historical control and then it's a different population, it's a different group, and it is really not comparable.

So, there is a huge difficulty in using historical controls. I am not saying you can't always use it, but I would go for active controls.

DR. LOCKWOOD: Dr. Gibbs.

DR. GIBBS: Charlie, somewhere on the agenda this morning--I don't know if this is the right point--I would like to circle back to Dr. Berenson's point about age.

We spent a lot of time looking at women of high BMI, which kind of ricocheted off the issue of the women under 18. There is a great deal of sexual activity. They need contraceptives also, and I wonder whether we actually reached consensus on that. I didn't hear it.

DR. LOCKWOOD: Is there a consensus that there should be no lower age limit, a specific lower age limit?

DR. PETITTI: Can we come back to that and continue our active controlled conversation.

DR. LOCKWOOD: So, we will come back to that, that's a good answer.

DR. PETITTI: I didn't want to lose the thought in the conversation about age. I am having a hard time understanding why we are making active controls such a huge problem.

I mean, in other drugs' approvals, it is done all the time with drugs which are much--well, I guess we don't do active controls that much--but we put people on placebos and here what we are saying is we just want to put someone on what we think to be an adequate contraceptive.

I am going to make four recommendations or sort of suggestions; first of all, that I agree that in any active controlled trial of contraception, we need to have a large margin for non-inferiority. What we are doing now is we are assuming that there is some magical Pearl Index against which we are comparing everything and we have no idea who really has that Pearl Index when we do an uncontrolled trial.

The second thing is I am going to suggest

three different kinds of possible comparators for the FDA to give as options to companies. One of them would be to use a what I would call "benchmark oral," which would be an oral contraceptive that is widely used in the United States or in the world and that we feel we know a lot about.

The second one would be to have a market basket of orals, which would take the distribution of sales of oral contraceptives and randomize women to the market basket.

The third would be to take a direct comparator where the only thing you have changed is one thing--for example, the 2 gallons of norgestimate or whatever we were doing and the 20 versus the 2 gallons and 10.

DR. LOCKWOOD: I am sorry, the third?

DR. PETITTI: The third are the benchmark oral, the market basket of orals, and the direct comparator. I mean, the single change comparator.

DR. BLUMENTHAL: I think that the last point--and those recommendations are very good. Dr. Gillen and I were having a sidebar about this a

couple of minutes ago, and I think your first suggestion about the benchmark, to me it could be very useful and effective, and sort of from the Agency's point of view, that would be the "do it our way" perspective instead of "have it your way."

You know, like Dr. Stadel was saying that too often the drug companies choose their comparator and choose the one they like best or the one that they think will make the new product look best.

I think from the point of view of amassing a large database, so that we get more and more information about how a new contraceptive performs relative to another one, having a benchmark contraceptive that is current--I think as Dr. Trussell was mentioning, that is current and well accepted, and could serve as a benchmark and really enlarge the database. And the Agency would prescribe which contraceptive or maybe there might be two in the case of a different--you might have one from each progestin group.

That seems to me to be a very logical and

probably useful way to enlarge the database, get good information, and eliminate the choice of the comparator from Pharma.

DR. PETERSON: I think a lot of the discussion in the last little bit is related to the bottom line question of what question are we trying to answer. Most of what we have been talking about is real world effectiveness and trying to estimate that from the clinical trials.

The beauty of the active controlled trial is the issue of comparability. It really doesn't help you a whole lot on the issue of generalizability. I think Bruce's point about the Phase 4, and some other comments about it, that the extent to which we really want that information, and Dr. Gillen's point about effect modification, if we want to know if there is a real difference in effectiveness by body mass index, that is going to have to come later on.

So, if we get back to the value of the active controlled trial to answer that question ultimately, it really gets back to this issue of

comparability and what do you know about the thing you are comparing it to. So, the more understanding you have about that ultimate question for the thing it is being compared to, the more valuable the comparability assessment is.

So, I think it gets back to Diana's point about picking something that you know as much as you can about that ultimate question for generalizability and say, well, let's compare it to that.

DR. LOCKWOOD: Dr. Price.

DR. PRICE: I will pass.

DR. LOCKWOOD: Dr. Tobert.

DR. TOBERT: Following on from Dr. Petitti's comment, I think I agree with nearly everything you have said except for the basket. I think it is very important these trials be done double-blind wherever possible, using the double-dummy technique, which means you have got to pick a single control entity.

I don't think the FDA should mandate that. Rather, I mean, companies should go to the FDA

with a proposal at an end of Phase 2 meeting, and no company is going to suggest some tiny product that has got a 1 percent market share.

Also, active controlled trials are done in other fields of medicine. The one that springs to my mind is antidepressants where, of course, you can't give a depressed patient a placebo, so you have to pick an active control. There are a variety to pick from. I am not quite sure what is picked these days, but you face the same problem and it is solvable.

I would add one thing. I have a slight nagging concern which is the EMEA is still saying that open trials are okay. The EMEA, like FDA, is a sophisticated body and I am wondering why, in 2005, they came out with that recommendation. That was one of the background documents that was cited and maybe somebody from FDA has some insight into this.

DR. PRICE: Question. In selecting the comparator, I would like to ask--intuitively, we think that the 20-microgram tablets are as

effective as the 30-microgram tablets. Does any one of our experts have any concept as to whether indeed the 20's are as effective especially in relationship to method failures?

The balance has always been that we have always thought that the 20's intuitively would be safer than the 30-microgram tablet. Is anyone aware of any data that would suggest or strongly says that the 20 micrograms indeed for what appears to be less effectiveness would indeed give you greater efficacy if you had a higher-dose pill?

DR. LOCKWOOD: Can I also expand on Phill's question a bit, which is should this benchmark concept be specific for, for example, the formulation, a triphasic agent versus a monophasic agent? Should it be specific for the route of application, a vaginal ring versus a pill, extended dose versus extended dose? Just how specific and how many benchmarks should there be, or should there just be one benchmark?

Then if we could also discuss that 20-microgram efficacy issue.

DR. PETITTI: Could I briefly clarify my proposal, because I think when I say "benchmark," since oral contraceptives are the most widely used form of hormonal contraception, I would say that the benchmark would be a benchmark pill, and that is the fallback.

If you don't have any good reason to choose something else, then you fall back on a combined or standard oral contraceptive. I don't care what it is, but probably 35 or 30 micrograms of estrogen and some progestin. It doesn't really matter because you accumulate data on that benchmark, and then, ultimately, you would be able to go to non-active controls, non-active studies.

There would be an alternative to the benchmark, which would be to make an argument to choose something that is a direct comparison, like one vaginal ring versus another vaginal ring, one transdermal against the other transdermal, one extended versus another extended.

So, I mean, you give options. I think there need to be options. I mean, I know long

enough in FDA and the government, as much as you would like to tell people exactly what to do, that that doesn't work, but there be good options, and the industry and the FDA could work something out.

DR. LOCKWOOD: If we could just respond to Phill's question, I will start it by saying that I think there is reasonable data using the surrogate of follicle size to suggest that the inhibitory effects on follicle size may not be quite comparable. But I don't know of any data to suggest a substantial clinically significant--I will use that fudge term--difference in efficacy, but please comment.

Dr. Trussell.

DR. TRUSSELL: I'll reason by analogy. I think the best evidence on whether lowering the dose decreases efficacy or effectiveness comes from the very ancient Oxford FPA study that Martin Vessey did, and there, clearly, 50-microgram pills had a lower pregnancy rate than lower than 50, and it was a huge population to base it on.

I believe that you will find a similar

decrease in efficacy as you keep lowering the dose, but probably you would not find it in even a randomized trial, because it would probably have to be too big to see it.

DR. LOCKWOOD: Dr. Monroe.

DR. MONROE: I was just going to say that I think Question 16 in the next session, this was a topic that we specifically wanted to address. It's one of our specific questions.

We could certainly do it now if you feel it fits in more appropriately, but I think it was Question 16 under Risk/Benefit that specifically is addressing the question that we are talking about at this time.

It is your prerogative certainly, as the Chair, but it may be that the presentations that we would have prior to that in the next session would be helpful in our considerations because. in talking about effectiveness, we are going to get into the issues of Pearls and confidence intervals and life-table analyses.

I think all of these are important

considerations when we try to talk about at least efficacy as we can define it in a limited clinical trial.

DR. LOCKWOOD: Do you want to address this now?

DR. MONROE: You can if you wish. I just wanted to bring to your attention that it is certainly something we are going to address and consider a very important issue that needs to be addressed, so however you wish to do that.

DR. LOCKWOOD: I will just read this. The question is, is there a pregnancy rate that would be--we are talking about Question 15 or Question 16. 16. The question is, should the Division approve lower-dose products that have apparent decreased efficacy and possible decreased risk of serious adverse events as compared to higher dose products--for example, 20 microgram versus 30 to 35 micrograms--so very much in line with what we are discussing in terms of benchmarks, and also what is the tradeoff that we are willing to accept in terms of known effects potentially on efficacy versus

much harder to quantify effects on safety.

Dr. Gillen.

DR. GILLEN: So, this is just going back to the choice of a benchmark and really I am looking for feasibility or advice on this.

Is it feasible--and I may have misunderstood when we first mentioned benchmark--to consider what I would call a time-invariant benchmark, so, at some point, if you come up with an oral contraceptive that is a lower dose that has a similar safety profile, it becomes unethical to randomize people to a benchmark that has been approved later, in the past, at some point it seems to me.

Maybe there is advice out there in terms of how quickly these therapies are evolving for somebody that is not a clinical researcher in the field. Are there thoughts on the idea of a time-invariant benchmark or is the benchmark more so something that is dictated by an authoritative position in terms of what you can be comparing to?

DR. PETITTI: Certainly, we have a history

of allowing to be continued to be marketed, oral contraceptives that we believe we have data that showed they are less safe based on the fact that they are safe enough.

I do think that, given the enormous experience we have with these drugs--I mean, this is probably the most well studied drug in the whole world, you know, in the whole universe, and we know a lot, and we have some products that have been on the market since the 1980s that are still pretty good products.

We have generic versions of them and then we have from the '90s, and you could pick any of, what, 20, maybe 10, that could become a time-invariant benchmark.

DR. GILLEN: So, I guess my question--to me, it raises an interesting ethical dilemma. So, a product may be on the market that doesn't show the best safety profile, but is it ethical to actually randomize somebody to that product.

DR. TRUSSELL: And nobody is going to do it anyway.

DR. LOCKWOOD: Again, what Diana is saying is that she is giving both the sponsor and the FDA some leeway. She is giving a menu of different options, benchmark, you know, most widely used agent, whatever that might be, a market basket of different agents that would be appropriate presumably to match in terms of, you know, they are all recently approved and they match the dose, and so forth, or a really specific match-up where you just change one parameter.

I am certainly no expert in medical ethics, but if an agent is currently being used widely, I don't see that there is an issue with beneficence or autonomy or justice with randomizing patients to that agent.

DR. GILLEN: Well, yes. So, you are making the statement that an agent is currently being used widely, but the key word there is "currently." So, I guess I am asking about how often we would be thinking about having to have this benchmark changed in time. So that is my again concept of a time invariant benchmark.

I mean, is it something that we would be changing as years progress repeatedly, or is it something that would stay stagnant over time.

DR. LOCKWOOD: Dr. Stadel.

DR. STADEL: I think this is the key issue in comparative trials is how the comparator is chosen. I was listening and thinking of my own experience with how difficult it is to get something unestablished in the Federal Government once it is established. So I have some angst about the time-invariant benchmark because I tried unestablishing some of those.

One possibility here is that the Division might want to consider asking for comment or proposals from industry on this issue; that is, what approach should be taken, because if we are going to work here with the industry that is producing these pills, we need to arrive at an approach that we all agree on the consumer side, on the industry side, is a reasonable approach to active comparators.

A shift in the direction of active

comparison trials, I think, is really valuable in this field but I think it is only going to realistically occur if we work out something that is agreeable and operational from a number of different points of view.

So, I would encourage entertainment of proposals, written proposals, about active comparator trials and some special thought if the Division chooses to move more in the direction of active comparator trials--that some thought be given to a specific discussion with industry and others on the various--I think some great ideas have come up, but I think it is a really critical issue.

Thanks.

DR. LOCKWOOD: To take the pulse of the panel, it sounds like we have evolved from the discussion of whether there should be active controlled trials to there should be, but choosing the control is a difficult task.

I think that Dan's approach performs at least the task of forming an excellent place to

start from, and that the industry ought to have a voice in the final formulation of that.

Dr. Trussell.

DR. TRUSSELL: I want to ask a question, because I don't understand whether we are talking about some theoretical possibility or something that actually exists.

Has the FDA had in the last decade a proposal from a pharmaceutical company for an active control that you thought was just inappropriate, and if the answer is no, then perhaps we don't need to spend so much time on it.

DR. PRICE: The answer is no.

DR. TRUSSELL: No, okay.

DR. MONROE: I think, though, there haven't been any proposals recently to do active controls for at least registration here now. Any product that is approved in Europe has had a limited active controlled trial, because the EMEA does require a small trial, a six-month trial, and they are generally looking I believe at endpoints more related to bleeding and hemostatic effects,

and things of that sort, not efficacy in the broad context or large safety issues.

It is a fairly limited trial in scope. The value of those data are subject to interpretation, and we haven't required that. But as far as any recently approved drug in the U.S., I don't recall where a company has proposed an active control beyond that limited study which is required for registration in Europe.

We have, as you have suggested to many, if not most, manufacturers--suggested that, if they did that, it might offer them some protection, so to speak, because you could go back and compare it if you had the misfortune of, for some reason, coming out with a result that made the drug look less effective than perhaps--or more or less absolute standards, because the work is sort of relative absolute standards but are not clearly delineated and it is a question of whether, by going to active controls--and again I think the panel has very clearly mentioned that there are a lot of issues that need to be worked out and there

certainly are merits to such an approach.

But in answer to your question, no, we haven't had that opportunity to say that is not a reasonable comparator. Perhaps, Phill, you would know because you have certainly a longer history than I.

DR. SLAUGHTER: I think that the important point is that the comparator data that we have been presented with has been related to some smaller issues, and not related to efficacy. We have no discussions often prior to these trials coming in at all.

DR. LOCKWOOD: Dr. Berenson.

DR. BERENSON: I just wanted to say that it may not always be possible to compare a ring to a ring, because how would we have gotten the first ring if we had these requirements. What if someone wants an intranasal spray contraceptive next month? We have nothing to compare it to, so we can't necessarily be that strict about what the comparison group will be.

DR. LOCKWOOD: Dr. Tobert.

DR. TOBERT: The Evra patch was compared to an oral contraceptive, which I thought was perfectly reasonable. I don't think you necessarily have got to compare like with like here.

DR. LOCKWOOD: Certainly, when you are starting, there is no comparison group, but what if now another patch is brought to the market, presumably you would want to use the patch.

DR. TOBERT: Yes. I think perhaps one could go both ways on this one.

DR. LOCKWOOD: I want to move on. Very briefly, can we assume there is consensus that electronic diaries ought to be used in pivotal contraceptive trials?

DR. BERENSON: Have they been demonstrated to be more valid in any way than paper diaries, or are they just more technologic?

DR. LOCKWOOD: That is a very good question. Does anybody have an answer for that?

DR. TRUSSELL: No, they have not.

DR. LOCKWOOD: I suspect someone will do a

randomized clinical trial comparing the two. I think that, you know, in the Tom Friedman world of flattening, we are obligated to have a web-based diary.

DR. SCOTT: Charlie, what are they? How do you use them? Exactly what is it? Is it a BlackBerry or something you carry around or what?

DR. LOCKWOOD: That is my sense is that the key is to try to get daily prospective or contemporaneous recording of events rather than doing it retrospectively.

DR. SCOTT: Do you use those in Third World countries as well?

DR. LOCKWOOD: That would be probably more useful in some ways than the paper, given the fact that Third World countries now are linked with fiberoptic cables and microwaves but not necessarily land lines and ways to communicate. Effective postal systems and so forth might actually be more effective in the Third World. But maybe the FDA can better clarify what they had in mind with electronic diaries.

DR. MONROE: Well, I think some of the issues have come up that on the paper diary. You have many limitations. You don't really know, first of all, when the data is actually recorded.

I think at least an electronic diary can help you determine whether it is being recorded within the time frame that you want so you can time-date, you know when it was actually entered, things of that sort, because we all know for those that have conducted large clinical trials, sometimes they are not filled out in real time, they are filled out retrospectively.

Now, whether that leads to more accurate results in the long run, I don't know, and that is the point you have raised, and what is accurate in the big context is not necessarily synonymous with recording an event within a given time frame.

There are many things you can do with it. You can limit the time that you can enter it. You could be more broad, but yet it's time dated, so you can go back and get some idea of the effectiveness.

Some of the other issues, though, if you have electronic diaries--and we have thought a lot about them. I mean, in some cases, companies have even proposed that these diaries alert an individual that they haven't done an entry on that day. Well, then that is alerting them that they haven't taken their pill. So there are many nuances. I think a global statement yes/no is a simplistic and we had hoped--and it depends really on the experience of the panel members that they may or may not have had with such instruments, to get some guidance as to yes, that's good and why would it be good, or no, and why would it not be good.

So, maybe a little more than just a simple yes/no, but again that depends on your experience.

DR. LOCKWOOD: Dr. Stadel.

DR. STADEL: The review by Mishell does specifically recommend that prospective studies of electronic versus paper diaries are needed although they are recommending in favor of them. So, from what I read there, it looks like there is not good

documentation from studies.

They also raise the question--the question ought to be noted that security issues were raised with electronic diaries that don't come up with paper diaries about making sure that the data are correctly secured.

I favor the electronic diaries. I think they probably ought to be studied more. Some of these kind of issues like this one and some of the other issues may be appropriate for sponsorship by organizations like the NICHD Reproductive Health Branch that the sponsor studies if it's not product-specific. We are raising some pretty general issues here, some of which I just note in passing probably are appropriate for consideration by organizations that sponsor research on contraceptive effectiveness and safety generally.

DR. LOCKWOOD: That may be particularly true for issues like BMI, variability, and so forth.

DR. LOCKWOOD: Paula.

DR. HILLARD: With regard to electronic

diaries, I will show a slide in a bit that looks at electronic diaries just with regard to timing of pill taking and number of missed pills, so, the study that Linda Potter published in '96, looking over a three-month interval of how many pills were missed in a given cycle, and a comparison with the paper diaries.

So, we do have that, and I will show that slide in a bit. But that is not the broader issue of recording everything else that would be recorded in a paper diary. Clearly, in this particular study, basically, what happened was when the pill was punched out of the package, the time was recorded, and that just said it was punched out of the package. It didn't say it didn't go down the drain.

It didn't record when it went in her mouth, or if it went in her mouth. But it is better again looking at a comparison between that and what women said about the number of pills that they missed--and, clearly, there is a difference there and I will show that. But I don't think we

have much else looking at the broader question of other data that would be recorded in a paper diary.

DR. LOCKWOOD: Although what you say does suggest that compliance would be better with the electronic diary, and that might affect perfect use.

DR. HILLARD: No, that is not what it shows, and if you want me to talk more about it--

DR. BERENSON: I think the key word on that Question No. 6 is "recommended," and it could certainly be recommended if you felt the data was better. As long as that word doesn't change to "required," because that becomes very strong.

I don't know about the feasibility of purchasing and handing out all these electronic devices because you will not get them back.

DR. HILLARD: That was true in Potter's study, she didn't get it back.

DR. LOCKWOOD: Dr. Gilliam.

DR. GILLIAM: One thing about Potter's study, that is an electronic pill pack, and I think what we are looking for is how do we get the best

possible data. For example, with diaries, people can wait until the end of the week and fill out all the paper diaries, or the end of the month and fill it all in, and people are using things like did they change the color of the pen that they were using to guess at that information.

So, I think intuitively we think that an electronic monitoring would be better. My only concern is that, in diverse populations, technology is not always the same for all people, and it is not always as readily assessable, so we just have to think about the down side for certain populations.

DR. LOCKWOOD: The consensus there is recommended, but not required.

Question 7. The Division has typically used premature termination rates as an assessment of patient satisfaction in clinical trials. Would information obtained from validated patient reported outcome instruments be more useful in contraceptive trials?

DR. BERENSON: Absolutely. You have to

look at the reasons the patients went off the medications. It could be many different barriers to using their medications other than patient satisfaction.

DR. PETITTI: Is there existing a PRO that has been validated?

DR. MONROE: Probably not, and when you say "validated," it means validated for what specifically. We did include a very lengthy document in the background document. Dr. Lockwood has brought that to my attention on multiple occasions just to show the scope, and it is in your package. It is how many pages, Dr. Lockwood--

DR. LOCKWOOD: At least 5,000.

DR. MONROE: --that the Agency has been requesting to, quote "validate" something. So, in answer to that, there isn't anything yet that has been accepted to be validated from the Division's point of view, and validation of a PRO instrument, if you follow I guess current standards, is not a trivial exercise. It is quite complicated.

DR. ESPEY: This seems like something that

would be extremely useful. Just as we are going to talk about the standardization of vaginal bleeding language--I mean, there is nothing out there. For those of us who have tried to do this kind of study, there is really nothing out there that you can compare across different studies, because there is no standardized instrument.

I mean, I think if this were an outcome of this meeting, to create something like this, that would be hugely useful for researchers and for women.

DR. LOCKWOOD: I would say that, while there is no validation necessarily in the context of contraceptive therapy, there has been substantial work done in validating different aspects of the instrument.

So, for example, scales--which scales work better in which settings, and so forth. But I do think that these would be far more useful than just relying on termination as a reflection of patient satisfaction with the agent.

Dr. Blumenthal.

DR. BLUMENTHAL: I agree with Dr. Espey in that I think that the more standardized an instrument you can get, that everybody buys into, the more uniform the data are and that means more comparable in the long run anyway. So, if it is a charge to the Agency to lobby Congress to get us a validated instrument, I think that would be useful.

This actually could even relate to the concept of what we were talking about before with the electronic diaries, because again software being somewhat difficult to create, when you move to electronic diaries, there is probably a lot less variation in the format of these diaries, because ultimately, everybody will buy one diary system, and data across the board, no matter which company, no matter which product, will be a lot more uniform. So that will also enhance our ability to compare products across studies and within studies.

DR. WESTNEY: I just want to comment from the field of urology where, of course, it is critical, I think, to have a validated instrument for whatever it is that you are looking at. It gets

to the second step where there are 10, 20, 30 different instruments which are all validated and then studies are using whichever instrument that they prefer, and, even though you still have some measure of patient preference, it still doesn't solve the problem of looking at different studies because they are utilizing different instruments.

DR. LOCKWOOD: It might be another opportunity for a market basket.

I think the consensus is that if such instruments were available, they would be extremely useful, and that the ball is back in the FDA's court to perhaps help develop that instrument or lobby for research in that area, and so forth, but that it seems that going beyond just a termination rate, determine satisfaction would be very useful.

The 8th question is: Could a validated PRO instrument, Patient Reported Outcome instrument, be used to obtain secondary labeling claims for superiority, for example, better cycle control?

Dr. Gilliam.

DR. GILLIAM: This is sort of addressing

both 7 and 8. I think validated instruments are very important, but we have to realize their limitations. They may provide internal validity to your study, but we still have questions of external validity.

So, if it has been validated in a specific population, we can't assume that it applies to all populations. So I think that applies to Item 8, as well, that a claim of superiority in a study that has internal validity does not necessarily mean that you would be able to make that claim without the external validity, as well.

DR. LOCKWOOD: By definition, assessing cycle control requires patient reports. There is really no other way to do it. So, they are all PRO instruments from that perspective, so it is being done anyway.

I think we are going to get into more detail about specifically what standards ought to be applied for defining cycle control a little bit later.

Dr. Perlmutter.

DR. PERLMUTTER: You have just usurped what I was going to say because, until we have better definitions of what our bleeding patterns are and what our side effects are, I don't know that we can answer this question.

DR. LOCKWOOD: We have reached the point where I think a break would be in order for several reasons, one having to do with urology, and the other having to do with the fact that I suspect the next topic, which is the discussion of life-table analysis versus the Pearl Index, may get contentious or may not, but may consume a lot of time.

So, with that, we will take about a 15-minute break.

[Break.]

DR. WATKINS: Next, Dr. Trussell's presentation.

Topic 2 - Efficacy and Risk/Benefit Assessment

DR. TRUSSELL: Thank you very much. I will confess in advance that Dr. Gillen and I were not given the opportunity to collude about what we

were going to talk about, so we don't know what each other is going to say.

In fact, none of here who are giving talks have any idea who else was talking or what they were talking about.

[Slide.]

In the next few minutes, I want to cover these issues about measuring contraceptive efficacy. The first is efficacy versus effectiveness. The second is typical versus perfect use. The third is the Pearl Index versus the life table. Non-completion of a trial and the effect that has on interpretation. Common errors in the literature. Results from the literature. And then communicating the risk of failure to clients.

[Slide.]

This is a review for many people here, but just so that we are all on the same page, efficacy measures, how well a method works under ideal conditions, and effectiveness, how well it works in the real world.

Efficacy would typically be measured in a clinical trial whereas effectiveness would be measured in survey or chart review, or something like that.

[Slide.]

Where do we have data? In the United States, our data come from the National Surveys of Family Growth, which were conducted in 1973, 1976, 1982, 1988, 1995, and then most recently in 2002.

These have the advantage that they are nationally representative, unlike any clinical trial would ever be. They have a disadvantage that they are retrospective. In particular, women are asked for each month, going back in the past for five years, which contraceptive method they used. Now, if you are like me and you had a vasectomy many years ago, that is not going to be very hard. But for many people this can really be a problem.

If you actually look at these data, which I have been doing with colleagues from the Guttmacher Institute, some reported patterns of use are just unbelievable--a month of condom use,

followed by a month of pill use, by condom use, by pill use, and there are every unimaginable combinations of these.

So, going back in time and remembering what one actually did each month can be a problem for some women. There is definite underreporting of abortion. Fewer than 50 percent of abortions are reported in the National Survey of Family Growth, and there may be overreporting of a contraceptive failure leading to a birth because it is our natural tendency to blame something on something rather than on ourselves. So, in fact, there may have been no contraceptive used even though it was reported in that month.

Clinical trials have the potential disadvantages of the Hawthorne effect, which is named after the Hawthorne Electric Works, which basically says that people might behave differently when you are observing them. That is hardly surprising.

Therefore, inference beyond the trial setting, even if they didn't behave differently,

does the population that you are studying have anything to do with the population that is actually going to use the method, which we talked about at some length already today.

Cycles of perfect use can be identified and pregnancy rates during perfect use can be estimated but, of course, adherence is self-reported. So, we only know what people tell us they did.

[Slide.]

Well, let's see about the underreporting of abortion. We see here from the last NSFG, not the 2002, but the 1995, these are uncorrected and corrected for underreporting of abortion.

It doesn't make a lot of difference for the pill, but it makes a whopping difference for spermicides when you add back in the estimated underreporting of abortion. Of course, that is tricky trying to get that right because the data are coming from two different sources.

[Slide.]

Self-reporting of adherence; we have

already heard about this. The classic study by Linda Potter, where there are self-reports on missed OCs compared with electronic recording on punched pills for 103 women for 3 cycles.

Now, these were birth-control pill packets where when you punch the pill out, it recorded the date and time. The results are absolutely disheartening. There was agreement in less than 50 percent of the days on whether a pill was punched out between what was reported on the paper diary and what the computer punch-back said.

There was overreporting of no missed pills, probably not surprising. No missed pills were reported by 53 to 59 percent of women on the paper diaries, but, in fact, only 19 to 33 percent of women had no missed pills.

There was underreporting of missing 3 or more pills, 10 to 14 percent on the paper diaries versus 30 to 51 percent on the electronic ones.

So, adherence, tough, tough to measure.

[Slide.]

Typical use versus perfect use. A

contraceptive failure during typical use can be measured in the clinical trial or in a survey.

It just means contraceptive failure during perfect use has been measured only in clinical trials since retrospective reporting of adherence in surveys is likely to be just terrible. It was terrible even in that group of women who were studied with electronic pill packets when they were supposed to be measuring it each day. But think back four years ago whether you used or missed a pill in a particular month could be problematic.

[Slide.]

What is typical use? By definition, a woman is a user whenever she considers herself to be using the method. Hence, typical use of a barrier method does not imply that it is actually used at every act of intercourse, and typical use of the pill could mean that I ran out of pills last month and I haven't started this month because I don't have a prescription but I am still using the pill.

So, typical use includes both inconsistent

use and incorrect use, as well as perfect use.

[Slide.]

In contrast, perfect use of a method requires actual use according to the directions for that method.

Perfect use of a barrier method, for example, would require that it be used correctly at every act of intercourse.

Perfect use does not imply no pregnancies, just that the rules were followed.

[Slide.]

Now, typically, what happens in clinical trials is that pregnancies are divided into method failures and user failures, but next is where the error occurs.

Suppose that, in a contraceptive trial, there are 100 years of exposure to the risk of pregnancy, and there are 15 pregnancies that occurred during a cycle of imperfect use and 5 pregnancies that occurred during a cycle of perfect use.

What is the method-related pregnancy rate,

the pregnancy rate during perfect use?

[Slide.]

The traditional answer is 5/100, or 5 per 100 women years of exposure. But that is wrong. There is a logical error here.

The denominator cannot be all exposures since, by definition, a method-related pregnancy can occur only during perfect use. So, we have to know how to divide the exposure as well as divide the pregnancies.

If there are 50 women years of perfect use, then the correct answer would be 5/50 or 10 per 100 years of exposure.

Now, this is a very common error in the literature. In fact, if you look in the January issue of Contraception, you will find a paper that makes this error again.

[Slide.]

So, there is a flaw in the design of clinical trials in addition. Information on perfect--correct and consistent--use is typically obtained only for cycles when pregnancy occurred.

Indeed, women in some trials are interrogated mercilessly to find out whether, in fact, they actually used the product or not.

But if you don't find out, except in pregnancy cycles, whether there was a perfect use or not, then you cannot estimate the perfect-use pregnancy rate.

[Slide.]

So, I will give a simple example here. The green cycles are those of perfect use, the red cycles are those of imperfect use, and the P means the woman got pregnant. The correct way to do this is to say that, during perfect use, there is one pregnancy, but there are 15 cycles, so the perfect use pregnancy rate is 1/15, and the typical failure rate in this example would be all or 2/18.

So, this is not a difficult concept, but the error persists in the literature.

[Slide.]

What about the Pearl Index versus the life table that we have alluded to earlier today? The Pearl Index is the pregnancy rate. It is

pregnancies per 100 women years of exposure, and it ranges in theory from zero, if there are no pregnancies, to 1,300 if every woman got pregnant in the first cycle of use. So it does not go between zero and 100, it goes between zero and 1,300.

But much more importantly, it is a rubber yardstick. Women most likely to become pregnant do so early leaving behind a pool increasingly consisting of the more compliant and the less fecund.

So, if you wanted to get a very good Pearl Index for your product, you would run your trial forever. If you could afford to do it, that is what you do, because you would drive it right down to zero.

Now, life-table methods produce estimates of the percent of women becoming pregnant within specific durations, like 6 months or 12 months or 24 months since initiating use.

The problem here is if you are going to compare across trials, there is not a problem

really. If you are going to do a randomized trial of pill A versus pill B, you can use the Pearl Index. Everything will work out just fine. No problem at all.

But if you want to compare a trial that ran for 6 months with a trial that ran for 12 months, then the trial that ran for 6 months should have a higher pregnancy rate even if the pills and the women are identical.

It also comes up in a different manner that was touched on this morning. Suppose that you have direct pill switchers. So, I have been using pill A for 9 months and I enter the trial and start using pill B. Now, I am really in my 10th month of exposure and, if I were doing a life-table analysis, I would be entered in month 10, not in month 1.

If you enter me in month 1, it is going to make my life table look much better. Likewise, if you are comparing historically, a Pearl Index from a pill when there are only fresh starts to a pill today, where 75 percent of the women have already

been using the pill for durations of zero to 15 months, the the Pearl Index is going to be lower for the current pill because those most likely to get pregnant are gone.

[Slide.]

Now, the Pearl Index again is a rubber yardstick. I and my colleague used the same data and obtained pregnancy rates of 7.5 and 4.4 during 100 women years of condom use.

One, me, who got 4.4, allowed each woman to contribute up to 5 years of exposure, whereas the other, my colleague Jane Menken, got 7.5, allowed each woman to contribute only up to 1 year of exposure.

Now, who is correct? There is no correct here. They are both correct. They just are measuring different things, and you cannot compare a Pearl Index from a 6-month trial to a Pearl from a 5-year trial.

[Slide.]

What about non-completion of a trial? Well, in an ideal world, all women would either

become pregnant or complete the trial without becoming pregnant. That is the goal. No trial has ever been conducted where that was the happy outcome.

In fact, a high fraction stop for other reasons. They could be lost to follow-up. They could stop for medical reasons. They could stop for personal reasons. Mostly what people do is focus only on the lost to follow-up, but that is also incorrect.

The problem is in everyone who does not complete the trial. What is the consequence?

[Slide.]

Well, in a life-table analysis, those who are censored--that is, those who just disappear. they are lost to follow up or they quit because they don't like the bleeding profile--they are assumed to have the same failure rate as those who are observed had they remained in the trial. So, what you observed is assumed to be what you would have observed had those people stayed around.

But that may not be true. In fact, if

they are more likely to get pregnant than those who remained in the trial, then you are going to get an underestimate of what the true failure rate was.

Now in a Pearl Index, it becomes more complicated. If those who are censored would have had a higher risk of pregnancy, the Pearl Index could be biased upward or downward.

[Slide.]

Here, I just give an example. If you work through the math of the consequence of those who leave the trial, having a higher risk of pregnancy than had they stayed in the trial than those who stayed in the trial and, in this case, the life table is certain to be biased downward. And it is. But the Pearl Index actually goes the other way.

[Slide.]

What about factors that influence failure? Well, one important one is the inherent efficacy of the method. IUDs are inherently more efficacious than spermicides. It wouldn't matter who you tested then on, they are going to be better than spermicides.

The next most important factor is imperfect use, the extent of which will depend on the motivation to avoid pregnancy and how easy it is to use the method perfectly. Although spermicides, you have to use them at every act of intercourse, that is not true of an implant which, once you put in, you leave it in and you don't have to do anything.

Frequency of intercourse makes a difference and it does decline with both age and marital duration. The plot of frequency of intercourse by age is what my colleague Charles Westhoff calls the saddest curve in the world. It looks like a train going off a cliff.

Now, efficacy will also depend upon the underlying level of fecundity. If we do a trial of only 49-year-old women, we will get a lower failure rate than if we do a trial of 21-year-old women. It also depends upon the competence or honesty of the investigator.

[Slide.]

What are some common errors? Well, the

incorrect calculation of method failure is rampant in the literature. Other little errors include multiplying cycles by 1,200 instead of 1,300 to get pregnancies per 100 women years of exposure, not including the learning phase in your contraceptive trial where all the bad ones can get wiped out, can get pregnant, or discontinuing non-adherent women. If they are not good laboratory patients, then you get rid of them.

[Slide.]

Common problems. A high percent not completing the trial is a really big problem and it would be common for less than 50 percent of the women to actually complete the trial--that is, to make it all the way to the end without becoming pregnant or become pregnant. Underreporting of abortion. Use of the Pearl Index when comparing risk of failure among methods where the trials were of different lengths.

[Slide.]

Clearly, there are problems in comparing methods. The results can come from different

sources. Where available, I take them, during typical use, from the National Survey of Family Growth adjusted for underreporting of abortion.

Another huge problem is, of course, that women choose which methods to use and are not randomly assigned to methods. Women who choose to use spermicides are very different from those who choose to use IUDs.

[Slide.]

These results are summarized in each edition of Contraceptive Technology. If I look at the master table from the next edition, which is in press, then we would draw conclusions that all clinicians know.

[Slide.]

Methods regarding adherence generally show a big difference between perfect-use and typical-use failure rates. The most effective methods during typical use are those not requiring adherence and the most effective methods are not those that protect against sexually transmitted infections.

[Slide.]

How about communicating the risk of failure? There have only been two studies that have examined how well do women understand contraceptive failure rates and how to communicate contraceptive effectiveness, and that result is a chart that appears in the new WHO Global Handbook for Family Planning shepherded through by Bert Peterson, and the next edition of Contraceptive Technology and I will just leave that with you.

[Slide.]

It has four bands of effectiveness. It doesn't attempt to distinguish between methods in each band, but it appears as though women can pretty well capture that those in the top band are the most effective and those in the bottom band are least effective. It is a convenient way to summarize a complicated result.

That is the end of my presentation. I don't know whether you want to take questions now.

DR. GIBBS: Dr. Trussell, this is a small side point. Do you have information as to whether

underreporting of abortion varies by country? Is it more underreported in the United States than, say, in Western Europe?

DR. TRUSSELL: It is underreported everywhere that it has been measured. You can measure underreporting only if you have another source to know accurately what the number of abortions actually is, and the source in the United States comes from surveys of abortion providers done by the Guttmacher Institute. But in every study of which I am aware, there is underreporting of induced abortion.

DR. SCOTT: Dr. Trussell, I listened to your presentation and the way the results are presented. I wonder if you could make a case for a very simplified way of reporting results assuming you can go to the accurate data to physicians and patients.

In other words, why not just say the perfect method where you have documented they took all the pills, 3 pregnancies in 100 women in 2 years. Imperfect use, actual use, 8 pregnancies in

100 women in 2 years, or something comparable, that people could actually understand.

DR. TRUSSELL: Well, we have tried to do that in each edition of Contraceptive Technology with that master failure rate table, which the Division certainly knows. This result comes from an empirical study of what women could actually understand.

It turns out that the most difficult to understand is a table. The easiest to understand is a picture, and this is what we finally wound up with, but it was a long process.

DR. SCOTT: I think it has been well shown that odds ratios and confidence intervals, and so on, are very confusing for physicians to counsel patients, and so on. I think it has been shown that most people understand, whether it's a number needed to treat, or whatever it is, per 100 patients or per 1,000 patients, and so on.

I am just wondering whether some sort of a system could be devised that way.

DR. TRUSSELL: That is what we think we

have done here. It is true--unambiguous in the literature is that people understand 3 out of 100 rather than 1 out of 33 where the denominator keeps changing. So, it is important to have a uniform denominator and that is what is tried to be shown in this picture here.

DR. TOBERT: Dr. Trussell, I think most of these errors come from the literature, which predominantly, as we have discussed, has been with uncontrolled trials. So, how many of these errors would potentially remain if you do a randomized, double-blind trial, analyze it using life-table methods and the intention-to-treat approach.

Would that solve all the problems or would it create new ones, and are there some problems that would still be there?

DR. TRUSSELL: If you used intent-to-treat and you didn't try to report method failure rate, then the error in the method failure rate wouldn't occur, but there have been randomized trials where the method failure rate is incorrectly calculated in the standard wrong way.

The really famous biostatistician Mindel Shepps wrote a paper once in which she concluded that the Pearl Index--I will paraphrase it--is completely useless and measures nothing that one can be certain of. Nevertheless, the Pearl Index in a randomized trial is going to be just fine in comparing method A to method B.

DR. TOBERT: Isn't it the Pearl Index is the crude incidence rate?

DR. TRUSSELL: With a randomized trial, you don't need powerful tools. A numerator and a denominator is just fine.

DR. TOBERT: I don't know. I always thought crude incidence rates were pretty much obsolete these days if you are measuring any kind of outcome. I guess we will get to that later with Dr. Gillen.

DR. TRUSSELL: It is an incidence. It's the number of pregnancies divided by exposure.

DR. TULMAN: I would like to get back to the issue of this picture here. Looking at it, it seems there is quite a bit of leeway and that, on

the top row, we have less than 1 pregnancy per 100 women year use which still involves a mathematical interpretation on the part of the woman, going down to the less effective where it is 30 pregnancies.

But one of the things the picture doesn't help us to understand is, is it that the second row is 2 per 100, or is it 20, or is it 29, going down to 29 and a half going down to 30. It doesn't tell us since it is essentially a ranked data type of thing. It doesn't tell us what the steps down are among those four levels, and it doesn't tell us within that category just how comparable they all are. So, is the lactation method the same as an injectable, or is there some differentiation?

DR. TRUSSELL: What you wind up with--we go back to it, and I publish it in every edition of Contraceptive Technology--you get the table and the problem with it is that the people who read the table couldn't understand it. Even after looking at it, they could not tell you whether the IUD or the condom was more effective.

So it was certainly not possible for the

humans that were a part of this WHO Working Group to devise something that conveyed all of this information which could be understood.

DR. TULMAN: Could you tell me a little bit more about who this was tested with?

DR. TRUSSELL: Yes. The first test was done in the United States and the second test was done outside the United States, and it was really the crudest. It wasn't asking for a really in-depth understanding of what it showed. It was simple questions like which is the more effective, the condom or the IUD after looking at the chart--before looking at the chart and after looking at the chart. The picture worked better than did the chart.

Then there was one test in which we tried to show people both typical and perfeyct use information, and that did the worst.

DR. TULMAN: Yes, I understand that. But, I mean, in terms of the sample, what was the average level of education or sophistication or whatever? Who was in the study I guess is the

question I had.

DR. TRUSSELL: In the United States, it was intended to be the population of people typically who would use contraception.

DR. ESPEY: I think this gets to the whole issue of health literacy, which is getting more press these days and, despite education levels, you know, health education is more difficult to communicate. But it is really clear that pictures are on the hierarchy of what people understand are way up there.

MS. SHANKLIN-SELBY: Are the people who this is intended, are they aware of what--I mean, is it explained to them what typical use--I mean, do they know is this perfect use, or is this kind of putting in like some leeway for--

DR. TRUSSELL: These come from the typical use figures, and the way that the perfect use was intended to be conveyed is how to make your method most effective, which is the extreme righthand column over there.

MS. SHANKLIN-SELBY: So that would

address, I mean, that they understand that they have to follow these directions. That is where that is addressed as far as any forgiveness for--

DR. TRUSSELL: Well, yes. I mean, how to make your method more effective, take a pill every day. For the pill, it is a pretty reasonable idea.

MS. SHANKLIN SELBY: So, that seemed to be what worked the best then.

DR. TRUSSELL: Yes.

DR. ESPEY: Well, James, presumably, you are actually talking to these patients, as well as just giving them--

DR. TRUSSELL: I don't have any patients. I am not talking to them, and Bert can explain. I mean, this is a global handbook for clinicians or family planning providers; correct? And this is a chart that is suggested they could use for their clients.

DR. WATKINS: Any other questions?

[No response.]

DR. WATKINS: Then we will move to Dr. Gillen's presentation.

DR. GILLEN: Thank you.

[Slide.]

So, the old adage is it's the job of every good statistician to wander into the light-spreading darkness, but I will try to refute that. So, we will see how I do.

As Dr. Trussell had said, we didn't get a whole lot of chance to communicate in terms of coming up with this so you will see a little bit of duplication here. I guess I could say ditto minus the vasectomy. Sorry, but I will go ahead and go through this anyway, and we will see where we get.

[Slide.]

When I was asked to present here, the first thing I did--you know, I mean, all my research is in clinical-trial design in general. So, I kind of went to the way of contraceptive trials just as I would with any other trial and this was the first slide of my clinical trials course to students; what are the minimum requirements that we need for a clinical trial.

Well, we need an appropriate target

population and the three that I am going to really concentrate on are the use of appropriate comparison groups, the use of an appropriate outcome measure, and the ability to maintain statistical criteria for evidence. So, talking about p-values and Type 1 errors, Type 2 errors, talking about power effectively, 1-minus power.

[Slide.]

So I am going to run through each of those last three kind of in order here. I am going to start actually with the outcome measures and talk about Pearl Index versus life-table methods, and show you a couple of examples. These are things again that Professor Trussell has alluded to in his talk, as well. We will just beat the Pearl Index to death while we are up here and just go ahead and get it out of the way now.

And then I will go on to comparison populations. In particular, something that has come up obviously early on in the morning is historical versus active controlled trials, and I will give some thoughts on each of those, and then

finally defining statistical evidence. So, the concept of testing for superiority versus non-inferiority is something that we have to consider if we are going to go to active controlled trials.

[Slide.]

So, first, with outcome measures.

[Slide.]

So, again, the Pearl Index is the number of pregnancies per 100 women years. It's a common measure. It has been used to summarize contraceptive effectiveness. However a drawback of the Pearl Index is that, in most situations, it is dependent upon the length of follow-up, on time, as Professor Trussell had just mentioned.

I am going to go through a quick example to kind of spell out exactly some of the issues that can go wrong, and this is a pretty mild example actually.

[Slide.]

I am going to consider that I have got two populations or two groups that consist of my study

population. So I have a low risk group that comprises 90 percent of the population, and, just for simplicity, I will assume that they have a constant risk of pregnancy. So, the one-year probability of pregnancy is 5 percent.

Then a high-risk group which comprises 10 percent of the population, which has a constant risk of pregnancy with that one year of probability of pregnancy being 50 percent.

[Slide.]

Let's think about what happens when we calculate the Pearl Index in this particular situation. So, the expected number of pregnancies is going to be my 90 percent times the 5 percent of those getting pregnant, and then the 10 percent in the high-risk population times the 50 percent probability of getting pregnant in the first year. Multiply that times 5,000 and I have 475 expected pregnancies.

The expected person years at risk--assuming that I am censoring at pregnancy, I am going to go ahead and assume that pregnancy

occurs uniform over the year. So, on average, people that become pregnant contribute half of a year into my study over that first year.

So, I have 4,525 individuals that contribute the full year. Those are the individuals that did not get pregnant again--I am going to assume no dropout here--age contribute one year. Then the 475 individuals contribute half of a year. So, my Pearl Index then is going to be 9.97 pregnancies per 100 per year in that case.

[Slide.]

Now, what happens when we move to calculating the Pearl Index over 2 years? Well, we need to consider the impact of censoring. That's the whole idea here; who is in the risk set is what we say in survival analysis.

So, by the end of the first year, the number left in the low risk group is going to be 4,275. Again, 5 percent of them, on average, will become pregnant or an expectation will become pregnant.

The number left in the high-risk group is

going to be approximately 250 in expectation, so now they only comprise about 5.8 percent of my sample going on from year 1 to year 2 rather than the 10 percent that they comprised early on.

[Slide.]

So, if I calculate the Pearl Index over years 1 and 2, so now I have come up with my new sample sizes. I expect to see 344 pregnancies between Year 1 and Year 2 with an expectation of 4,352 person years over that time. So, now my Pearl Index over Year 1 to 2 is now 7.92. So it has dropped by 2 at this point.

[Slide.]

Then, when I go to do the cumulative over 2 years, again, the expected number of pregnancies is just the sum of the first and second year. The expected number of person years at risk is the sum of person years at risk between 0 and 1 year, and 1 to 2 years.

So, now, my Pearl Index calculated over the 2 years is roughly 9 so it has dropped by 1. Again, as Professor Trussell noted, send your trial

out to infinity, the Pearl Index will drop to zero. You only have people that aren't going to become pregnant at some point.

[Slide.]

So, when is the Pearl Index independent on studies for support? I mean, so when can we actually interpret it as being a time-invariant measure? When will it not change?

Two cases. One is the rate of pregnancy is homogeneous across all possible subgroups in your study population. Not going to happen.

Number 2 is this rate remains constant with time, which I assumed in my previous example. But again, it's a pretty big stretch.

[Slide.]

One thing to note is that, in my previous examples--so some might say okay, well, if you want to keep the risk set to the people at risk in your trial to be relatively consistent, well, one thing I could do is go and identify a high-risk person and bring them in each time one of my high-risk individuals has a pregnancy.

Can't do that, because we can't identify them in general. Okay, that's not possible. The other thing is, well, maybe I can put them back into the trial after they have given birth, et cetera, et cetera. That is not going to happen either. They won't have contributed time at risk at that point. Okay. So, I can't even go with that. I am still going to have this reduction in Pearl Index as time moves along.

[Slide.]

Another issue--again, for the non-statisticians in the house, I apologize for the next few seconds, but I have to rant and rave, because another issue with what is going on with the Pearl Index is that confidence intervals in general are calculated incorrectly.

There was just a 2003 paper on the European Journal of Contraceptive and Reproductive Health that discussed how one would calculate a confidence interval for a Pearl Index and they said, well, okay, assume the Poisson distribution.

That is a very particular distribution.

It assumes that the mean--so the mean rate of your Pearl Index is equal to the variance, and the variance is how we quantify uncertainty in our outcome measure.

Well, it turns out that rate data typically get characterized by stemming from what is called an "overdispersed" Poisson distribution; in other words, the variability that you observed is bigger than the variability that you would assume by a Poisson count. So, you are more uncertain about that particular estimate.

How does that arise? Well, it arises by having mixtures of populations. Again, if people have different underlying rates of pregnancy, you don't have a single Poisson distribution. You have a mixture of a bunch of different Poisson distributions. So my last example had two.

[Slide.]

So, in our previous example, let's just go back and see what kind of an impact that some of the calculations people have been making can have.

So, again, I have got my low-risk group, which

comprises of 90 percent of the population, and I have got my high-risk group, which comprises 10 percent of the population.

[Slide.]

So, if I go to the Pearl Index over 1 year again, remember, the true Pearl Index, what I would expect to see is 9.97 pregnancies per 100 years. Let's suppose that I ran my study and I actually observed 457 pregnancies over that one year, and I observed 4,763 years of follow-up. So I calculate my Pearl Index and I get 9.6 pregnancies per 100 per year. Okay.

[Slide.]

Well, if I assume that Poisson distribution, which has been advocated in the literature, in the EMEA, then I get a 95 percent confidence interval that runs from 8.73 to 10.51, and, if I take into account the fact that I have got a heterogeneous population, it turns out that my variability is about 20 percent larger than I had assumed it would be with the Poisson distribution and it's just wrong.

[Slide.]

So, we have underestimated the variance and that translates, for all of us that like to read confidence intervals, to meaning that the confidence interval is shorter than it actually should be. It doesn't have the correct coverage probability is what we say.

So, it turns out that the true 95 percent confidence interval, or a correct or consistent 95 percent confidence interval, runs from 8.63 to 10.55. So it is about 8 percent wider than the previous interval.

Now, some of you are saying, well, okay, those numbers don't look that different, et cetera, et cetera. But this is about the impact of doing corrections for interim analyses in clinical trials, and we definitely demand those. So it does have an impact on what we are doing particularly when we are studying superiority and non-inferiority bounds.

[Slide.]

So, how do we deal with the fact that the

risk set is changing because that is really the problem with the Pearl Index. We need to take into account that our risk set is changing as time moves along because people are dropping out of the study that have different baseline risks.

Well, survival analysis actually kind of conquered this quite a while ago, and Potter was the one that led to a lot of this work in contraceptive trials. We just consider conditional probabilities.

So, I changed my function of interest from the rate to a cumulative probability over some period of time, and I acknowledge that follow-up time is part of what I am trying to estimate. So if I want to talk about T being the time of failure, then I can say okay, what is the probability of failing--i.e., an unintended pregnancy within the first two years.

Well, that is 1 minus the probability that existed in the past years without a pregnancy, and it turns out that statistics can handle that. I can just condition it upon the fact that I survived

past the first year. So, that incorporates that changing risk set. It says, hey, let's just take the people that are still here with us and recalculate that probability at that point.

If I do that, I run through and I get roughly 17 percent.

[Slide.]

So, that guy is, in fact, called a life-table estimator. That is the whole point of the life-table estimator is just to condition upon those changing risk sets.

It turns out that in contraceptive failure trials, most of our conditional probabilities are typically--that is a very crude life-table estimator. It is done at each year, the one that I just showed. Mostly, we do these at each monthly or each cycle in order to more accurately incorporate the changes in the risk as time moves along.

[Slide.]

I am going to refer to this guy--this is the way most of your statistics refers to it--as

the Kaplan-Meier estimator, and the Kaplan-Meier estimator is equivalent to the life-table estimator.

Let me just bring those intervals down to each time somebody has a pregnancy. That is called the Kaplan-Meier estimator. That is what I will refer to it as from this point on.

[Slide.]

So, one of the questions that came up was, well, are there any benefits of using the Pearl Index. Well, it has been in wide use for a long time. Okay, so why has it been? We need to examine that.

Well, the real reasons I believe are, number one, people like the ease of accrued rate interpretation. But I think that if you start looking at the Kaplan-Meier estimator and just talking about the probability of an unintended pregnancy over a given period of time, that is quite an interpretable or clinically relevant parameter as well. So, we can overcome that with practice as we learn to teach people to interpret

these statistics.

For historically controlled trials, well, we have a good deal of data that summarizes Pearl Indexes, so it gives us some sort of a reference. Well, it may not be the right reference. We have already talked about that. But are we truly estimating the same Pearl Index from historical controls that we will be today. That is an issue, but this is one of the reasons. When people point to actually putting forth the Pearl Index, they say, well, I have data that talks about the Pearl Index. Well, maybe or maybe not--depending upon cohort effects. Are you comparing apples to oranges?

So, again, that is going to change as the popularity of the Kaplan-Meier estimate grows, as well. We will have more and more data on Kaplan-Meier estimates, which are again interpreted according to times of follow-up.

[Slide.]

Another question that had been brought up in the Backgrounder was to say, well, can we

incorporate changes in the treatment regimens. That is a very interesting question. I pondered it for quite a while actually. So, the idea is that patients can discontinue use or use additional contraceptives for some interval of times. So they go off treatment and then they come back on later in time; can you recover them back into your study.

Well, technically, yeah, you can just put them back into the risk set and bring them in as intervals as time moves on. It is just not clear to me, though, how one should make a judgment as to when to reenter them into the risk set and I will give you a quick example.

[Slide.]

Let's take somebody that uses a back-up contraception between an interval of times, say t1 to t2--and I am going to assume that zero is my start. So, t1 is bigger than zero here. So, this individual could be considered, then, at risk for the interval from zero to t1 and then reentered back into the risk set at time t2. So they are just interval censored, is what we say, between t1

and t2.

However, when we do that, we implicitly make the assumption that that person's hazard, which is the way we define a risk of pregnancy or an instantaneous risk of pregnancy, at time t2 is the same as everybody else that has been at risk from time zero to t2. To me, that is not a reasonable assumption.

There are reasons that people go off a contraceptive. There are reasons that they go onto an additional contraceptive. I wouldn't necessarily make the blanket statement or blanket assumption that they actually have the same hazard as an individual that has been on treatment all the way from zero study start to t2.

[Slide.]

So, my short answer to the incorporation; well, I think in the trial phase, no, we could do some sort of post-hoc analyses, after things are said and done. I wouldn't use it as my primary analysis or my primary study design. But we could do Kaplan-Meier estimates where we stratify for

individuals that have gone off treatment and estimate different conditional probabilities or different life-table estimates.

We could also go into a regression framework, and some of you are probably familiar with the Cox proportional hazards model. That would allow us to allow people to go into different treatment groups as time moves along and actually estimate what their relative hazard would be if they went off treatment versus on treatment

Again, that is data modeling, so that is not something that I would, a priori, propose in a clinical-trial setting because you are actually going to have to go through, model the data. It is not prespecified.

[Slide.]

One thing that comes up is regardless--and this was already mentioned in Professor Trussell's presentation--is regardless of the measure you use, you have to define what a failure actually is and who is at risk.

So, for all new interventions, we need to

consider safety--in other words, are there adverse events that clearly outweigh any potential benefit--efficacy--can the intervention reduce the probability of unintended pregnancy in a beneficial way--and effectiveness; i.e., whether the adoption of the intervention as a standard reduced the probability of unintended pregnancy in the population.

[Slide.]

So, one difference between evaluation of efficacy and effectiveness is in what defines the failure and who should be included in the risk set. We have already talked about this a little bit.

In the clinical-trial setting, we really can't truly evaluate efficacy because of possible selection bias of the patients that are entering our trial. However, it may be useful to evaluate--and I am really just quoting one of Professor Trussell's papers from Contraception in 2004--intervention failures, rates during actual use including inconsistent or incorrect use, and intervention failure rates during perfect use.

[Slide.]

You were just given a very nice presentation to say okay, well, if you are going to go with the perfect use method, you really need to consider who you are including in the risk set. So, in other words, I don't want to include, at risk time, where someone isn't actually at risk for a method failure under perfect use.

So, I have to think about when an individual is actually at risk for the particular event that I am considering.

[Slide.]

So, historical versus active controlled trials, something that we have already had a bit of debate about here.

[Slide.]

So, again, in the past, many methods have been assessed via historical controlled trials. From the Backgrounder, some of the Pearl Indexes quoted were 1.5 or, more recently, a Pearl Index of 2, have been used for the efficacy criteria and, again, such criteria stems from experience of

historical controls.

We have talked about this already. Lots of bias can result from using historical controlled studies, particularly in this particular setting where samples are not comparable with respect to baseline risk or covariate values that are running across different groups, the evaluative measure of outcome, how are you following up on patients with respect to failure, and duration of study. So, if you are comparing Pearl Indexes, are you comparing apples and oranges by comparing your Pearl Index of 2 years versus a Pearl Index of 1 year.

[Slide.]

So, if we are going to go with the historical control--I am just kind of laying out the pluses and minuses of all of these things--if we are going to do a historical controlled trial, one of the big things that I would push is that we really need to acknowledge uncertainty of the estimates.

So, the EMEA requires sufficient sample size to guarantee the width of the 95 percent

confidence interval for the Pearl Index to be no larger than 1. That is how they are defining a sufficiently powered study. But they don't mention where the bound of that confidence interval needs to be in relation to a point estimate. It really is just looking at the point estimate.

It is better, in my opinion, that you require, possibly in addition to this efficiency in terms of power and sample size, to require that the upper bound of the confidence interval is less than an observed threshold if you are talking about the Pearl Index.

So, what alternatives have you ruled out? What Pearl Indexes have you ruled out with your particular study? In either case, that is the whole reason I put the last section in on correct computation of confidence intervals. If you are going to use a Pearl Index, and you are going to be using a confidence interval to define superiority and make a decision, you need to be correctly computing that confidence interval.

[Slide.]

Because it's impossible to guarantee comparability between historical controls and current study samples, it is almost always advantageous to employ randomization when ethically feasible. So that is my standpoint.

Of course, we can't do a placebo-controlled trial here. However, we can and should at least consider the use of an active control when comparable interventions are already in use.

The nice thing as well that goes along with having the randomized trial, if we are just talking about benchmarks with respect to the life-table estimator, we can talk about the cumulative probability of failure over one year, the cumulative probability of failure at two years.

If we have a randomized trial and I have all the data with me so I know exactly when each person failed, I can compare the entire survival curves or the entire failure rate curves, if you will, over the entire period of follow-up. It has been well used. Oncology uses log-rank statistics

all the time, and we can go into a proportional-hazards framework if we want to do a covariate adjustment, as well.

So that is another plus that we get from the randomized experiment is we don't have to just use a benchmark of a single number to compare back to historical control. If we have all data on both groups, we can actually compare those Kaplan-Meier estimators over the full length of follow-up.

[Slide.]

So, if we are going to go to active controlled trials, we have to think about the difference between a superiority trial and a non-inferiority trial. So I just want to lay out some of the issues there.

[Slide.]

So, the statistical criteria for evidence in a superiority trial; well, that is evidence to rule out equality of effect as measured by the chosen parameter. So your chosen parameter might be the Pearl Index, a one-year survival estimate or, if you are going to do the actual comparison of

survival curves, it might be a hazard ratio. So compare the instantaneous risk of death over all times, basically average them.

Our contrast; let's, for example, say that we have a one-year difference in failure rates as measured by the Kaplan-Meier estimator. So, here I have got the Kaplan-Meier estimator computed at one year for the treatment group minus the Kaplan-Meier estimator for failure and the active controlled group at one year.

So, negative here would be good in terms of the treatment, say a smaller failure rate at one year. So, a classic hypothesis that we would be testing in the superiority phase then would be, okay, difference in the Kaplan-Meier estimators of failure at one year, greater than or equal to zero, versus the alternative of less than or equal to zero, and rejection of my null hypothesis deciding in terms of efficacy for the treatment versus active control would correspond to an upper bound of my confidence interval for that difference being less than zero.

[Slide.]

So, what is the difference when we go to a non-inferiority trial? Well, a non-inferiority trial means that we need evidence to rule out some margin of efficacy less than the active control, we are willing to accept something.

So, let's go back to our contrast of 1-year failure rates. I have to define what we call "non-inferiority margin" now, some delta that I am willing to accept in this particular trial.

Again, we are going to have a little discussion of this, I would imagine, but that delta is not trivial. You have to take into account safety, the risk/benefit profile, secondary endpoints, and how you are performing on those, as well.

In this case, rejection of the null hypothesis would correspond to an upper-bound confidence interval being less than--I apologize, cut-and-paste has gotten to me--that should be a delta sitting right there. So, if anybody is taking notes, change that to a delta, and I will

change it for the official slides.

So, in other words, you have ruled out a difference of delta or greater with your confidence interval in that study.

[Slide.]

So, we have had a little discussion already about when is it reasonable to consider a non-inferiority trial instead of a superiority trial. Well, again, I go back to the ICH Guidelines.

First of all, we need an active comparable--I should put in there active controlled treatment that must be truly active in the study population. If the active control is truly active in the study population, I ask myself two questions when I am going to do a non-inferiority trial.

The first question is can I define a margin to define non-inferiority to be established. I have to be able to do that if I am going to go into--superiority is easy, zero versus naught. But can I define a clinically relevant non-inferiority margin?

The second is that if the active control is a standard of care--which a lot of us are dealing with--then is the new treatment also superior on secondary endpoints? So, is it a lower dose, we expect a better safety profile, et cetera, et cetera, so that we are willing to accept some small deviation with respect to efficacy and in contrast or benefit for the secondary endpoints and possible safety profiles.

So, if the answer to either of those is yes, sure, a non-inferiority trial, in my opinion, can be performed ethically.

[Slide.]

So, what are the issues then that we need to consider in setting that non-inferiority margin? Well, what measure compares the distributions? We have already talked two of the commonly used ones, Pearl Index or Kaplan-Meier life table.

Is the treatment effect random? Okay. So, are you see different treatment effects in different populations as you are going through? How much of a decrease in the effect is acceptable?

Again, that is often a hard one to quantify.

Then how do we account for variability in the estimates from our historical controlled trials because that is going to come into leading us into this non-inferiority margin.

[Slide.]

So, what is some of the precedence that has been set in different trials that I have worked on and those that I have been a part of? Is the treatment effect random? Well, ideally, you can use meta-analysis data for multiple trials.

You have to be careful here, though. If you are talking about something like a Pearl Index, do all the trials have the same duration of follow-up? Do they have reasonably generalizable patient populations? Are you just measuring two different parameters--in other words, often trials coming up with different estimates--or are you trying to estimate the same thing and noticing different variability across groups.

How much of a decrease in effect is acceptable? Well, I have been on trials where we

have considered 10 percent decreases in the active control effect all the way up to 50 percent. It depends upon what the safety profile is, what the advantage is on other secondary endpoints.

Then how do we account for variability in the estimates from historical controlled trials? One thing, if you want to be very conservative, is use the worst-case scenario from a historical and 95 percent confidence interval--that's the most conservative--or explicitly account for variability in historical controlled trials if you actually have the data at your hand. That is another possibility.

[Slide.]

So, just a quick summary.

[Slide.]

We need to define an appropriate target population, comparison group, outcome measure, and maintain statistical criteria for evidence.

With respect to the outcome measure, the Pearl Index is usually, almost always I would say, implicitly dependent upon the length of follow-up

whereas the Kaplan-Meier estimates make that dependence explicit, and you are talking about the conditional probability at a particular point in time.

In either case, we need to obtain correct inference and the definition of the risk set needs to correspond to the definition of failure. You can't be included in the risk set if it is not possible for you to have an event--I mean, if you just completely exclude it from the numerator.

When ethically and logistically possible, I advocate active controls for a lot of the reasons that we have discussed already.

Again, I put some prefaces on here. I don't say that all the time everyone should be doing an active controlled trial, but when it is ethically and logistically reasonable, we should be looking at that option.

If historical controls are going to be used, then we need to account for uncertainty in terms of defining the superiority criteria through the use of confidence intervals, hopefully

correctly calculated confidence intervals.

DR. LOCKWOOD: Questions on presentation?

DR. GILLEN: Spread the darkness?

DR. TRUSSELL: Because this is a question and because you addressed it, it seems to me that you can do a sensitivity analysis by including people who go off treatment and come back on treatment. You can enter them in your example. You have a choice of entering them at t1 or t2, and you can do it both ways and, at least when I have done it, it hasn't often made a huge amount of difference.

DR. GILLEN: Yes; and certainly the stratified analysis can help you to assess that. I guess my issue is, if you are doing trial design, in the protocol, you need to specify right upfront what you are going to be doing with people as they go off treatment, as they come on. I personally, a priori, in a trial design situation, would feel uncomfortable saying I am going to assume that this person's hazard is the same at time t2. So, I am going to reenter them at t2, or is it the same at

time t1, because I am assuming some sort of non-informative censoring that is going on there.

So, from the trial standpoint, in prespecifying exactly what I am going to do, I would feel a little bit uneasy specifying upfront that that is exactly the assumption I am going to make.

DR. TRUSSELL: But, if you have a non-informative censoring, then, no matter what you do with them, it's wrong.

DR. GILLEN: What is that now? If you have informative censoring, you mean.

DR. TRUSSELL: Correct.

DR. GILLEN: Exactly. Yes; if you have informative censoring, all survival analysis methods without a doubt rely upon the assumption of non-informative censoring; in other words, the reason you are leaving the trial is not indicative of when you are going to fail.

So that is going to be an issue. But do I think that those people have the same hazard even, or do I think that they are somewhat inherently

different for their reason for leaving or using a back-up contraceptive.

I can't, a priori, say, or at least I don't feel comfortable, a priori, saying, yes, they should have the same hazard as those who have been in the trial and being compliant at the same time.

DR. GILLIAM: When you talk about when someone reenters into a trial, I can make sense of that if I think about a 1-month injectable. But I have more difficulty understanding how you would judge, for example, pill use when people could miss one or two. There seems to be a subjective factor to even judge when treatment restarts. How would you do it then?

DR. GILLEN: That is a scenario that I haven't even gotten into is how do you actually measure compliance in real time--I mean, and this is something that you guys are talking about with the diaries that we have been discussing.

I think that is a very, very difficult scenario. You know, this question was raised more like a generalization. Let's suppose in the

best-case scenario we knew exactly when somebody goes off treatment and exactly when they come back on, do I leave them out and bring them back in.

Even in our best-case scenario, I feel uneasy doing that, because I am making these implicit assumptions about their baseline risk.

DR. TRUSSELL: It is more, Melissa, you have got a pill trial and you have got women who, on certain months, use, in addition, condoms. Or you have a barrier contraceptive trial and you have use of emergency contraception. What are you going to do with those people?

DR. GILLEN: And I guess what I am advocating is saying that those people are censored at the time that they go off, you know, this prolonged exposure to an additional contraceptive or coming off of their current contraceptive.

DR. TRUSSELL: And the down side of that is that you rapidly run out of people in your trial.

DR. GILLEN: I know, without a doubt. I mean, certainly sample size and censoring here is

going to be an issue. The down side of the other is that you may be estimating the wrong quantity. So, you know, you have plenty of people to do it, but it may not be the correct thing.

DR. PETITTI: I have a question for the FDA.

About how long are these trials currently scheduled to run? Are the subjects recruited supposed to come in and stay for six months or a year or two years, or is there some standard?

DR. MONROE: Most trials, they are recruited with the intention that they stay in for a year. Sometimes they are longer. But for an oral contraceptive, it has generally been with the expectation that they would be in for a year.

DR. PETITTI: And the reason for the one year?

DR. MONROE: I don't have any scientific rationale. It has just been the way it has been done for many years. Some trials, as we say, will continue, they may allow them in for much longer.

I think, obviously, both you and Dr.

Trussell have raised the fact that that will have a significant impact presumably on the efficacy assessment based if you see something as simplistic as a Pearl. But again I don't have a scientific rationale for that. It has been deemed along the way that that was a reasonable time.

For much longer intervals, I think--well, I mean, there are some reasons for not going much longer is that many of the adverse events of interest tend to occur within the year, particularly--at least a small number of thrombotic events, and so on. I think a lot of data show that you are at an increased risk during that period of time. But I can't give you any more information than that.

DR. GILLEN: I have actually dealt with this question a lot in survival trials, in general. I guess the basic idea is that clinically, we are interested in long-term efficacy of trials.

So, the way a survival--this is considered a survival trial. Survival trials means that, basically, an event occurring, a time to an event

is your primary outcome of interest.

The way that the power is dictated for a survival trial is in the number of events that you observe. So, if I am trying to estimate the time to unintended pregnancy, if I have got 10,000 people but I haven't observed any unintended pregnancies, I have zero information on the time to unintended pregnancy. I haven't actually observed it on anyone at this point.

So, you know, you can power a survival study in two ways. Let's say I need 500 events or unintended pregnancies to obtain adequate power for my study. Well, I can obtain that by bringing in 100,000 people and following them for four days or I can obtain that by bringing in 1,000 people and following them for three years.

I am estimating different things with those. You know, am I estimating efficacy over three days, which I may not clinically care about, or am I estimating clinical efficacy over three years, which may be more of a gold standard. So that is part of what has to go in in terms of

defining the length of follow up for a trial. Again, the physical power is no longer dictated by the number of people in your study. It is dictated by the number of events that you actually get to observe.

DR. PETITTI: The reason I raised the issue is because I think this situation is actually quite different than a typical trial, the kind you are talking about.

I mean, here you know that you have a pregnancy hazard which is not in any way constant over time, and you also know from your experience that you have problems that affect the validity of the estimate that you are primarily interested in, which is efficacy, which get worse the longer you follow people because they become less and less compliant and you have all these dropouts.

You also know that you are actually not getting any data on safety. So I would say that some of the things that we have heard here argue for--at trials that are shorter with more people, although, of course, you have to balance the cost

of recruitment versus the cost of follow-up. But I would wonder--and, of course, there are lots more people who do actual trials in this field in this room.

I would think that, in this situation, the cost of follow-up for any given subject might exceed the cost of recruitment.

DR. GILLEN: I guess the one thing I would say is that we actually deal with the two scenarios that you describe in oncology trials all the time. We have changing hazards over time, and we have different lengths of care and which we deem clinically relevant. So, we do deal with this in other trials.

The one thing I will say is that, from a patient standpoint, you are not looking to take an oral contraceptive for only six months and you are not looking to know what the efficacy is over necessarily six months.

If there truly is a time-bearing effect of that oral contraceptive in terms of how long you are on it, that may be a compliance issue or

something else, but from a patient standpoint, you may want to know what the long-term benefit is.

So, you are right. I mean, logistically, maybe it is cheaper. But, at the same time, one has to weigh that against the clinical relevance of what they are truly measuring.

DR. PETITTI: But that would presume that you are using a life-table method of analysis and presenting data where you show the hazard as a function--or failure rate as a function--of time. Given that the Pearl Index is presented, those people actually know nothing about their long-term outcome.

So, if the field doesn't change from a Pearl Index to a life table, then running longer trials is actually useless.

DR. GILLEN: Well, I mean, that Pearl Index is really an average of each of those short-term intervals. I mean, it is a crude average, I agree, it is not a good summary measure.

DR. PETITTI: But you also trade off the withdrawal and the dropout and the non-compliant.

DR. GILLEN: That's right.

DR. LOCKWOOD: Just one comment. It may not be necessarily cheaper because you have to have more people over a shorter period of time, and it may actually be more expensive.

Dr. Tobert.

DR. TOBERT: I wonder if the one-year tradition came about in part because, as we have heard, that is a way to drive down your Pearl Index by going on for longer. If you do an active controlled trial, then you can have a shorter time. It doesn't matter really whether you use a Pearl Index or a life-table method because you have got a control.

So, companies usually like to have shorter trials. So, I think this is yet another reason to have an active control and use the life-table analysis.

DR. LOCKWOOD: Dr. Trussell.

DR. TRUSSELL: The same issue was discussed about a decade ago in the Devices Division and it was driven by the very clear

reality that people didn't use these devices for a year. I mean, half the people had quit by six months.

So, it is increasingly common to get six-month trials for devices.

DR. LOCKWOOD: Dr. Gibbs.

DR. GIBBS: Our conversation this morning has been noticeably devoid of reference to sexually transmitted diseases and other consequence of sexual behavior.

So, a national priority is to encourage safe sex. My question is, in design of trials, how does good trial design encourage safe sex and how do you account for safe sexual behaviors and measures of oral contraceptives.

DR. GILLEN: My short answer is you go into a trial with a primary endpoint. That is your primary endpoint, and you have secondary endpoints in mind. So, you would effectively be treating sexually transmitted diseases as a secondary endpoint if you are truly concerned with efficacy being unintended pregnancies in that case.

It is nearly impossible to power trials and do adequate inference. The one thing that statistics is horrible at--and we have been doing it for a long time and we are still not any good at it--is the multiple comparison problem because we don't know the correlation between tests.

So, we don't know the correlation between your tests on secondary endpoints and the efficacy endpoint in many situations. We can make assumptions about that.

DR. GIBBS: It was really unfair of me to ask you that question. It was really a clinical question. Rather than having STDs as a secondary endpoint, I am talking about prevention of STDs, which basically is going to mean barrier contraceptive in addition.

So, has that aspect been incorporated into trials?

DR. LOCKWOOD: I think we are going to talk actually more about what happens when more than one method is used in the analysis, and that is coming up.

Dr. Stadel.

DR. STADEL: That was a very nice discussion of the full range of the statistical issues. I have a simple question that comes from having worked on drug labels and trying to communicate between people from different backgrounds.

Suppose you did a comparative trial. By the way, I think one year has a certain appeal in terms of communicating commonly to people. People think in terms of what happened last year. So, over a year is not a bad way of communicating. You have to have something as a primary endpoint.

DR. GILLEN: Absolutely.

DR. STADEL: And you have got to have some way you can make the cut. Suppose you took as a primary endpoint--you do a randomized, comparative trial, and you take as the endpoint the total pregnancy rate among people who complete one year with consideration of whether the dropout rate was meaningfully different between the two arms.

That is a review issue and it becomes

somewhat of a qualitative judgment ultimately, and was it acceptable or not. But suppose it is acceptable, that there is not a big dropout. So you have a pregnancy rate among people who complete one year, and then you have the question of whether it is significantly different between the two arms according to the established criteria.

That gives you one endpoint to function at, one that I think most people would find fairly easy to understand; what was the rate of pregnancy among people who completed a year or six month or nine months.

DR. LOCKWOOD: How is that possible, I mean, if you are pregnant, you can't complete the--so how would that work?

DR. GILLEN: That's complete. I mean, you have follow up on that person, so they are no longer censored that year. You know when the actual pregnancy occurred.

DR. LOCKWOOD: Other questions?

The new plan, so that we are full of energy and enthusiasm to deal with the Pearl Index

versus life-table analysis question and others, will be to break for lunch now and reassemble at about 1:10.

(Whereupon, at 12:10 p.m., the proceedings were recessed, to be resumed at 1:10 p.m.)

AFTERNOON PROCEEDINGS

[1:15 p.m.]

DR. WATKINS: We will jump right back into Discussion Questions Part 2, and that is on Contraceptive Efficacy Assessment.

DR. LOCKWOOD: I thought maybe we could start with a perhaps very brief lay summary by Dr. Trussell and Gillen of the issues involved--with no discussion of any statistical endpoints--of why it is important, why the issue of how to deal with dropouts--and potentially this would also apply to people that use multiple contraceptive methods during a prolonged interval of observation--why that affects life-table analysis.

A very eloquent series of statistical formula were put up there. But I am an obstetrician, so perhaps that could be described in a succinct fashion.

DR. TRUSSELL: The issue of dropouts is really very simple; it's what do you assume about what would have happened to the people who dropped out had they stayed in the trial.

Now, you could assume that they all would get pregnant, in which case your pregnancy rate will be 80 percent so that is not going to be a very useful thing to assume.

You could assume that none would get pregnant. Or you could assume that they get pregnant at exactly the same rate as the people who stayed in the trial and it is that assumption that is the one that is always made. It means that the reason that those people left the trial didn't have anything to do with whether or not they would have become pregnant. They just sort of randomly leave.

That is a different problem than what do you do in a trial where you have people who use contraceptives other than the one that you are studying. The two cases that have come up most frequently would be the use of condoms among people who are using them for protection against sexually transmitted infections.

You try to guard against that in the trial by saying that the people who come into the trial shouldn't be at risk of sexually transmitted

infections. They should be monogamous. But, of course, that makes them not like the rest of the population. So, if your goal is to make them like the rest of the population, you would like for them to be like the rest of the population.

Even if you say that, you still have to counsel women that, if they are at risk of sexually transmitted infections, they should use condoms. Then, the question is what do you do with them because if everybody, all the time, used a condom and used a pill, you should get a very good trial result.

So, that's it, and there isn't a magic answer about what to do. There are several things that you could do. You could say, okay, the first time a woman uses a condom, she is out of the trial. You censor her at that point. The same would apply to emergency contraception because people are using that, too, even in trials of hormonal contraceptives. When they miss pills, they can use emergency contraception. That is one suggestion for handling them.

Another is just to recognize that this is the way the product is going to be used and sometimes people will use condoms as well as pills, and that is just what you get, because the real world is that way.

DR. LOCKWOOD: And don't analyze it any differently.

DR. TRUSSELL: Don't analyze it any differently. You just look at it is an intent-to-treat analysis. The problem is again considerably lessened if you have an active control because condom use ought to be the same in the two groups.

DR. LOCKWOOD: Okay. So, let's address the first question, No. 9, Pearl Index versus life-table analysis, what are the relative merits. I think a lot of that has been covered. I didn't really hear the relative merits of the Pearl Index, but maybe somebody can reiterate that. I guess the relative merits honestly are simplicity and ease of presentation.

Are there situations where one approach

should be favored over the other, and, if so, what are they, and how should divergent pregnancy rates, calculated by the Pearl Index versus life-table methods, be considered in the approval process and in labeling?

DR. TRUSSELL: I think it is time to retire the Pearl Index. I mean, you could not publish a paper in an academic journal with a Pearl Index with the single exception of Contraception, where it is still done. But, I mean, statisticians abandoned this years and years ago. Why should we keep doing it?

DR. GILLEN: Just one other point. I mean, if you have such divergent rates between those two, you have either one serious mixture of distributions that is going on in your study that is leading to this in terms of the changing hazard rate, or you have a time-bearing treatment effect that is going on over time, and both of those things should raise a little bit of red flag in terms of what is happening here if you see that contradiction. I mean, if they are that divergent,

you have more issues.

I mean, the Pearl Index is a very crude summarization of that rate over a given time period and, if you don't have at least consistency in terms of point estimates going one direction or the other in those two methods, there is definitely a subset of population that this is not acceptable in--I mean that something is going wrong.

DR. LOCKWOOD: I think we all get the latter that clearly, rates of pregnancy tend to aggregate toward the beginning of the cycle, more fecund and the higher frequency of intercourse, less compliance, and so forth. So, there is sort of natural selection and thus, if you are recording a disproportionate number of cycles in one trial compared to another, the Pearl Index would be less valid. But in a controlled trial, an actively-controlled trial, why would it matter? That is one question.

The second question is the first part of your argument against it. I can't get my hands around, hazard-ratio issues, et cetera.

DR. GILLEN: It is simply if the risk--let's assume that a contraceptive method for the first month was perfect but then its efficacy started to decline over time.

So, as time marches on, people are having higher and higher probabilities of becoming pregnant. Then, you start to average those things with the Pearl Index. So, you are not taking into account--you are just basically giving me a crude measurement of what is happening across time exactly.

DR. LOCKWOOD: The opposite argument.

DR. GILLEN: Exactly. Exactly.

DR. LOCKWOOD: So why, in an actively controlled trial, would it matter?

DR. TRUSSELL: It doesn't in an actively controlled trial. But, if you want to accumulate a sort of database of things, then, why not have the life-table or--I mean, you can get the whole survival curve. The probability of getting pregnant by one month, two months, three months, four months, five months. I mean, any first-year

graduate student can calculate that.

[Laughter.]

DR. LOCKWOOD: Okay. It sounds like there is really no one here willing to put their neck out to try to salvage the Pearl Index, so I think that is the end of that conversation.

Let's move on to the next question which is--and it sounds like there aren't any situations and people would prefer the life table if there were divergence.

Question 10: How should divergent pregnancy rates obtained in the U.S. and non-U.S. populations be considered in the approval process and in labeling? That sort of gets a little bit at what we were discussing before regarding the acceptance of non-U.S. studies.

DR. TRUSSELL: I would see it as somewhat different. I mean, you can do two trials if the FDA recommends that you do two trials in certain circumstances. You could do two trials in the United States and get two different answers. In fact, you would never get identical answers.

So, it is hardly surprising. If you look at the summary of the literature on the same product, all the clinical trials that have been done, you get numbers that are all over the place, in part due to poor trial design, but in part due to the fact that you have got different people in the trials.

So, as long as the FDA is going to consider--again, I think all of this goes away if you have an active control. If the FDA continues to consider all OCs as the same, then, you do an active control, you get equivalency, and you don't put in the individual pregnancy rates from the trials. You could. But you are still declaring all pills to be the same unless you change your mind.

DR. GILLEN: That is the question. Are they all the same.

DR. LOCKWOOD: Dr. Johnson and then Dr. Pettiti.

DR. JOHNSON: Actually, I was back on the Pearl Index, but my only question actually was to Dr. Gillen, and I will go ahead and pose it because

he made it very clear to me. I asked him this earlier.

If we go from one to the other, if we go from the Pearl Index to the life-table analysis, then, can we use the old data? Is all this data that we have--is it still valuable? Can we use it? Can we compare new studies to old studies? What kind of problems would that raise?

DR. GILLEN: In order to go back and compute the life-table estimates based upon historical data, you would have to have individual level times as censoring for individuals so that you know exactly when they are in the risk set at each individual month, for example.

If you had that data, then you could reconstruct the life-tables from them. But oftentimes what we have are summary statistics that is coming from each of those trials.

DR. JOHNSON: So, it is really going to be starting from anew.

DR. GILLEN: I think that there is some precedence for having data on life-table methods in

recent trials certainly, so there is some historical data that we could probably get a hold of in terms of full Kaplan-Meier estimators of what a curve looks like over a particular support.

DR. TRUSSELL: Many of them are published.

DR. GILLEN: Yes.

DR. LOCKWOOD: But, again, if you had an actively controlled trial against an agent that you do have the Pearl Index for, then you would have that data,

DR. PETITTI: I just wanted to--on this Question No. 10, I think when you see a "divergent" pregnancy rate, you have to ask three questions. Is it due to something that has to do with things you can control, like differences in the duration of the study? What we have heard is that you will always get a different pregnancy rate when you have studies which are different durations because of this problem of the hazard being not constant. That is controllable. You could always have every study have the same duration. I am going to get away from the active-control idea because I don't

know if that is going to--I think we should think of both.

The second one is whether or not you always have a different pregnancy rate when calculated either by the Pearl Index method if you have different dropout rates over time because you have again a time-dependent variable, and you have a hazard rate that is dependent on time.

Now, you cannot control dropout rates and you cannot make them comparable. But when you have divergent pregnancy rates in the U.S. and non-U.S. populations, you can ask the question of whether or not the differences are explained by these two things and, by doing a life-table analysis for the U.S. and the non-U.S. populations, you can determine whether or not those two things are contributors.

The other thing that I want to know is--you really have to, of course, ask whether they are different due to just differences in sample size and then I think that this is the place where there might be prespecified subgroup analyses

comparing the U.S. and non-U.S. populations according to characteristics that you believe might be modifying the pregnancy rate and particularly body mass index. So, that should be a prespecified subgroup analysis whenever there are studies that are going to enroll populations that might differ on that variable.

DR. LOCKWOOD: Dr. Stadel.

DR. STADEL: Dr. Petitti just said what I was going to say. Dr. Petitti just addressed really what I was going to address, but I will just say it again. First off, I think that, since the FDA has jurisdiction in the United States, that U.S. data takes primacy if there is a conflict between data from U.S. and otherwise--that is just my opinion--and that adjustment for covariates, if it explains the difference; for example, if one takes the foreign data and adjusts it for body mass, standardizes it to the U.S. data, or something like that, if that essentially explains the difference in findings, then one has an answer. And, if it doesn't and one has a real conflict

between U.S. data and foreign data, then, I think the U.S. data takes primacy.

DR. LOCKWOOD: Dr. Gillen.

DR. GILLEN: I think, you know, there was a little hint earlier about the difference between random variability between study results and inconsistency between study results. I mean, when the FDA is requiring two confirmatory trials, they are looking for consistency of treatment effects across study results in order to generalize to the population.

If you start seeing divergence in the sense that you have point estimates going in different directions, that starts to make people worry because that either means that you have confounding in a sense where you have differences in your covariate distributions across populations where your treatment is not looking nearly as good in one group than the other. That means you have got to look at those subgroup effects.

Number two, if you do have consistency, then at least it is telling you that, in some

sense, maybe one is underpowered versus the other, et cetera, et cetera, and you are talking about some sense of random variability. But you are still looking for that point estimate to be at least within the ballpark of range and to be consistent with the first trial.

DR. LOCKWOOD: No. 11. So the next question is should "on-study pregnancies" be defined to include only those pregnancies that occur while subjects are within the treatment cycle or also include those pregnancies that have an estimated date of conception, that may have occurred within a certain number of days at the end of hormonal therapy, 2, 5, 14 days, where the treatment cycle is defined to include pill-free intervals following active treatment.

So, for example, using that 14-day rule that seems to apply currently, 7 days presumably would be counted as still part of the treatment day, and then 7--you are sort of given a 7-day grace period where there should still be some residual contraceptive effect.

Is that reasonable? Is it evidence-based and, if not, should it be varied. Should, in fact, that we set up a cutoff that the sponsors would be held to.

DR. TRUSSELL: I don't think that anybody thinks there is a residual effect that lasts 7 days into the next cycle. In fact, if you miss two pills or three pills, depending upon WHO guidelines, depending upon what dose of pills you are looking at, then, you are supposed to use back-up contraception.

So, I think the question really is uncertainty in dating of when the pregnancy occurred. I mean, in principle, if you knew exactly, then I would think that you would not want to count any pregnancy that occurred after the last treatment cycle, the pill-free interval being included in that treatment cycle, the full 28 or whatever days.

DR. LOCKWOOD: So, 7 days beyond the end of hormonal treatment.

DR. TRUSSELL: Well, but it's because it's

a part of that same treatment cycle. I mean, you would count--

DR. LOCKWOOD: To zero.

DR. TRUSSELL: Zero. Zero beyond the end of--I mean, you still--suppose it's 21 active days and 7 placebos, you count the 7 placebos, as well. But in the real world, there is a problem dating the pregnancies with confidence to know whether they occurred on treatment or not.

This is erring toward counting them--the 14-day rule errs towards counting them as being on treatment.

DR. LOCKWOOD: I mean, I take your point. I think you are actually being more generous than current standards apply, as I understand them, because you are saying 14 days, and this argument is once you are beyond the treatment period, you are no longer--we don't expect efficacy from the agent.

DR. JOHNSON: I must confess, as a reproductive endocrinologist, when I read 14 days, I thought, but people can ovulate and conceive

before 14 days. So I agree. At the end of the treatment, whatever that treatment is, that is when the endpoint should be from my viewpoint.

DR. LOCKWOOD: Dr. Stadel.

DR. STADEL: I just endorse that. I think if a company has proposed a product with a defined pill-free interval, and if that protocol is acceptable up front, then one certainly has to treat pregnancies differently if they fall outside that predefined interval than if they fall inside it, in my mind.

DR. LOCKWOOD: Dr. Gillen.

DR. GILLEN: I think one thing that comes up here--I agree with the counting inside, but one other notion that comes up is that people need to be followed up after end-of-study or after they have discontinued use to make sure that a pregnancy hasn't occurred so you can go back and actually backdate it. Now, if you count it when it was on treatment or off treatment, fine, but you just need to know whether it occurred within some interval of time post-discontinuation.

DR. MONROE: Just one clarification. If it's a 21/7, and normally, you would have that 7-day placebo, you would still consider that as an on-treatment pregnancy the last time through, because a pack has 28 pills.

DR. LOCKWOOD: Those 7 days.

DR. MONROE: Then anything that falls outside of that 7th day, you would not call as an on-treatment pregnancy. Is that what I am hearing several people saying?

DR. LOCKWOOD: I mean, theoretically, a proliferative phase could last 7 days. I mean, if someone normally has a 21-day cycle, they ovulate at day 7, 14 days constant luteal phase, so it theoretically could happen.

So, I would agree, it doesn't make any sense, in fact, to demand that the product protect someone when there is theoretically no reason it should. So it seems to me that it is a plausible argument.

DR. STADEL: This would not [inaudible] denials in pills with shorter than a 7-day,

pill-free interval. Again, if that is what the company has put forward, they ought to be counted that way.

DR. BLUMENTHAL: But in those cycles, it is still a 28-day total cycle, so that, if it is 24 and 4, then the assumption is that at the end of the four days you are done.

DR. STADEL: That is all I am saying, I agree with that.

DR. LOCKWOOD: So, there is consensus, I think.

The next question, Question 12. How can the life-table analysis of pregnancy rates be adjusted for the use of back-up--this is going to be the controversial one--back-up contraception midway through the exposure period--for example, back-up contraception used only during treatment cycle 6 in a 13-month treatment cycle?

So, do we delete that cycle? Do we delete everything beyond there? Do we include it? This is exactly what you were talking about earlier, you know, what can we recommend?

DR. TRUSSELL: To my mind, you really have three alternatives. One is to censor the woman at whatever--treatment cycle 5, so she would contribute 5 cycles. Now, you are going to be throwing away a lot of data by doing that.

The second one is to skip cycle 6, so she would contribute in the life tables to months 1, 2, 3, 4, 5, 7, 8, 9, 10, 11, 12, 13.

The third assumption that you could make is that she contributes to cycles 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12.

I mean, logically, I think that is the only three choices. You can do it all three ways and see if it makes a difference.

DR. GIBBS: Charlie, is the corollary question here what do you do if a woman uses a barrier contraceptive to protect herself against STDs also?

DR. LOCKWOOD: Yes.

DR. TRUSSELL: It's the same question.

DR. MONROE: We heard Dr. Trussell give us three options. Do you have any recommendations as

far as the options, because we can analyze them all three ways, and as you say, each has some merit and some down side.

DR. BLUMENTHAL: Yes; Is there a hierarchy that ought to be applied to those three options?

DR. ESPEY: It seems as--if what we are trying to get at is real-life effectiveness, then leaving them all in would be the way to go.

DR. TRUSSELL: There is a fourth option, which is to recognize that that is the way it is going to be used and and not leave out the cycle at all, but keep it in there.

DR. LOCKWOOD: That's three, that's your third.

DR. TRUSSELL: No, no, no.

DR. LOCKWOOD: Including all cycles?

DR. TRUSSELL: No; There are four. There are four options. One is to throw her out.

DR. LOCKWOOD: Eliminate the cycle.

DR. TRUSSELL: The second one is going to be to count all 13 cycles and understand that that is the way that people use pills that were

occasionally using back-up.

DR. LOCKWOOD: Right.

DR. TRUSSELL: The third one is to omit cycle 6 in the life-table analysis, and the fourth one is to count the first 12 cycles. She only contributes 12 cycles, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, not leave out 6. You go 1 to 5, and 7 to 13, or 1 to 12.

DR. ESPEY: I think what I was supporting was Option 4, to just leave them in. Again, I think the whole move here is looking towards real-life effectiveness as opposed to efficacy, and that would be the most likely way to achieve that.

DR. LOCKWOOD: And the assumption of the control trial.

DR. ESPEY: Right.

DR. PETITTI: You are going to open up a bit of a can of worms if you start to look at condom use because condom use is not reliable. And so now what do you do if she used condoms half the time during month 6? It might be easier just to leave it out.

DR. GILLEN: Again, I think a lot of this goes back to are you doing a randomized comparative trial, or are you doing a historical controlled trial.

If you are doing a randomized, comparative trial, there should be no reason to believe--well, hopefully, there would be no reason to believe there would be differential use in terms of back-up contraception between the two arms given the randomization.

If they are doing the historical control, then you have got to define exactly what you are using as a threshold and be certain that your historical control is representative of the back-up methods that they were using regardless of how you are putting them back into the risk set now to make a fair comparison between the two studies.

DR. LOCKWOOD: Theoretically, you would have to do whatever they did, whatever the historical control did.

Dr. Stadel.

DR. STADEL: I think from the standpoint

of having been a reviewer, I would probably say let's look at it with the back-up people in and with them out and does it make a big difference and, if it doesn't, then, whew, the problem is over, and if it does make a difference, then you have to really dig in and engage in the review and you have got to come up with an interpretation and a position. In that situation, it is a matter of judgment, and I don't think there is a formula answer in those circumstances.

DR. PETITTI: Perhaps the one wrong answer would be to censor them and throw them out of the study completely. Did I hear that from--

DR. LOCKWOOD: I think there is consensus, then, that people or sponsors ought to analyze the data both ways but that, if there is a hierarchy, if I sense a hierarchy, it is to model the real world which would mean to include all the data and assume that it reflects typical use, one, and two, that it is going to be reflective in both the treatment and in the control arms to an equivalent degree assuming it's adequately powered, and so

forth.

Next question. How should the analysis of pregnancy rates be adjusted for the use of back-up contraception in extended cycle contraceptive trials? For example, in an 84/7 dosing regimen, should an entire 91-day cycle be considered non-evaluable, or should a 28-day portion of the cycle be excluded from consideration of at- risk cycles?

DR. STADEL: I will put forth the same thing I would do is I would look at the data with it all out, and then with it back in, in varying definitions, and determine how large an impact it had.

The one opinion I would express is that I have a little discomfort with relying primarily on the total data with back-up methods in because of what it might encourage in terms of behavior during trials and the running of trials.

I think that is a reality you have to think about is that. if the primary emphasis is on data that includes the use of back-up methods, what

does that do with regard to what kind of device people are given, what the whole context is. I think there is a very delicate issue here having to do with trials and success.

DR. LOCKWOOD: It certainly could introduce a bias if a sponsor were encouraging safe sex and avoiding sexually transmitted diseases.

Paula.

DR. HILLARD: So, the other issue here is in reporting of condom use. So, if we have concerns about diary reporting versus electronic reporting of pill, consistency of pill use, we also have the issue of reporting of condom use, as well, so what sort of a marker do we have for that consistency?

We don't have an electronic marker for that, as well, but we throw it all of that into the mix and I think we have to consider that as we think about both of these questions here.

DR. LOCKWOOD: Dr. Peterson.

DR. PETERSON: I think part of it is going back to this issue of apples and apples, and

oranges and oranges. With the active controlled trial, you do assume that that major factor is going to be equally distributed between the groups and, in a sense, it takes away the problem. But it creates a generalizability problem when you try to take those findings and then say what--in effect, did these people use two methods of contraception and to what extent did that happen when you try to interpret the findings of that study for the real-world effectiveness part.

DR. LOCKWOOD: So I have actually been of more of a statistical nature. It was pretty clear your three options with traditional 28-day cycle treatments, but you really don't have that option of excluding the cycle with extended dose regimens because you only have one cycle basically.

So, in this context, I guess the options would be only two. You would censor the remaining days of that extended dose-regimen cycle, or you would eliminate the patient from analysis.

DR. TRUSSELL: No, I would say that there are still four options. One is to censor her at

the end of her last cycle before she started to use back-up. The first time she uses back-up, she is out of the trial.

The other is always to count it, because it's a pure intent-to-treat analysis. And then the other two would both involve throwing out, in this case, a cycle, which would be 90-whatever days.

DR. LOCKWOOD: Dr. Tobert.

DR. TOBERT: I mean, it seems to me in terms of the previous point about interpretability of the data, as long as it says in the label how many of the cycles had back-up contraception, or how many women use back-up contraception, and is a randomized trial, then the prescribers have got all the information they need, I think, however you analyze it. But I certainly don't favor throwing out, wasting data.

DR. GILLEN: So I just want to make a quick comment on putting them back in the risk set on one of the options that was stated. So, let me just take an example where a woman is on treatment from 0 to 3 months. She is off treatment or what

have you, on a back-up contraceptive for 3 to 9 months, and then she comes back on again.

One of the suggestions was to just go ahead and tie her back in and assume that, if you were going to 12 months total, that she was on treatment from zero to 6 months.

When you are calculating that Kaplan-Meier estimator, you are assuming--you are putting her back in the risk set and conditioned on the fact she has been--you are saying what is the probability she would fail in the 4th month given she was at risk for 3 months.

She hasn't been at risk for 3 months. She has been at risk for 9 months at that point. Hazards change over time and if somebody survived for 9 months without being pregnant, that is a lot different than them surviving for 3 months without having become pregnant. Her baseline risk is different.

So, I would actually strongly urge against throwing her back in in a continuous fashion because then I do think you are mixing risk sets,

and that can be a dangerous practice.

DR. TRUSSELL: What if, in fact, there was no sex in that month? That does happen. Should you be including months where there is no risk at being exposed?

DR. GILLEN: I still don't see the issue, though, with putting her back in the risk set as soon as she comes back into the trial, based upon study time. Right. So, you can still include her back in at 9 months, 10 months, and 11 months. Let's assume there was no sex for the inner 6 months, but do you really want to put her back in at the 4-month interval?

DR. TRUSSELL: As I said earlier, my first choice would be to put her back in where she comes in. But there is an alternative. One can make an argument that you could put her back in and make continuous use. You can run it both ways and see if it makes any difference.

DR. GILLEN: I guess from the standpoint of, a priori, stating what you are going to do, I am just illustrating the argument against putting

them back in, in a continuous fashion.

DR. TRUSSELL: I would, a priori, state I am going to do it all those ways.

DR. TULMAN: I have a question for the FDA. When these trials are being conducted and being set up in a methodology for subject recruitment and the procedure, what are patients or subjects being told about safe-sex practices, or are they being told anything vis-a-vis using condoms?

DR. MONROE: I think today that patients are recruited with the expectation that they will not use a back-up method. But, on the other hand, they are not precluded because of these various concerns about STDs, and so forth. So, it is something you can't absolutely control. I can't say what investigators are actually telling the subjects, but I am sure that no investigator would counsel that somebody should absolutely not use back-up contraception.

So, we are faced with a practical problem is really why we are asking this question, because

the intent is not that a sponsor enroll a large number of women that clearly they are going to be using back-up contraception primarily for protection against sexually transmitted diseases. But, in the real world, that happens.

Then one has to address that in the context of the data that you receive from a clinical trial, and that is the reason we are really posing this question to the committee is to see what your recommendations are in terms of handling the reality instead of an ideal situation because our trials are conducted in the real world, at least the data we get from the trials that others conduct.

I can't answer your question really. You would have to ask a sponsor of a clinical trial exactly how they are counseling patients. But, in protocols, they are not precluded, clearly. I don't see that as an exclusion. Is Dr. Price here? You have heard many of these. I think that they are given that as an option to use.

What has been your experience, Phill, in

the protocols you have recently reviewed vis-a-vis that question?

DR. PRICE: Just exactly what you said.

DR. TULMAN: I guess my point is whether--is that, at some point, who has ultimate responsibility when you give the woman the set of pills, the pack of pills or the packs of pills, in terms of, from an ethical point of view, what do you tell the person in terms of safe sex practices and is there any uniformity--should there be any uniformity, should the FDA have any role in that. Does it go back to the institutional IRB, or how does that work?

DR. MONROE: I would like to really throw that question back to members of the panel because we have many investigators, I think, sitting at the table, and why don't we have them answer your question because they are the ones who are actually involved with the clinical trials. There may be differences amongst how they advise patients, so I am not the right person to ask.

DR. PETITTI: I don't want to go back--I

want to go back briefly to the life-table issue, but I don't think this should really sidetrack the conversation about what to do about counseling for STDs.

I think that I finally found something to disagree with James about, which is I think that, given that we have determined that a Pearl Index is not the best way to analyze data and that the life-table methods are always preferable and, given that we agree in general that fertility has a changing hazard, that you would actually want to put the women back into the life-table analysis in the month in which they would have been had they not taken the break.

It comes up again and the reason why I even belabor it is we are going to sort of talk about all this extended dose-regimen stuff again tomorrow, and I do think for the extended dose regimens, that they would come back in at that level.

Now, that is not to say you shouldn't do the analysis every way, but I have a little bit of

discomfort when you say do it this way, do it that way, do it however--you know, five different ways and see if they agree, if that is the best advice. I would prefer in this instance to give specific advice that given the hazard issue, that they come back in at the month they would have been at.

They have a personal biological clock ticking at number of cycles, and you never get off that clock.

DR. TRUSSELL: That would be my first choice, which is what I said, but what would you do with people who one cycle did not have sex, what is happening to their personal biologic during that cycle? It's the same issue.

DR. PERLMUTTER: I would like to be a little more practical on this. If we have somebody who is taking pills on a 84/7 dosing, and we know from Potter's study that, in fact, women miss pills and they do so on a regular basis, then, to get into an ideal situation of somebody taking their pills every single day, I think, is ridiculous because they are not going to do that.

If we are going to get into reality, then we have to think about missing pills and we have to think about using some kind of back-up contraception, and I will even take it away from sex.

What if somebody is sick and uses antibiotics, and they start spotting, and the recommendation is that they are the ones most at risk of pregnancy, and they use condoms for that week, how do we handle that? That is reality.

DR. LOCKWOOD: I do think, though, that in order to calculate perfect use and typical use, you really do need to analyze it both ways. But I think it is important to analyze it both ways because you do want some model, some surrogate, some reasonable assessment of what typical use will be.

I don't see any way around it. We do want perfect-use data, I think--and I think you have to analyze the data in such a way as to exclude this confounding from multiple contraceptive use.

DR. PERLMUTTER: Can I just respond to

that? I agree with that totally, but I think your numbers are going to have to be exceedingly high, then, in order to get the numbers that you need for perfect use. You are going to have to have huge numbers.

DR. LOCKWOOD: Depending on the frequency of use of other methods and missed pills, and so forth.

DR. GIBBS: Charlie, two separate comments.

The first question is how reliable is reporting of back-up contraception. My guess is that from what we have heard about reporting of abortions and lots of other things, it may not be all that reliable and, if it is not that reliable, then we are going to have a lot of miscategorization of patients, those who did and those who did not use back-up contraception.

So, I think that is kind of muddying it, and I think the best thing to do is just take all the data as it is and then do subcategorization as you like.

The second commentary is what to do about condom use. So, I think it is incumbent upon investigators to encourage all potential research subjects to practice safe sex. That is just ethical. And, if you are not going to enroll condom users, then you are going to be limiting your research enrollees to monogamous women and that creates problems of generalizability.

DR. LOCKWOOD: Dr. Tobert.

DR. TOBERT: As this very enlightening discussion is proceeding, I am wondering is this situation really different, the situation of back-up contraception, to what pertains with a lot of trials in different areas of medicine.

The one I am most familiar with is cardiovascular outcome trials where you test the treatment, of lipid-lowering treatment, say. You don't care if somebody starts aspirin, if they have got a beta blocker, if they start some other treatment. I mean, you don't censor them when that happens even though their hazard is different.

That is the real world and you basically

ignore that. Nor do you discourage that either. Now, you might say the effects are not as great with a condom, but still in principle is the same thing. So, I am for ignoring condom use basically.

DR. LOCKWOOD: That sort of gets at the fundamental philosophical question of should we actually be trying to calculate perfect use. We don't calculate perfect use as regards to MIs, you know, if we are using a statin or a Plaxil or some other trade name I am not supposed to use.

So, why do we do it for--I mean, I think I know the answer--but why do we do it for contraceptives?

Dr. Gillen.

DR. GILLEN: I had two comments actually. I guess the first would be I agree with you. Correct me if I am wrong, but it sounds like your experience is coming more from comparative trials, though, right?

DR. TOBERT: Right.

DR. GILLEN: Exactly. If we are in a comparative-trial situation, that is not going to

be an issue because you should be nondifferential on each arm, But, if you are going to the historical control, again, you have to decide what you are going to do with these people because you are measuring two different quantities, whether you leave them in or whether you take them out.

DR. TOBERT: But the historical method is history as far as this panel is concerned, isn't it?

DR. GILLEN: And if everybody is content with the threshold being set upon having back-up contraception in there, then that is fine. Then you go into it with your eyes open and you say this is a threshold I am setting, given that I am going to analyze the data in this way, and this is the parameter that I will be estimating. So, I would agree with that.

My other question actually--so there has been talk of I would analyze it both ways, and I would, a priori, state that I would analyze it both ways. I was just wondering if the FDA could comment for a second on choosing the summary

measure and how it would be defined up front.

I mean, for example, I have been in a trial where I said, okay, I am going to compare a hazard ratio, and I will show you what the median survival is. I am going to choose both of those to be my primary endpoints. Really, that is what we are doing when we are defining the summary measures and we are saying we are calculating them in different ways. We are potentially estimating different parameters and we are testing different parameters here.

It seems to me that we are specifying multiple endpoints at this point, and I am wondering what the FDA's thoughts are on this. I mean, it seems like we should be trying to come up with a consensus in terms of saying how are you going to analyze your data at the end of the day.

Now, other secondary endpoints definitely need to be looked at, and subgroups need to be looked at, et cetera. But that is not where our primary analysis stands from my experience. So, I was wondering what the thoughts were on that.

DR. LOCKWOOD: Don't all answer at once.

DR. KAMMERMAN: Hi. This is Lisa Kammerman. I am a statistician with the Center.

I agree we need to prespecify the endpoints upfront. We always look at other analyses as a form of sensitivity analyses to see if there are discrepancies, but I think it would be very helpful to get some consensus on what the endpoints should be.

Do we want to use the Kaplan-Meier estimates, say the proportion of women who became pregnant within the first six months, within the first year? Do we want to look at the shapes of the curves in getting there--for example, the log-rank where we are comparing the shapes of the curves-- regardless of the one-year endpoint? I think that is what Dr. Gillen is getting at, but that would be a very helpful contribution.

DR. GILLEN: If I can just respond real quick. Yeah, I mean it's that and it is even slightly more subtle than that to say, yes, I want to look at the Kaplan-Meier probabilities at six

months or one year. And then it comes to a question of how do I calculate those Kaplan-Meier probabilities. Should I put people back into the risk set when they left? Do I put them back in when they returned?

So, what happens in a hypothetical situation where you have some sort of conflicting or inconsistency across those two methods? It seems to me that what the panel would like to do is to come up with a consensus first and say this is what we are going to be looking at as a primary endpoint, this is how we are going to be calculating it, this is exactly what our outcome is.

Then other things become secondary endpoints in support of evidence and sensitivity analysis at that point.

DR. KAMMERMAN: I think my personal opinion is that we want to look at the intent-to-treat. The women, assuming we have a controlled trial, are randomized to one of two treatment arms with the intention that is the protocol they are going to be

following for the year. And there are always mitigating circumstances no matter what study, what drug product.

So, because of the issues of measurement error and back-up contraception, as we saw in one of the earlier charts, isn't always so great, and right now, if there is a pregnancy that occurs during back-up contraception, we are counting that. But otherwise we exclude those cycles.

So, it is my opinion, if we do have a controlled trial, we need to include all the cycles. However, if you think it is better to exclude women who aren't a risk, understanding there is going to be a lot of measurement error and misclassification rates, then I agree that the women should reenter relative to the start of the trial.

So, if she misses the first middle two cycles and was compliant the first three, then she would reenter at cycle 6. Is that what you were getting at?

DR. TRUSSELL: I would strongly support

the primary endpoint be all cycles and all pregnancies, and all the rest of these were secondary. But I didn't realize you were asking what should be the primary outcome measure because the primary outcome measure, at least in all the trials I have seen, has been the intent-to-treat populations, so I have no reason to change that.

But these other secondary analyses tell you whether it makes a difference how you handle cycles of no use or cycles of dual use.

DR. KAMMERMAN: Just since I am here, also, I just wanted to address that when you talk about the active controlled studies, it is also important to keep the hypothesis in mind. Is the idea to show superiority and efficacy? Is it to show a dose response? Is it important to show that there is comparability and efficacy in order to show an improvement in safety?

So, when we throw out the term active control, it is always important to keep in mind the general hypothesis and what the study objectives are.

DR. LOCKWOOD: So, to summarize, we are being asked--really, the statisticians are being asked--to advise the rest of the panel. The actual study design characteristics that are required, intent-to-treat seems to be universally agreed upon. I think everybody on the panel would agree that that is the ideal way to approach it.

Life-table analysis, no further discussion needs to be done on that, but the specifics of that life-table analysis, the specific type of life-table analysis and Kaplan-Meier, and then how to handle subgroup analysis. I guess, beyond just the issue of back-up contraception, theoretically, you could also parcel out high BMIs and other aspects to that.

And then final question, non-inferiority versus superiority, or is that up to the sponsor?

DR. BLUMENTHAL: I was going to come back to just the concept of all pregnancies, all cycles, I think that if we look back at just the morning discussion, well, we have sort of gotten rid of the Pearl Index and one or two other things and it

seems to me that we are rapidly approaching the point where we are getting rid of perfect use. There is no perfect use.

In this day and age, with back-up contraception, whether it's emergency contraception and hormonal, or whether it's condoms, or whether it is use of condoms to prevent STDs, it is unlikely that there are going to be any real perfect-use cycles anymore.

So, the concept of all pregnancies, all cycles, our real intent-to-treat analysis is likely to be the most clinically useful, as well as useful to both the industry and the Agency and that might be a sea change just in general and make the chart in contraceptive technology a lot simpler, too.

DR. TRUSSELL: That is already what they are doing.

DR. GILLEN: Well, I think that is the criti--I think that is key.

DR. TRUSSELL: A perfect-use analysis is a secondary analysis before the FDA.

DR. GILLEN: I think it is unlikely, and

again I think that is why we need the active controls trials because we can't use any of these historical controls in this context, in the context of all these back-up methods being used, and so forth, and so on.

So, in a sense we have to rebuild the database anew and reeducate both our colleagues, the public and industry, to recognize through the non-existence of perfect use, the real intent-to-treat analysis and life-table utilization, and I think people will pick up on it pretty quickly.

DR. LOCKWOOD: Dr. Stadel.

DR. STADEL: If primary emphasis is placed on the efficacy data that includes the use of back-up contraception, it seems to me that that pleads again for a comparative trial where the active comparator is known, its efficacy is known, without back-up; that is, that you have some idea of what the actual efficacy of the product is.

If you don't have that, if your comparison involves a lot of back-up, it seems to me I would

want to augment that trial data with some surrogate-outcome information, and I think probably some work is needed on the extent of consensus about the use of things like ovulation suppression because if we are in a era--and I was just struck by the comment of passage of time from when I entered this field many years ago about the difference in the issues involving sexually transmitted diseases for the recruitment of patients into these trials.

So, my last comment is on that. It seems to me there is a shared responsibility between investigators, sponsors and the FDA, and that the investigator is encouraging the patient into the trial, the sponsor is supporting the trial, which gives them a major responsibility for what they are supporting, and they are supporting it in response to Agency needs for information.

So, it seems to me there is a shared responsibility for seeing that the advice that is given to patients when they are recruited in these trials is ethically acceptable.

DR. HILLARD: Just following up on the previous two comments, I think recognizing the current realities in terms of risks of STDs and even the realities of clinical trials, we are potentially, in clinical trials, recruiting women who believe themselves to be in a mutually monogamous relationship, and that may or may not be the case.

The woman may or may not be aware of partners' other partners. So, I think we have to recognize that as well. I think we also have to recognize the realities of patterns of sexual activity particularly for adolescents, but also for young adults, and those are patterns of serial monogamy.

So, over the course of time, depending on what your time interval is, whether it is six months or a year, that individual may well be in a different relationship at the end of the year. My patients, the women I see, recognize that,if they are in a new relationship, then the current advice is use condoms, not for back-up contraception, but

for minimizing risk of STDs.

So, I think we have to recognize those real-world realities for women that are using our contraceptive methods, but also women who are being recruited into clinical trials.

DR. TOBERT: Mr. Chairman, I think you said that these active comparison trials would be analyzed on the intention-to-treat basis, but I don't think there actually can be a pure intention-to-treat--Dr. Gillen might want to comment--because after all, if a woman stops using the treatment--say, she wants to get pregnant, or for whatever reason, she stops, she has an adverse effect, then you don't include pregnancies past that point. So, that violates the intention-to-treat principle. So, I don't think there can ever be pure intention to treat.

I see Dr. Gillen shaking his head, which gives me further confidence to go on, which is to try and answer the question that we had from the FDA about what the hypothesis should be for these comparative trials, and I think it should be

non-inferiority in most cases because, after all, the standard is pretty damn good. You are not likely to be able to beat it in terms of efficacy, possibly in terms of adverse effects. I think we will be talking about that tomorrow. But I think the non-inferiority margin should be quite wide. Otherwise, the trials would be impossibly large.

DR. LOCKWOOD: Dr. Scott.

DR. SCOTT: I understand the importance of effectiveness rather than efficacy, and that is important.

I just wonder, though, whether there are studies to show what is the best way to get efficacy. In other words, I know that there are studies to show you can't even take penicillin 10 days in a row, that people stop it, and so on. But somebody brought up the question about shorter trials once.

Are the data more reliable in a shorter trial--in other words, as far as the pregnancy rate is concerned? A lot of these things I think are solved with the active controls. But nevertheless,

if you are going to compare a new preparation, say, with a different preparation, and a woman who is monogamous is going to depend on the efficacy, and you have different women in the active controls, would the efficacy actually be lower for a monogamous patient who is not using the condoms and other methods, too?

In other words, you see what I am saying? Maybe there is a way, there are some guidelines to say what are the best ways to get efficacy rather than effectiveness also.

DR. LOCKWOOD: Dr. Gilliam.

DR. GILLIAM: I think there are three distinctions or three levels of these trials. If you assume that all women are biologically the same, then there should be an inherent efficacy of a drug, but that is different than what we are measuring when we measure perfect use.

We are not measuring the inherent efficacy that is somewhere out there. We are measuring what happens to some extent when you put it into real-world use. We are not able to account for

whether women are actually having intercourse. We are not able to account for how fertile that woman is. So, even by the time you actually put it in a real-world context or even an ideal-study context, you are already moving a step away from the inherent efficacy of the drug.

The next level is what happens with typical use when we put in all of the messiness of human behavior. I think, with those distinctions, we probably can't get the ideal efficacy but we can--at least with some secondary analysis and taking out condoms and deciding how we are going to use that, we can calculate what the perfect use might be in comparison to the typical use.

DR. SCOTT: Is there any information on a shorter trial, if that is more--the data are better than if it is a longer trial, or the way it is conducted?

DR. GILLIAM: A couple of times the idea of using biologic endpoints has come up, but I would think, for example, if you were going to look at BMI, you might just look at ovulation

suppression.

That would show in this woman of this given BMI, does this dose of contraceptive suppress ovulation. That would tell you the efficacy of that drug, not how she is going to use it, or what will happen with long-term use.

DR. SCOTT: I am a little suspicious of surrogate markers, though, even though it has been mentioned several times.

DR. GILLIAM: I understand the problems with secondary, and it doesn't give you a lot of clinical use. But you might be able to get some sense of safe dosing for a contraceptive method that way.

DR. LOCKWOOD: I think if we are moving toward consensus, it is that the concept of perfect use is probably an anachronism, that there is no sort of perfect person, that there is substantial variability and fecundity related to age, related to several mucus factors, male factors, et cetera, et cetera, and that, even with perfect uses, there is likely to be significant variability in

different populations, that there might still be a valid effort to make in terms of secondary analysis. Primary analysis ought to be on actual use.

I think we have covered the statistical approaches that ought to be used ad nauseam. There is universal acceptance of this modified intent-to-treat although bearing in mind that, unlike an MI, some people actually may want to get pregnant--some people may also want to have an MI, I suppose, but fewer likely--that there are lots of messy conditions in real world with monogamous relationships, and serial monogamy and age factors that relate to monogamy versus use of barriers. But, again, a lot of these wash out with use of active controls, intent-to-treat, and life-table analysis and the ability to then do subanalysis.

I want to move to the next set of questions. We don't want to move yet to the next set of questions?

DR. PETERSON: I don't want to hold things up at all, but I hope that the concept that Bruce

put forward is one that helps us move ahead because so much washes out with the randomized trial design. But at the end of it, what do you know to inform the physician and the patient about how well this works.

Bruce's point, I think really saves the day on that one because, if you know how well something works that you are comparing it to, you have a much better answer to that question.

An absurd example would be if half of both groups used an IUD. Well, that would wash out, but would you know how well the pill works? No. So, it is really imperative that we know what we are talking about when we are comparing, to ultimately translate to how well does this work.

We know it works as well as that does, but we don't know what to tell the patient in absolute terms about how effective it is unless we know that for what we are comparing it to. So, I think Bruce sort of helps us get out of a lot of the dilemmas that we have been discussing.

DR. ESPEY: Well, maybe it would convince

them all to use an IUD, which is a much better method anyway. But, I mean, I do think we have some consensus about this. I know there are some one-person splitters, but the overall concept, I think, just taking as a given that we all approve of the idea of using active controls and that, in that context, you know, throwing anybody in there is still--you know, it is going to give the most powerful data of what actually happens out in the real world where condoms are used but I think, as Paul has mentioned, inconsistently and, you know--

DR. LOCKWOOD: And inconsistent reporting.

DR. ESPEY: And a difficult reporting issue.

DR. LOCKWOOD: Dr. Monroe.

DR. MONROE: Yes. I would like a little clarification on expanding upon your concept that initially you raised and how the active comparator, at the end of the day, for instance, would help us write a meaningful label to convey the actual or typical effectiveness or efficacy, whatever term we want to use, to both the professional healthcare

provider and then the consumer.

The very challenging question, which we didn't really resolve, is what this comparator might be and by whatever standards were used when this comparator was approved to be a safe and effective compound or drug. How they compare to what we are doing today, the population, and so on, it is hard to know for certain.

So, we agree hypothetically that we are going to use Drug X as the comparator because, going back to your example, it is either widely used or it's part of a basket of drugs, or whatever else. But, in today's population, we don't really know its absolute efficacy except for this new trial we are going to do because it may have been approved 10 years ago or 15 years ago, 20 years ago. So we run it in this randomized, active controlled trial and it comes out with whatever the number may be.

Again, it's a little bit problematic because--and we have the new drug, as well, in here--and then we have to decide whether this is

going to be a non-inferiority or superiority. In most cases, I am sure a company would try to go with the non-inferiority because to show superiority over products that are already very good would be extremely difficult.

I am not quite sure how that is going to solve all the problems. We come out and we can just say it was non-inferior to some drug or previously approved drug, and to make that statement. Then we give, what, the results of the trial which was just conducted? We give the results of both drugs? We give the results of just the drug that is up for approval?

I wonder if the panel could help me better understand, as we carry this through the process, how this is then going to lead to something that is going to be easy to interpret and help us to understand really the level of protection that it might give the average user in terms of prevention of pregnancy.

If you could perhaps discuss that a little bit. Then the last question I hear from that end

of the table, I am not sure who is saying it, well, we have to use a wide margin to show non-inferiority.

What do we mean by a wide margin? Are we talking about a couple of percent, are we talking about absolute percent--because when we are talking about effectiveness of oral contraceptives, let alone injectables or implants, which are probably even more effective, we are talking about drugs that are approaching, I would think, 98 percent effective or somewhere in that range, maybe 99, at least in clinical trial environments.

So, where is our room to go, and if we are going to say it only has to be non-inferior by--are we going to be talking about relative percents or are we talking about absolute percents. If we are talking about absolute percents, and it has to be only within 1 or 2 percent, we are talking about a product that might be only half as effective as the standard.

So, if you follow this whole concept through, it is leaving me very confused and very

challenged, I guess here because it may seem, from some of our questions, that we don't have standards in place, but we do.

They are really fairly cautious standards, conservative standards, and I am afraid that from some of the things I hear, it might actually lead to a relaxation.

That's what I--I don't know whether we can address it today, but certainly before this is over, at least I would like to hear this. And I don't know if the rest of my colleagues here at the table would like further clarification because some of these are nice concepts but I would like to learn how they are going to--or how you folks would suggest they actually be applied in the situation of a contraceptive.

It is easy to apply this for a therapy that maybe you only get a 30 percent response rate. But we are talking about response rates that are very high at the extreme in terms of protection and how they would apply in these circumstances.

I will just stop now, but if you could

consider that.

DR. LOCKWOOD: I want to rebut the challenge that we are proposing a relaxation in standards. I don't think that is the case at all. I think it is more reformation of an extremely ornate structure that maybe doesn't really have any real-world basis for its justification.

The argument I think, if I can summarize the group's consideration of this, is that it is a reality that safe and effective contraceptives exist, that they do a very good job, particularly when used appropriately for the purposes of contraception, and that a lot of the previously used measures of efficacy may not be relevant, Pearl Index, or even perfect use, frankly, just because there is no perfect use, and that what has been suggested is that a lot of potential problems, washout with the use of active controls, which it sounds like everybody is doing anyway but we are now saying should only be the case perhaps with the rarest of exceptions, and that rather than having the sponsor identify a comparator that would right

away be to their advantage, so to speak, in terms of having a relatively low efficacy, that it ought to be the job of the FDA to recommend specific comparators, and that the basis for choosing that comparator ought to be widely used current drugs that have relatively recent documentation of efficacy, not something that was approved in 1968 but something relatively recently approved that, obviously, we believe also has a good safety profile and that the options could be to provide to the sponsor several alternatives, the standard comparator, which is a drug--I am making this up now--but a drug that was approved within the last 10 years, that has a very high market penetration, not quite the gold standard, but it's a benchmark, it is something that is widely used, the customers have voted with their feet--they are buying it, whether the customer be the doctor or the patient--and that that would be one option.

Then a second option would be a comparator that differs in only one aspect--let's say a different progestin or a different concentration of

the progestin--again relatively recently approved, you know, a decade, to set a number for you, and that that would be a reasonable alternative.

A third option--this is Diana's rules here--would be a market basket of drugs in a similar class, you know, third-generation progestins, using relatively comparable estradiol doses, and so forth, and that they would have the option of choosing one of those three alternatives and that they would then engage in a randomized clinical trial which would look at real-world entry criteria and judge efficacy based on this already accepted drug, that you could, if you wanted to, try to get some sense of, quote "perfect" use by doing subanalyses of certain subpopulations, but that, in fact, that approach is much simpler, will be easier for the sponsors to grapple with, understand, and conduct trials with, reduce costs potentially, improve access to new contraceptives by women.

I don't think it's a relaxation. I think it is actually a simplification. It's reformation,

not relaxation.

DR. TRUSSELL: It would be a relaxation unless you are going to require very big trials in the following sense. If it is the case now--if it is the case that the FDA has now put a cap of 2 on the Pearl Index, and you are not going to prove anything above 2--it used to be 1 1/2, but let's just say it's 1 1/2--you are not going to prove anything over 1 1/2.

Now, you design the equivalency trial where you think that the gold standard has a pregnancy rate of 1 1/2, and you are going to declare to be equivalent something that is, say, another 1 1/2, so that would be up to 3. Then, you can wind up approving a contraceptive with a failure rate of 3 and saying that it is equivalent to 1 of 1 1/2.

DR. LOCKWOOD: Because, in fact, the, quote "1.5" Pearl Index turned out to be 5 or 4 or 3.

DR. TRUSSELL: No, no, no. I am just saying that--so that is exactly what I understand

their questions to mean and, if you really want to do--in order to get adequate power, you are going to have to have a very, very large trial.

DR. LOCKWOOD: But how can it be defined as relaxation when, in fact--

DR. TRUSSELL: Because you will approve something with a pregnancy rate of 3, whereas, before you would not have approved anything over 1 1/2.

DR. LOCKWOOD: But what it really is is saying that the failure rate was actually higher than has been suggested by the original clinical trial. We are quibbling over semantics over here.

DR. TRUSSELL: Oh, no; it is not semantics at all.

DR. PETITTI: I think that, if you really think that the failure rate, the real failure rate that we are using now as the benchmark, is 1.5 in typical use defined as we have now said it should be defined in a comparative trial, then I don't think you would want to put your money on that trial.

What I think we get here is--I think what we are saying is that we believe that the failure rate of women in a trial who are followed and whose data is analyzed appropriately will not be 2.

DR. TRUSSELL: Then make it 1.

DR. PETITTI: It won't be 1.

DR. LOCKWOOD: No, no, the other way; right.

DR. PETITTI: The other way, make it 4.

DR. LOCKWOOD: 4, 5.

DR. PETITTI: 4. And the standard upon which the FDA is now approving contraceptives based on, let's say, 200 women followed for one 28-day cycle, or 10,000 cycles, what is that? That is one pregnancy; right? Isn't it? Isn't the Pearl Index, if you assume that you have one pregnancy in that trial if the true--

DR. LOCKWOOD: 1,300 cycles--

DR. PETITTI: 10,000 cycles is how many hundred women years? Per hundred women. Per hundred women. How many pregnancies do you have in that trial if the true rate were really--

DR. LOCKWOOD: It's 10,000 divided by 1,300 would be the number of 100 women years.

DR. PETITTI: So, you would have 6. The way I understand--and you can correct me about how you currently analyze the data--is you take that number in the trial that you have done and you correct the number of pregnancies by throwing out all the pregnancies that occurred for some reason you can explain. Now I would--no? Okay.

DR. MONROE: No, we don't throw out any pregnancies, at least--okay. The drugs now--let's backtrack because--

DR. PETITTI: So, it's 6, 6 pregnancies.

DR. MONROE: I am not sure about your calculation, but I will leave that to our statisticians here. But the way our drugs recently have been labeled is we have used the actual observed pregnancy rate. Recently, we have not put in our labels perfect use, number one.

We calculated it's a secondary input and what we are saying--and this is the observed rate--what we do is we count all pregnancies even

if they occur in a back-up cycle. What we are doing is--you may not say it's fair, but it's a conservative analysis--we will remove from the denominator those cycles where back-up contraception has been used or things like that. So, if anything, it is going to make the Pearl look worse, not better.

So, we take those out as at-risk months, unless you get pregnant and then you get the credit for the pregnancy, and I am saying this sort of facetiously here--so, if anything, the labeling is a conservative kind of label of what is used.

So, let's say that--and I wouldn't call this, by the way, typical use as would appear in your chart. That is a very different thing. We have, perhaps, perfect use. We have observed use, and what we get from a clinical trial is the observed, and it may include some components of typical, but we know clearly, a patient participating in a clinical trial is not a typical user in that she is seeing a healthcare provider, she is being supplied with drugs, she doesn't have

to worry whether she is going to be able to pay for her drug this month. So, again, it is somewhat of a contrived environment. So, it is somewhere between a perfect use, whatever that may be, and a typical use, as you include in your studies where you showed rates of 7 or 8 percent failure rate.

So, this is sort of where we are today, so just everybody understands. I hope I have explained the way it is. So, what comes out in a label is a fairly conservative estimate of the true, or at least of the observed, efficacy within the confines of that clinical trial.

DR. LOCKWOOD: Dr. Tobert.

DR. TOBERT: I certainly wouldn't characterize what the panel is proposing as a relaxation at all. I mean, to go from uncontrolled trials as a predominant support for approval to properly randomized controlled trials can't be described as relaxation, I don't think.

I think that the inferiority margin should be applied across all the trials--in other words, the sponsor should be able to do a

mini-meta-analysis to include all the trials--and if the FDA--and also other supporting data that the sponsor will no doubt provide like suppression of ovulation, plasma levels of the estrogen and progestin that will allow the FDA to be more confident that the thing actually works and, if there is any doubt, I mean, the FDA can call for a Phase 4 trial.

The FDA did ask whether the control results should be shown in the label or not. I think the answer to that is a definite yes without showing--I mean, that is the way it is always done in other branches of medicine, drugs for other things. Without the active comparator data, the data would be uninterpretable so those must be shown.

DR. BLUMENTHAL: It seems to me that the Committee has been asked to help the Agency determine ways of doing contraceptive trials that are more statistically valid, more methodologically correct, more clinically useful or maybe more clinically meaningful.

In terms of all the things that we have discussed this morning, and perhaps outlined best in the soliloquy that Charlie provided a few minutes ago, it turns out that effectiveness is lower--just because of the way this is all going to work out is lower than what the standard of the Agency has been in the past. Well, that is just the way it goes and that previous standard was just based on an ideal that doesn't compute in reality.

So, we are going to have a new adjusted standard and, if you calculate, if you look at effectiveness rates across a time line, all of a sudden there is going to be a blip and the rate is going to go up. And that is going to be a factor of a change in how we evaluate the drugs.

Does that mean that that is necessarily a bad thing? Does that mean it's a relaxation? To me, not really. It means that we may have provided data now that are more meaningful and more informative both to providers and to patients than we did before and it has to be interpreted in that light.

DR. LOCKWOOD: Dr. Gillen.

DR. GILLEN: There is some talk of how--and I think this is a very difficult issue is how to come up with non-inferiority margin. One thing I would propose is to kind of take it from a top-down procedure.

So, let's assume, first of all, that we have our active control and we have decided upon that for a second. The easiest possible scenario to come up with a non-inferiority margin is to assume that you are certain about what the summary measure, outcome measure, is for that active control.

So, let's assume it's a Pearl Index and it's a Pearl Index of 1.5, and there is no variability. Now, let's talk about what we are willing to accept as a difference, and we have got to talk about the contrast, so what are we willing to live with in terms of non-inferiority relative to that 1.5.

If we can't come up with that number, there is no hope for us from this point on because

there is variability in that starting number. So, that is the very best-case scenario.

What you have got to ask yourself is are you really trying to prove superiority versus a placebo, in which case that non-inferiority margin is huge, or are you really talking about non-inferiority relative to an active control that is out there with a similar safety profile.

So, let's assume that we are able to decide upon, okay, I am willing to go to 2.5. Okay, that gives me a difference of 1.

Now, I have to start thinking about what is the inherent randomness in my estimate of that active control because now, as it goes and shifts from 1.5 which I thought it was, now in my trial it is actually 2.5, and so I am willing to accept a Pearl Index of 3.5 at this point. Really not the one.

I want to take the worst-case scenarios in there, and what you can do--effectively, what you will need to do is go from meta-analyses to see what the variability is from study to study for

that active control. As you get the worst-case scenarios based upon those 95 percent confidence limits, you have to be willing to live and die by what you are ruling out now that you have set that non-inferiority margin.

Again, if you want to be on the conservative side, taking the lowest worst-case scenario from those historical controls for the active, based upon its 95 percent confidence intervals, would be potentially one of the most conservative things you can do if you are wanting to eliminate the possibility of obtaining non-inferiority results where you have an observed Pearl Index of, say, 4, because it can happen because of the random variability in what the active control measure is.

So, what I would suggest is, if you are going to go down the active-control path, is to go from, again, the step-down method where you assume what you are willing to accept,giving no randomness, build in the inherent randomness as you go along, and then work from there and talk about

what your worst case scenarios might be.

DR. LOCKWOOD: Maybe this is a good point to ask 14. We are dancing around this question.

For historically controlled trials--which we are not going to do anymore--should the consideration for approval be based on the point estimate of the pregnancy rate, the upper bound of the confidence interval around that point estimate.

So, let's modify that question by saying that, in this context, what should that upper bound be? Should it be 95, especially for non-inferiority? Should it be the 95th percentile? Or should it be the 60th or one standard deviation? Maybe we are not comfortable with such a wide confidence interval.

DR. TRUSSELL: I certainly would favor the upper bound of the confidence interval, whatever it is. But I want to follow up on the previous point because I think it is something that it just hasn't dawned on anybody yet.

Let's suppose, using exactly the example we just have of where we think the truth for the

gold standard is 1 1/2, and you are willing to accept a band of 1 point, so that is 2 1/2.

If you run through the math, you are going to need thousands of patients in each arm of that trial and that is a lot more than is currently being called for. So, that is the implication of what you have just said. I mean, you are talking really thousands.

DR. LOCKWOOD: Dr. Berenson.

DR. BERENSON: This is a remark as a clinician. I am concerned about the idea that this group has now decided there is no such thing as perfect use, because perfect use just means 100 percent compliance with your medication;the patient took their pill every day within a 2- or 3-hour interval at the same time, or at least a pill a day.

I do have patients that do that. So how do I counsel my patient? If they now label it based on typical use, that the pill is only 93 percent effective, and I have a patient that would use it every single day, it is not correct for me

to counsel her that she needs to get an implant or an IUD because that is more effective. That is why we have perfect-use and typical-use tables so we can counsel our patients appropriately.

DR. PRICE: On that same subject, it has been documented in diary data and we just say whether--how well you believe diary data. But manufacturers have submitted this data where, quote, unquote, this subject has used her medication perfectly.

We have pregnancies that we are still looking at where the supposed patient missed her cycle, her pill, by six hours or one hour, or one day, and she was counted as a user failure. So there are subjects out there who take their pills perfectly.

DR. PETERSON: Just following up to the last couple of points, what is helpful to the FDA in answering Dr. Monroe's question, and what does the FDA need to know about effectiveness, what does a provider need to know, and what does the patient or client need to know and how much of that should

happen with pre-market approval process, and how much should follow in the real-world effect in this part, in the post-marketing surveillance part, because we already started with the understanding that for reasons James just mentioned with sample size, that things like is this pill going to be substantially less effective, let's say a 20-microgram in an obese woman. Well, you are just not going to know until presumably the large studies are done post-marketing.

But what is important to know in the pre-market approval process--and if we go back to Dr. Monroe's point, about 98 or 99 percent, then that is really relating to Abbey's point about what happens with taking a pill every day and do we really need to know for approval or for provision what one could expect if they took the drug as indicated. Then we have also, I think, all agreed that we ought to have some understanding of how effective the drug is as commonly used.

So, the question would be how effective is effective and what degree of discrimination needs

to be made when we start trying to answer 15 and 16 if we need to determine that we are going from 98 to 99 percent.

Let's say that if somebody took a 50-microgram pill every day, the risk of pregnancy was 1 in 1,000, and a 30- to 35-microgram, it's 1 in 500, and then in a 20-microgram, it's 1 in 100, well, is that difference important and, if so, and it's important pre-market, then the studies have to be designed accordingly, and sample sizes are huge.

On the other hand, if that is not an important difference to be determined pre-market but it is one that is important to be determined post-market, then our colleagues at NIH and others--you know, that is a research agenda.

But I think that would be helpful to us in trying to be helpful in answering your question about what it is that is important pre-market in distinguishing between the level of effectiveness as indicated and as typically expected that is important to discriminate between prior to approval.

DR. MONROE: I think what we want is to get the thoughts of the panel really as to what some of these parameters might be. We have had many discussions amongst ourselves. We have a large department and we have a range of opinions. What we would like are your thoughts because this is giving us a group of experts, people that are involved with patient care, and we would like to hear from you.

That is why we actually have that question out there. So, again, the considerations are, and I am just going to think in terms of the Pearl, because that is the way everything is labeled today, admitting that it has many pitfalls and maybe it's a way of the past.

But the numbers we have quoted, both in this document and elsewhere, are the point estimates, number one. That is why we have asked about again should we be talking about point estimates or upper bounds of some confidence interval, because the point estimate is only an estimate and there are certain ranges of certainty.

Perhaps Drs. Trussell or Gillen would like to address that, are there bounds that you feel beyond which--I think one of our questions addressed that, Question 15, that you just, as a clinician, doesn't feel, or would not feel, it is appropriate to have a hormonal contraceptive that didn't meet certain levels for efficacy at least again as best we can measure in the clinical trial, whether we want to talk about perfect use or not.

Again, the more parameters we put on it, I think the less we know with certainty. I think we can come out of a clinical trial and know with a higher degree of certainty how many pregnancies occurred. Whether the circumstances of those pregnancies are always associated with perfect use or not, we have to go back to diary data.

I think we are going to hear some presentations on how accurate diary data are--they are perhaps less than 100 percent accurate--or the use of pills, so that maybe you will feel more comfortable because we have given you a number, but how valid that number is, I am not sure.

So, we don't want to give numbers that aren't going to be useful. So, again, I have to put the question back to you folks because, at the end of the line, we approve drugs really on a risk/benefit ratio. There is a price for use of these drugs whether it be in the very rare but serious adverse events or other kinds of things.

So, we have to balance all that, because I think everyone will recognize that you could have a pill, an oral contraceptive, with a sufficiently high dose of estrogen and progestin that you could approach levels of effectiveness that might get close to an implant or something of that sort. But would we find the safety profile that goes with that acceptable, and maybe your answer is yes. I don't know.

So these are--it is not an easy question. If it was easy, we wouldn't be here asking you to help us come up with an answer.

Please, Shelley.

DR. SLAUGHTER: I also think, Scott--I think that we try to provide in our label the best

information that helps you as the prescriber counsel your patients, So we have to turn it around, what sort of information do you need to counsel the patient and is it important to say that, in a clinical trial setting in which everything is controlled, this is probably the best you are going to get, and then we go down from there.

Again, I would like to hear what information you think in terms of whether it is the Pearl or the life-table analyses that should really be presented, so you can best counsel your patients.

DR. LOCKWOOD: We are going to go through these three questions, because I think they are critical to this process, and then we will take a little break, and then we will get into the presentations.

I would like people to comment specifically on whether there should be a point estimate or an upper bound of confidence interval, or both, in actively controlled trials that ought

to be the basis for making a decision.

DR. TRUSSELL: Well, if it's actively controlled trials, then what is going to matter here is the size of the delta. And you are between a rock and a very hard place here because, in order to get sample sizes that are actually doable, at least by what we normally think of as clinical trials of these kinds of contraceptives, you are going to have to make delta quite big, on the order of probably 3 or 4 percentage points.

But if you believe--but then you are stuck in the hard place again because do you really think that a contraceptive with a pregnancy rate of 1 1/2 percent is equal, clinically equivalent, to one with 5 percent, and probably nobody would really believe that. But, yet, that is what you are going to be--that is the regulator's dilemma is setting delta low enough is really going to drive up the size of these trials.

I think about this because I have been heavily involved in emergency contraception, and the failure rates are very similar for a year as

for per act for emergency contraception. So that is why these numbers are sort of in my head.

Do you consider an emergency contraceptive with a pregnancy rate of 1 1/2 percent the same as one with a pregnancy rate of 5? Well, no. So let's do 1 1/2 and 3, and then you are going to need about 8,000 women in each arm of the trial.

DR. TOBERT: Well, I mean, a lot does depend on what the rate is in the control arm and we have heard various numbers. But I regard you as the authority, Dr. Trussell, and your paper has 8 percent for the combined pill and mini-pill in typical use?

DR. TRUSSELL: That comes from data from surveys and that is typical use in the population. What I would say that you get out of analyzing all data in a clinical trial is typical use in the clinical trial because you count all of the cycles whether they are perfect or imperfect use.

In clinical trials, repeatedly, you are getting Pearl Indexes of 1.2, 1.5, 1.7, you know, somewhere between 1 and 2, and I don't see any

reason why that is going to change much even if you change the inclusion criteria for women who come into the trial.

Let's suppose it's 2, it goes up from a 1 1/2 to 2. Well, still, testing the difference, if delta is 2 percentage points from 2 to 4, do you really think that 2 and 4 are equivalent and, even with that, it is going to take a huge--it is going to be many times the size of the population in the current trials.

I don't know how to make it any clearer what the dilemma is.

DR. TOBERT: Clearly, this is a dilemma because you can't make the trials so impossibly large that nobody will want to do them. And I do take your point, if it really is only 2 percent, then you have to have a wide margin. Of course, the inability to rule out a margin of less than, say, 3 percent doesn't mean it is actually 5 percent as opposed to 2 percent. It may just mean that you can't do a trial big enough to do that.

I mean, you could have--to do a relatively

small trial, it could be 2 percent in both arms. But you still haven't eliminated the possibility it could be as large as 5 percent in the active arm. That is always the essence of the dilemma with these non-inferiority trials where the control has a low rate. But, I mean, the only alternative is to go back to the historical controls, which I thought we spent a long time eliminating.

DR. TRUSSELL: We did, but I think without understanding what the implications are. So, you are either going to have to have a large delta, or you are going to have to have extremely low power, and you are between a rock and a very, very hard place.

DR. TOBERT: You are going to have to have a large delta is basically it.

DR. TRUSSELL: Then that means that you could have the FDA approving a pill when the observed pregnancy rate for the new product, for example, is 6.

DR. LOCKWOOD: Well, no; the confidence interval is 6.

DR. TOBERT: If it was clearly inferior, if it was significantly inferior, to the control then, obviously, FDA wouldn't approve it. More likely you would have--you know, your standard would be 2 percent and your test might be 2 1/2 percent or something. But the 95 confidence interval might be 5 percent or 6 percent. It doesn't mean it is. You just haven't been able to rule that out.

DR. TRUSSELL: I would challenge you to work out the numbers.

DR. TOBERT: Well, in a preliminary way. I am not a statistician. Maybe Dr. Gillen has a better handle on the numbers.

DR. GILLEN: I mean, certainly, there is going to be sample-size inflation. I mean, it's notorious in non-inferiority trials that you are going to run into large values when you are trying to rule out with confidence interval limits particular thresholds.

Again, I think that you have to weigh what you are giving up going either way. I mean, there

isn't an easy answer here I think is what we have all come to.

When I was advocating them, I was kind of--you know, I don't want to get too technical, but in my statement I said, when ethically and logistically feasible, I advise doing active controlled trials, and that was the precursor. I was putting that in there because I realize that it does take a large sample size.

But also, if you are going to have this sense of comparability across trials, you are going to have to start somewhere and our gold standard is randomization. I mean, that is what it is. You know, if you are setting this delta limit too large, you are effectively allowing for a certain amount of threshold, okay, in terms of what you are willing to accept and the FDA has to weigh that.

They have to weigh ultimately what they are willing to accept in terms of non-comparability to historical controls versus potentially high Pearl Indexes coming from a comparative trial in an active control setting unless you are going to

force people to run these extremely large trials. I mean, that is the bottom line.

Ultimately, what the FDA has to do, in my opinion, is run through the types of non-inferiority margins that they would find acceptable under particularly valid circumstances, if they knew what the active control treatment effect was, and see if it's going to be feasible to require people to do this. And, if it's not, you have got to live with the fact that you are doing these impossible-to-compare historical controlled trials.

You cannot compare the--or you cannot solidify that you have comparability across groups in these trials and you have to live with that.

DR. LOCKWOOD: Okay. Short statements. Drs. Johnson, Stadel, and Berenson.

DR. JOHNSON: I can probably make a short statement. I guess I would ask the statisticians which is better, which is worse, to have a control trial where, yes, your power is not going to be great because you can't get enough patients to

really study it but, over time, you will have enough of these trials and then you can do some, you know, post-approval analysis of the pregnancy rates, or is it better to use the control group or the historic controls where it really is no comparison to how women use pills these days, or hormonal contraceptives these days.

It sounds like the better way to go is with the active controls and accept the fact that the power is going to be poor. But I could be wrong about that.

DR. STADEL: I think somebody needs to work out actual existing historical data and crunch the numbers and say look--and one could say to the given companies, you know, there is a lot of sentiment in favor of active controlled trials, why don't you look at what your experience has been with historical controlled trials and come back and say what you could do and what you would be interested to do.

I personally believe that to get good comparative data, you are going to have to augment

the active controlled trial pregnancy outcome data with surrogate outcome data because I think you are going to get the "n's" there with follicle suppression.

You can do the studies at a size where you can get a fairly rigorous comparison but I do recognize that that does edge onto the issue of whether there is consensus about the use of follicle suppression.

Thank you.

DR. BERENSON: First, I have a question for Dr. Monroe. If it was demonstrated that the efficacy of, say, a 10-microgram pill was less than that of a 20-microgram pill, does that necessarily mean that the FDA would not approve it at all, or does it mean that they could not claim to be as equally effective as a 20-microgram pill? After all, we write prescriptions for diaphragms and those are only about 86 percent effective.

DR. MONROE: Well, I think you have raised a very good point, and that is something we would like to hear from you because, again, we might want

to table all of this because we have later questions that address, if we have a product that is less effective, can this be conveyed to you, the patient, by labeling.

This is sort of what you are asking me. So perhaps we are really stuck, as to, I think, using Dr. Trussell--between the rock and the hard place here, because there is a cost and a gain.

I just want to clarify; I believe, when we wrote this document and we were talking about historical control, we are not talking about a comparison necessarily against another product. The basis for approval is that a drug be different than placebo. I mean, that is the sort of, I think, the bottom line here.

So, when we are talking about history, we are talking about the expected pregnancy rate in this population really not to be using any contraception. I think, at least, again, based on Dr. Trussell's table, and we would all tell patients that over a course of a year, we probably expect about 80 percent of women that are not using

any form of contraception to get pregnant over the course of a year.

So, that is what we had meant when we use the term "historical control." At least that is what I meant. My colleagues may not feel the same. So, we are not trying to say that this is better than a product we approved 20 years ago. That is not the control we had in mind.

We are talking about the background rate just like in certain diseases where it is unethical to do a placebo and, in essence, we are saying we can't do a placebo. So that is really all we meant by that term.

So, now, if we want to talk about comparisons against different products, that is a very different question and that is not, I don't think, what we are really asking you.

Maybe that is where we have gotten everybody confused here, because, clearly, we can't say a product approved 10 years ago, which had a Pearl of 1.5, did better or worse than a product that we are going to look at 10 years down the road

in a non-randomized trial.

That is not the question we are posing and maybe we have confused you somewhat. I think we raised this in a sense, and I think Questions 14 and 15 are really saying, what would you as practitioners feel an oral contraceptive must have in terms of efficacy--and, again, we are going back just to the confines of a clinical-trial because that is the best we can do--not in terms of, again, is this necessarily better or worse than that which was approved X years ago, because that product that was approved X years ago maybe was done in a context, and I think we have listed some of the conditions. Perhaps pregnancy tests were less sensitive. Perhaps women had lower BMIs. Perhaps whatever was going on may really have had a true demonstrated efficacy in that trial of a Pearl of even 0.5 or maybe 1.5, and maybe today that is going to be a 3.

Do we want products like that? That is really what we are trying to get at. So it's a question again that we didn't feel we could answer,

that we needed your thoughts as the practitioners out there that are the experts.

Dr. Slaughter may want to clarify what I have said or put it in a little different perspective. But I think that is what we are trying to talk about, what is really an acceptable rate.

DR. SLAUGHTER: No, I don't have anything to add to that, but that is how we had reviewed the historical controls. Again, when you go back and counsel your patients, what sort of limits or rates are acceptable to you or to your patients.

DR. LOCKWOOD: So, we are going to take your advice because we are stuck in mud here because we don't actually yet have a consensus on how high is too high. But maybe we will mingle during the break and chat more.

When we come back, though, we are going to have two presentations by Dr. Gilliam and Dr. Hillard, and we will move on to another set of questions. We will probably come back to this tomorrow morning and hat will give us plenty of

time to think.

DR. GILLEN: May I just make one comment?

DR. LOCKWOOD: Yes.

DR. GILLEN: Just maybe something to ponder is perhaps there could be a compromise in the sense that you must guarantee a Pearl Index below some threshold with respect to your confidence interval and also have a very wide margin on an active control, as well, so that you are powered--so that, on your primary experimental treatment arm, you have to guarantee that you have your Pearl Index above a particular threshold and set the non-inferiority margin much wider with respect to the active control to try and get at least some sort of comparability.

This is similar to what the EMEA is advocating. You know, they are effectively saying yeah, we are going to do the historical control, but we will have you do a smaller-scale active controlled trial.

You could kind of formalize this in some way to look at both endpoints to try and get at a

compromise of that and that would bring the sample size down some. It wouldn't give you obviously perfect power on the active control study arm, but you would at least have some sort of comparability. So, something to ponder.

DR. LOCKWOOD: I have been corrected. We cannot discuss this amongst each other. So we can think about what was just said and discuss it amongst ourselves internally. But we can take a break, so let's do that.

DR. PETITTI: When we come back, I really think that I absolutely have to have some clarification of the standard for approval of a contraceptive if the historical control is no contraceptive.

If that is really what you mean, then I don't see the purpose of doing any kind of studies in any more than about 10 people.

DR. BERENSON: That is not true, because you need to know how effective the method is so you can tell your patient.

DR. PETITTI: But, I mean, I think we

actually need some clarification of what the standard for approval is and what the kind of clinical trial we are trying to recommend to you is because we don't want to put women or the industry through any more trouble than they need to go through if there is a standard.

What I heard is that a margin of three pregnancies was too great against your historical control of 1.5, But now what I am hearing is your historical control of 1.5 is against no contraception?

DR. TRUSSELL: Furthermore, I mean, we have been told that the threshold used to be a Pearl of 1.5 and it has gone to something like 2. Well, I mean, clearly, that is not against a placebo of over 80 percent. So, they are inconsistent statements.

If you really would deny a pill on the basis that it had a Pearl Index of 3, then that is inconsistent with saying that you would approve a product as long as it prevents pregnancy relative to a placebo. I think that is what Diana meant.

DR. LOCKWOOD: So, break, and then Dr. Monroe will respond to that.

DR. WATKINS: Back in 10 minutes, please.

[Break.]

DR. LOCKWOOD: Okay. I am going to have to call another audible.

PANEL MEMBER: We don't know what that means.

DR. LOCKWOOD: You don't know what audible means? Oh, God, all my sports metaphors. So, we are going to change the play that was originally planned. It is even worse when I am in Europe and I am trying to use American sports metaphors, and they just look at me completely blankly.

At any rate, there really is a consensus, I think, to try to press ahead with the sort of critical set of questions and then we will move to Dr. Gilliam and Dr. Hillard's presentations.

I would like to ask Dr. Monroe when he returns to clarify exactly what is meant by historically controlled trials. I think the consensus of the group was, when we read this

question, at least it certainly was my view that we were talking about historically conducted clinical trials that had an endpoint, a Pearl Index to which they can compare as opposed to placebo.

Then I think we need to grapple with the issue of the size of studies that would be required if you are using non-inferiority and what would the upper confidence interval be and how realistic would it be to answer that.

Finally, what is the number--they want a number--what is the number that would be unacceptable as a pregnancy rate regardless of potential added safety that might be attributable to a new formulation. Okay?

DR. TRUSSELL: I just called back to my office to get the numbers that I was illustrating before, and I don't have 1 1/2 and 3, but I have 2 and 4. So if you think that you expect a pregnancy rate of 2, you declare delta to be two percentage points, so that you are saying really that 4 and 2 are clinically the same, there is no difference between them--they are clinically indistinguishable

or unmeaningful--and then with 80 percent power you would need 1,000 women per arm.

If you used 1 1/2 and 3, it would be well more than 1,000 women per arm. To just see how fast--I mean, to do 2 with a delta of 1 is 3,200; 2 with a delta of 0.5 is 11,500 per arm.

DR. LOCKWOOD: Is that for one year?

DR. PETITTI: Is that for one year?

DR. TRUSSELL: Actually, in a real trial, you would have to modify this, because I am not counting any lost to follow-up or anything like that.

DR. TOBERT: Based on one year duration of treatment.

DR. TRUSSELL: Yes; it is 2 percent over one year, yes. Or, if you were considering emergency contraceptive pill, it would be 2 percent per act.

PANEL MEMBER: Can we have that number again? Can we have those numbers again?

DR. TRUSSELL: Yes; you set up an equivalence trial where the expected pregnancy rate

is 2 percent and you allow delta to be two percentage points, then you would need 11,000--sorry; you would need 1,000 women per arm for 80 percent power, and there is no adjustment made here for dropping out or anything like that.

DR. JOHNSON: How about if you went up to 5, between 2 and 5, do you know the difference there?

DR. TRUSSELL: I don't.

DR. JOHNSON: Because I am wondering if we can pick a reasonable number of patients in a trial and pick a number that we all accept is reasonable.

I mean, we are kind of being asked an impossible question is what is an acceptable pregnancy rate. But if we can't agree to one that sounds like a reasonable trial and also a reasonably acceptable pregnancy rate, then maybe we can get to that.

DR. TRUSSELL: But, I mean, honestly, I mean, I would ask the clinicians here if you really do believe that 2 and 5 are equivalent clinically, it would make no difference to you whether you put

your patient on the 2 or the 5 percent pill.

DR. PETITTI: If you have a study with 200 women and 10,000 cycles, and you do a cumulative probability using a life-table analysis, what is the confidence interval on a pregnancy rate of 2? What is the upper bound?

DR. LOCKWOOD: Now, currently.

DR. TRUSSELL: I am just giving--

DR. PETITTI: No, no; I want to say that it is no different in a small study with only 200 women followed for 10,000 cycles--or 10,000 cycles, as is currently required by the FDA, and you have a measured 2 percent rate, and you calculate it correctly according to some kind of cumulative life-table method and not using this crazy Poisson where every single month counts basically as a observation, thus narrowing the confidence interval, the upper bound of that confidence interval must be around 5 or 6.

DR. TRUSSELL: I am answering only one question which is in a totally Poisson.

DR. PETITTI: Okay.

DR. GILLEN: In the follow up you are taking into account here, so you have got variability--

DR. TRUSSELL: You have got variability on both arms.

DR. GILLEN: You have got variability on both arms and so, once you take the difference in those two probabilities the variance is at, so it is twice as large if they were roughly equal.

DR. PETITTI: But what is the upper bound of that confidence interval on a cumulative life-table with 200 women?

DR. TRUSSELL: Big.

DR. GILLEN: It's 0.02 times 0.98 over "n."

DR. TOBERT: In any event, I mean, the sort of numbers you have given, 1,000 versus 1,000 are not undoable. I mean, this is not necessarily a single trial. The data can be pooled from all the trials in the marketing approval package, which normally would be 3- or 4,000 patients.

Maybe you could reduce the burden on the

sponsor by making them shorter trials since most of the pregnancies will occur in the early months we were told today.

So, I don't think the numbers you have come up with are undoable.

DR. TRUSSELL: 1,000 women in each arm?

DR. TOBERT: Pooled across all the trials. I mean, recent--what is Everett doing--I mean, they did 3,000--no, wait a minute, sorry, it was cycles, so I can't--this is the trouble, it is all in cycles and we can't merely convert it. But, I mean, I am just talking about packages in general, not necessarily contraceptive drugs, drugs in general, typical package, 2-, 3-, 4,000 patients in it, not for a year necessarily. But, I mean, that would be the number of patients.

DR. LOCKWOOD: Can I ask a question? It seems to me--and this is really reiterating the previous question and statement--that if you are doing a current study and you are comparing two different agents, and the sample size is 200 women over 10,000 cycles, there must be a pretty wide

confidence interval already with current studies that are ongoing. How is this any different? I mean, that is the size of the current studies.

So, they must be accepting pretty wide intervals for non-inferiority right now. I think probably the real issue here is exactly what would we accept as clinicians as the upper limit of a pregnancy rate for a product that we were comparing to a product we are comfortable with, we use all the time, we are familiar with, and we put patients on all the time.

So if, at the end of this theoretical trial, we find that--and I hate to use this--but we find that the Pearl Indices of the two different drugs were maybe not statistically significantly different, and the confidence intervals of the pregnancy rates were fairly wide, you know--let's say 0.2 to 4--how would that be different than looking at non-inferiority and seeing maybe a little bit wider confidence intervals where the means are pretty similar?

DR. TRUSSELL: Because they are not doing

equivalence trials, first of all, so they are not looking at those data.

Secondly, they are comparing the point estimates. That is what they are doing, that is what prompted Question 14.

Thirdly, at least as I understand it, there is a requirement not only for 200 women to complete the trial, but also for 10,000 cycles--well, 200 people completing the trial contribute 2,600 cycles. So, the rest of those cycles are coming from women who don't last the entire 12 months and, presumably, the requirement for having 200 women to complete the entire 12 months is just so you can get a one-year failure rate with a confidence interval that is not huge.

You could require 1,000 women, 1,000 cycles, but have only 20 women making it out to month 12, then you are going to get a much higher confidence interval for your 12-month rate.

DR. LOCKWOOD: So, you are comparing the means of the point estimates and assuming that it is adequately powered if there is no statistically

significant difference between the two, the agents are comparable? Is that what they are doing currently?

DR. TRUSSELL: They are not comparing two things. It is not comparing two things. The trials that are coming in don't require an active control arm and what they are looking at is the Pearl Index over 13 cycles. And my understanding is if your Pearl Index is above X, where X was something like 1 1/2, tough luck.

DR. LOCKWOOD: I thought you told us this morning that most of the current sponsored studies had active controls.

DR. TRUSSELL: A very small subset.

DR. SLAUGHTER: No. We don't have any that have been approved based on active control studies.

[Inaudible comment.]

DR. SLAUGHTER: That's right. Some of them include small comparator trials, comparative trials, but not for the purpose of establishing the efficacy. So our determination of acceptable

efficacy, so to speak, is not based on a comparison to another drug product. It is based pretty much on an accepted Pearl Index, and that is sort of where we took off with this, should we be looking at a Pearl Index, should we be looking at something else.

DR. LOCKWOOD: So, to use the language of a different kind of committee, do we want to restate our enthusiasm for the use of active controlled trials?

DR. BERENSON: I think that when those comments were made earlier, not everybody understood what an historical comparison group was, that that was a placebo group, so you may want to revisit that conversation.

It seems to me that there is an issue that are we going to state that every oral contraceptive that is on the market right now has good efficacy and, as long as they compare it to one of those, if they did a comparison trial, that that would be adequate or are you going to the question that James asked, where if there is a 3 percent

difference in efficacy, it is no longer equal? There seem to be numerous questions on the table.

DR. LOCKWOOD: I think what the FDA is interested in knowing is what is our tolerance of variable pregnancy rates, stated as simply as I think we can state, as it relates to approving a new agent.

DR. TOBERT: I think there is a little bit of potential confusion here between the confidence interval and the point estimate. I mean, if your test product had a pregnancy rate of 5 percent, I would think that would be unacceptable, and, you know, your control had the expected 2 percent. That is not the same as saying--but then you wouldn't be able to rule out non-inferiority of probably 7 or 8 percent.

But the likely scenario is you have got similar pregnancy rates, a couple percent in each case give or take a fraction of a percentage, but because you can't do the trials big enough, you can't eliminate a very small difference. You can't eliminate a half a percent difference. You can't

eliminate a 1 percent difference.

You maybe can eliminate a 2 percent difference. You can certainly eliminate a 3 percent difference. Is that correct, Dr. Gillen, with reasonably sized trials? That's with 90 percent confidence, which is the usual standard.

DR. GILLEN: Or 80.

DR. TOBERT: Or 80, you know, as long as it is 80, 90.

DR. LOCKWOOD: Again, I mean, this is a critical point, but is the consensus of the group that sponsors should be obligated to do active controlled trials?

DR. BERENSON: Yes.

DR. PETITTI: As a trial designer, I can get you, --I can promise you a 1.5 Pearl Index. I mean, you give me latitude to define the inclusions and the exclusions, and counsel women about how they are or are not going to use condoms, and I can promise you 1.5.

DR. LOCKWOOD: I take that as a yes.

Dr. Stadel.

DR. STADEL: I would only say I still think there is a lot of desirability of active controlled trials on a lot of data outcomes in addition to the pregnancy bleeding pattern, side effects, and so forth, if, in fact, people crunch their numbers. I have had companies who come back and say we can't do this for efficacy. Then one has the data in hand to make the decision rather than doing it sort of, you know, theoretically.

I still think there is a lot of reason to encourage active controlled trials and to augment the pregnancy ones with surrogate outcomes. I think there are a number of issues here that are important.

DR. LOCKWOOD: So I think this may be one of the few moments when we may want to take a vote as to the question of whether or not the FDA ought to require sponsors to conduct active controlled trials to have approval of new products.

DR. SLAUGHTER: Dr. Lockwood.

DR. LOCKWOOD: Yes.

DR. SLAUGHTER: I just wanted to say something about the word "require."

DR. LOCKWOOD: I am not surprised.

DR. SLAUGHTER: Because we cannot use that terminology. It is usually what we recommend in terms of the sponsor, how they look at the trials, and we say this is our recommendation based on certain sets of circumstances. but we don't use the word "require."

DR. LOCKWOOD: Thank you. I am seeking legal counsel here.

DR. GIBBS: Question.

DR. LOCKWOOD: Yes.

DR. GIBBS: Did we say this morning that there is any other division where the FDA accepts historically controlled trials for approval?

DR. TRUSSELL: What about devices. Contraceptives.

DR. GIBBS: Well, outside of the contraceptive world.

[Many comments off mike.]

DR. BERENSON: Isn't that because they are able to compare the placebo with other drugs and we don't consider it ethical to randomize women to a sugar pill?

DR. TRUSSELL: But you can't do--you have to do [inaudible] in most [inaudible].

DR. BERENSON: No; that is the reason why other antibiotics are not allowed to use historical controls because I think that it is not considered as ethical to randomize people to a placebo; is that correct?

DR. TRUSSELL: Well, there are plenty of drugs where you cannot randomize to a placebo. You have to randomize to whatever is the currently accepted product, at least with life-threatening drugs--I mean--

DR. KAMMERMAN: Occasionally--I have been with the FDA for 17 years, and there have been occasions where we have used historical controlled studies. I can think of cases where there have been maybe a lot of articles on published studies where--I am kind of grappling, I am not really good

at coming up with specific cases--but there can be circumstances where it is unethical to randomize subjects to placebo, and then you can use historical controls.

But they have to be well defined. You have to lay out all the rules upfront about similar populations, any caveats, do the trial designs differ. There is a document that--is it the level of evidence, or was it E9 that talks about historical controls and the circumstances they can be used.

So, in this situation, I would think that possibly there could be the use of historical controls in the way we know that, but the studies would have to have similar populations, and, as we have discussed here, there have been changes in ascertainment of pregnancies, changes in the doses and efficacy over time. So, that would be an issue.

I think sometimes, over time, the entry criteria has changed. Historical controls have often been open-label studies, so that is another

potential problem.

This particular case does not appear to be as clean in that sense as we might see in some other settings where historical controls have been used.

DR. LOCKWOOD: So, Question 5 posed by the FDA is: Is there a role for active controlled trials; if so, under what circumstances?

I would like to actually go around the table and get a yes or no as to is there a role for active controlled trials and a two-sentence--

DR. SCOTT: Charlie, I hate to do this, but can I just ask one quick question? Does the FDA approve a new product based on one trial by the pharmaceutical company that is proposing the new drug?

In other words, what I am getting at is, you know, it has been pretty well shown even with randomized trials that who sponsors the trial has something to do with the outcome, and are all the approvals based on what is submitted by, say, a company for a new product based on the study that

they have done? Are there independent studies that could be used for a meta-analysis, for example, when you need a lot of patients?

DR. PRICE: To answer your question, we have historically used, as I stated earlier, one trial, and for a product that was a non-new molecular entity. For a new molecular entity, we have historically required two or recommended two, and that was to reconcile the data if there was any difference in the data.

For more recent oral contraceptives, I am going to just say from '96 on, usually, it is one trial that has a minimum of 10,000 patients, 200 women completing those cycles, and it can go up to 12-, 15-, rarely 20,000 subjects.

DR. LOCKWOOD: Cycles.

DR. PRICE: Cycles.

DR. SCOTT: And it doesn't make any difference who did the trial?

DR. PRICE: No.

DR. LOCKWOOD: Okay. Do we understand the question? One more question.

DR. GILLEN: One more comment because--it is very hard to make a vote on this because there was an issue that was raised before we left and that is that there seems to be a little bit of a contradiction in the current standard for the threshold for historical control and testing superiority against placebo.

I think that the motivation of the FDA needs to be made clear. Are you really testing for superiority against placebo, and, if so, why use a standard of a Pearl Index of 1.5?

[Inaudible comments.]

DR. GIBBS: So, any Pearl Index we pick is going to be arbitrary, and I don't know that we are able to decide exactly what the right one is. Perhaps 3 is acceptable. Perhaps 4 is acceptable depending on other benefits of that product, and maybe we can't decide what is right for every practitioner and every patient. If we have a delta that is a little wider, well, that's fine.

DR. LOCKWOOD: Do you want to answer the question?

DR. GIBBS: No, but--

DR. SLAUGHTER: I think that what historically, what has been done is that consideration was given to what the pregnancy rate would be on those individuals not using contraception.

Relative to that, it was decided that the rate with a hormonal contraceptive should be less than that, on the order of less than 1 percent. That is how we came up with this 1-per-100-women-year Pearl Index.

I think, although we don't do direct comparison trials, over the years, the standard has been to compare to that Pearl Index of 1 which has slowly drifted up to a Pearl Index of 2.

Now, we are at a situation where we are trying to understand if that is really what we can do in terms of comparing drug products back in 1960 to 1980s where the estrogen levels were higher, trials were different, pregnancy evaluation was different to the present day scenario where the estrogen levels are lower, there may be better

detections, et cetera.

Should we be using this cutoff of 2 or less in a historical--what we have called historically based, based on that idea of 2 or less per 100 women years.

DR. LOCKWOOD: Without further ado, I really want to get to this critical question of is there a role for--we are being very temperate here in our wording--is there a role for active controlled trials, and a very brief comment, if so, under what circumstances.

DR. BERENSON: Charlie, can I ask a question?

DR. LOCKWOOD: No.

[Laughter.]

DR. BERENSON: Just before we vote, one question that I have been thinking about all day, as we make these recommendations to the FDA, is what impact will these recommendations have on the development and marketing of new contraceptives in the United States, because there is one thing to be a pure scientist and to put forth our ideal, and

there is another thing regarding feasibility.
I can vote on this easily as a scientist, but I have no idea what the feasibility is because I have no contact with the pharmaceutical agencies.
DR. LOCKWOOD: What I would like in people's comments, yes or no for the question, but then, in their comments, some sense of their tolerance of fairly wide confidence intervals of acceptability and any other comments they want to make that can inform the FDA as to the sentiments of the Committee members.
So, let's start with Dr. Johnson.
DR. WATKINS: And as you go around the table, please state your name so that the transcriber can accurately record your vote.
DR. JOHNSON: I would support using active controls. I think the biggest down side of that is that indeed we have to recommend to the FDA, I presume, what active controls are acceptable and again, in respect to Dr. Trussell, what range of confidence interval is acceptable.
Having said that, I think that that

provides much firmer data and, in the long term, I think will give us better oversight into how to approve new contraceptive choices.
So, is that what you were looking for?
DR. LOCKWOOD: Yep.
Dr. Stadel.
DR. STADEL: I think there is a role. I think the sample-size issues should be explored using actual existing data and that the outcomes that are feasible for a randomized comparative study should be defined after looking at the real data that are available on other sample-size implication.
I think that the surrogate outcome should be considered for randomized trials.
DR. LOCKWOOD: Dr. Petitti?
DR. PETITTI: I think that we should strongly recommend the use of active controls because it provides better information to protect the health of the public and to allow women to make truly informed choices about what they use for contraception.

DR. LOCKWOOD: Dr. Gilliam.
DR. GILLIAM: Melissa Gilliam. I think that randomized trials or active controls should be encouraged, but I think there has to be a role for examining the feasibility and also the potential safety and usefulness of a new method. If something is very novel or very different and will be highly acceptable to people who might be accepting of lower efficacy, then I think we have to take that into account.
DR. LOCKWOOD: So, the confidence intervals really should depend on the other potential attributes of the agent in terms of safety, and so forth.
DR. GILLIAM: Yes.
DR. HILLARD: Paula Hillard. I believe there is a role for active controlled trials, and I think this would give women and clinicians a firmer basis on which to make their decisions.
DR. PERLMUTTER: Johanna Perlmutter. I do believe there is a role for active controls, and I think that it is important for us, as clinicians,

to know the pros and cons. I don't think the numbers are as important as long as I know the numbers and I can give that to the patient when I am counseling them.
MS. SHANKLIN-SELBY: My name is Liz Selby and I agree there is a role for the active controlled trials. Just as a female, I would want to know the product that I was using had been compared to something that was currently being used, like a gold standard, so to speak, as compared to something from 30 or 40 years ago where, like you were saying. The incidence of obesity--I mean, just attitudes, the usage, I mean, was different 40 years ago. I would want to know that it was based on something a little more current.
DR. GILLEN: Daniel Gillen. I do believe that there is a role for active controls. I think a lot of this is motivated by the lower doses that are coming out and the moving benchmark that we have against historical controls and we need some sort of frame of reference for comparing new

therapies.

I realize that there are logistical constraints and I would contend that if the true goal is, you know, as was stated earlier, to really test superiority against placebo, that leads to a very wide margin for a non-inferiority trial and that can be taken care of in terms of sample size, I mean, if that truly is the issue.

On the other hand, if you are truly trying to compare efficacy against an active control, non-inferiority margins are going to be lower and the sample size is going to be needed to be there. But that is the only way that we can guarantee true efficacy against what is out on the market.

DR. BLUMENTHAL: Paul Blumenthal. I also believe that active controls have a role to play in contraceptive development and the contraceptive approval process. It may not be for the primary outcome, but perhaps for looking at specific subgroups or dissecting out potential confounders as we go through the approval process.

I don't have a predetermined limit on what

the effectiveness of a contraceptive up for approval should be. Rather, I would like to see the most valid and most highly generalizable data out there so that I can best counsel the patient about what she and I can both expect.

DR. GIBBS: Ron Gibbs. I also would recommend active trials for most circumstances except under selected circumstances where a trial might not be logistically feasible. After all, this is the standard for most drugs, the FDA has said, and we are all in the practice of evidence-based medicine. Wherever we can I think we should get the best evidence we can.

Regarding the delta, I have a high tolerance for a wide delta. After all, the patient and the provider would have a wide choice of contraceptives, weighing risks and benefits, some with very high efficacy and others with poor efficacy, and that decision ought to be tailored between the patient and her provider.

DR. TRUSSELL: James Trussell. I think we learn a lot from randomized trials that we

otherwise wouldn't know and I would see them as the gold standard. I do think here that we are going to face a problem of either lower power or pretty high delta and I am concerned about that.

I am less concerned if the results of the trial are actually given in the patient package insert so that they be there for people to see, and they can vote with their feet.

But I am particularly concerned about--I couldn't care about another "me-too" drug. It just doesn't bother me at all. So, if we discourage those, fine and dandy. But if we wind up discouraging really new products that either have superior efficacy or some other wonderful non-contraceptive benefit, then I would be discouraged.

DR. WESTNEY: Lenaine Westney. I agree with using randomized, controlled trials in oral contraceptive usage. I think that this is partially mandated by the expansion of the inclusion criteria to groups that were previously not included in older trials.

Additionally, I would hope that we would be able to identify what the efficacy is in subgroups and therefore allowing physicians to stratify who is at a lower risk and who is, for instance, a more compliant patient, possibly can be on a lower dose estrogen.

DR. ESPEY: Eve Espey. I also agree it should not just--it doesn't just play a role but it really should be the standard for investigation. It is really ironic that a medication that is used by so many women and that has such far-reaching consequences has such poor data.

I think for that reason alone, other than just sort of approving for efficacy, we need a critical mass of good data about hormonal contraceptive pills and the only way to do that is with a randomized controlled trial.

I agree with, you know, the wide confidence interval for efficacy because there are so many other things that women take into consideration when they use a contraceptive method including non-contraceptive benefits, and, as Abbey

pointed out, there are other methods that are out there being used that have much higher failure rates than oral contraceptive pills.

DR. PETERSON: Herbert Peterson. I agree there is an important role for randomized trials. I think the extent to which they should be recommended needs to be considered in light of the questions that those trials are trying to answer, and, if it is trying to address the question is it better than nothing, then the delta can be as broad as it can be just about. But if it is trying to answer Johanna's question about, I need to know how effective it is so I can counsel people, then the randomized trial may not get us there.

So, I think there needs to be a game plan to answer that question; if it is not pre-market, then post-market.

DR. BERENSON: I do think there is a role for active controls because many times these requests are to claim superiority over another agent or at least to state that they are equal, and you are only going to be able to do that with

active controls.

I think we can accept a very wide delta on these new methods as long as we know what the efficacy is so we can appropriately counsel our patients if the risk is higher for an unintended pregnancy.

DR. TULMAN: Lorraine Tulman. I also believe there is a role for randomized controlled trials in this. I was surprised to find out that, in fact, much of the research on oral contraceptives was not based on randomized controlled trials as I thought that was a gold standard for the FDA.

I think women deserve to have the state-of-the-art, sort of the science paradigm, which is the randomized controlled trial, and I think we also, in our discussions, need to clarify when we are talking about natural fertility rates versus historical controls, which I think needs to be cleared up in the FDA documentation.

DR. SCOTT: Well, I strongly support active controlled trials and, by that, I mean

high-quality, randomized controlled trials. In this era of evidence-based medicine, I mean, you just have to support this and there have just been too many misleading studies in the past with poor study design--I don't mean about contraception, but many things--and so we have to get to that stage.

I think that randomized controlled trials are very good for some things--efficacy, effectiveness, and so on. They are very poor for safety. So, I wouldn't say that that has to be part of the criteria.

In fact, I think, for safety, the only way you really get at very rare, serious complications and side effects is post-marketing. That is what has happened with other medications and they have turned up after they are used by thousands and millions of patients, and so on.

So, I think it should be randomized controlled trials primarily for efficacy and effectiveness. And I think that there are plenty of problems with these. But, you know, it is done with many other things, in multi-center studies. I

think that the problems can be worked out so I strongly support that.

DR. BUSTILLO: Maria Bustillo. I strongly support randomized, active controlled trials, and I think the most important is trying to figure out what is the tolerance in terms of the difference that you are going to be able to accept.

I think, thinking about it, probably it's not going to be too small because, again, what we have to do is be able to inform our patients about what that is and make an informed decision as to whether the benefit of the particular new pill is really worth it in light of what else is available.

I think if you don't do randomized trials, you are not going to have some of the other confounding things like bleeding, et cetera, that, potentially, the pharmaceutical companies actually are going to use as a marketing tool to make you prescribe that particular pill over another.

DR. LOCKWOOD: Charles Lockwood. Yes to the question, and I believe that using active controlled trials will address many of the concerns

that were raised in the morning session regarding applicability to the real world and changing populations, issues with lower doses and missed pills and BMI issues, and so forth.

I think, however, having stated that, this recommendation by the FDA should not create an undue financial burden on sponsors because we don't want to inhibit the potential for the introduction of novel and hopefully safer and more efficacious agents into the marketplace.

The total vote was 19 Yes, no Noes, no abstentions. Thank you, all.

I want to finish up by addressing Question 16 which is, should the Division approve lower-dose products that have apparent decreased efficacy and possible decreased risks of serious adverse events as compared to higher-dose agents in the classic model as the 20-microgram versus the 30-to 35-microgram ethinyl-estradiol compounds?

I think, as you discuss this, some comment--I am going to try to do this in 10 minutes--some comment about just how high a failure

rate would be acceptable for an agent with the potential for markedly better safety profile and particularly as it relates to thromboembolic disease.

DR. JOHNSON: Well, I think the key to that is knowing that there actually are fewer serious adverse events. I mean, do we know with certainty that lower estrogen levels do significantly lower the risk of VTE? So, I would think that, yes, it is reasonable to accept a lower efficacy rate but only if it is proven that, indeed, it has a decrease, significant decrease in adverse events.

DR. TOBERT: I agree entirely with that and, if you are talking about big trials, this is where you need a big trial. It perhaps could be a Phase 4 trial, but if you want to show that something is different with regard to a rare event like VTE, obviously, it is going to take a big trial.

MS. SHANKLIN-SELBY: I think, if there is a decrease in efficacy, I mean, that should be

conveyed to the woman along with the fact that the risk of being pregnant, particularly as you get older--I mean, for me, now, it is not an issue anymore. But, in my early forties, I mean, there would be more of a risk for me to be pregnant than there would be to take the pill.

I mean, I would want to know that I was protected against being pregnant rather than worrying about a rare, rare event because the risk would be much greater in pregnancy. So I would want that information conveyed to me. I mean, I think for some people, having a very low risk of side effects would be more important to them than the risk of being pregnant. But I think these should all be conveyed to the woman.

DR. LOCKWOOD: So, I think the sentiment of the Committee is a little paradoxical. If you are requiring evidence of safety, which would require enormous trials, talk about dwarfing anything we have talked about for using active controlled trials, prior to accepting a higher pregnancy rate, you will never get that opportunity

because no one will ever do these enormous safety trials.

Alternatively, if you are waiting for Phase 4 studies and other ascertainments of safety it would take years potentially to acquire, you will never have approved the drug in the first place.

So, let me rephrase the question and say, is there a number--and we will use the Pearl Index as much as I hate to use it--is there a number, a Pearl Index number, above which you would be uncomfortable going beyond assuming sort of current estimations of Pearl Indices in non-randomized, non-controlled trials.

Is there a number? Is it 3? Is it--you know, for a 10-microgram ethinyl-estradiol compound that has minimal effects or no effects on hemostatic parameters and have every reason in the world to expect that it would have a lower incidence of venous thrombotic sequelae, would you accept a Pearl Index of 3?

DR. PETERSON: I think that goes back to

the issue about any method being better than placebo. If you look at the--let's say it goes back to the 98 and 99 percent. If we really believe that if you take a pill every day, with a 35-microgram pill, you have got a 1 to 2 percent chance of getting pregnant. Well, if the 20-microgram, or let's say 10-microgram pill, was five times that, so it's 5 to 10 percent, that is still a lot better than not using any method and it is better than some other widely used methods, not so widely used, but approved-for-use methods.

So, the question would be sort of what is the benchmark against which that should be measured, is it against other pills, which then gets back to the James delta issue, and say, well, how sure are we that it's not--if 10 percent is too much, then how sure are we that it is not 10 percent, which gets into this issue of if we are using randomized trials of power and sample size and delta, or is it, well, it really doesn't matter in terms of approval.

It would be approved if it's more

effective than nothing, which is a slam dunk in sports metaphors. But if that is not true and it is an issue, it is really not being compared to spermicides. It is being compared to other pills and a 5 to 10 percent rate would be unacceptable given that there are other pills with similar safety profiles.

Then you start saying, well, how similar is it theoretically? Reducing it from 30 to 35 to 20 might reduce the risk of VTE but the limited data we have doesn't help us in that regard. In fact, some can interpret it as some limited evidence against there being an improvement.

So, I think it comes back to this issue of what are we talking about is the framework against which we are trying to make a judgment.

DR. LOCKWOOD: Dr. Monroe.

DR. MONROE: The framework is not really against placebo. Placebo is just telling us that it is effective. And I think we have certain expectations for any form of hormonal conception and it is certainly well below or much better that

what we would expect for barriers, and so on.

So, our expectation is not just that it be better than nothing, but that is what we meant when we used the word "historical" because, in order to approve a drug, it has to show benefit. But now when we are getting down to the specifics of a hormonal contraceptive, we are really looking a risk/benefit because some of these other methods, in terms of the method itself, have virtually no risk if there is a risk of pregnancy.

We have, as you have alluded to, these numbers, whether it be a 2 percent, whether it be a 3 percent, whether it be a 4 percent.

So, we are asking you really, in this very narrow range of 1, 2, 3, 4, what are your feelings about an oral contraceptive or transdermal or an intravaginal because those are the types of products we are asking you to address in terms of do you have a cutoff if it's conveyed.

If we do these non-inferiority studies, we wouldn't know for certain unless they are very large. If I understood what you have said, Dr.

Trussell, it could be a number really around 4 percent and there may be a balance where you can still work both with an absolute cutoff and a non-inferiority and perhaps mitigated in some way as Dr. Gillen did.

So, I just want to correct perhaps another misconception that I introduced. We are just not expecting you to agree that as long as it is better than placebo. We expect you to sort of continue to think in that context of what the expectations are for a hormonal contraceptive product, which is certainly highly effective, but how highly effective would you folks find to be acceptable.

If you could provide us with some range in that area because that is an important question to us in our moving forward because of the complexity that you had indicated.

DR. BERENSON: I think it's important when we consider Question No. 16 with regards to some of these pills may be advantageous for special populations. So, it may be not what you would recommend to your average patient. But perhaps you

are a breast-feeding mother when your other choice is a progestin-only pill. Maybe 10 micrograms of estrogen plus progestin is better than a progestin-only pill. Maybe your patient with lupus, if you don't feel comfortable providing her with a higher dose estrogen pill.

So, again, I go back to yes, I think they should be approved. But physicians and patients both need to know the risk they are taking when they are using a less efficacious pill, and the manufacturers. The real question should be what range can we use so the manufacturers can state it is equally efficacious.

DR. TOBERT: There seems to be an underlying question here also about whether you can simply say, because the pill has got less estrogen in it, it is going to cause fewer VTEs. I think people are saying that trials to show that would be too big so you would have to take that on faith. But can you take that on faith.

I mean, it seems reasonable but it certainly isn't a slam dunk, I think. There is a

paper in the background package showing you couldn't show a difference between 20 and 35 microgram estrogen pills.

DR. LOCKWOOD: I was using it as an example. I mean, let's assume in this theoretical discussion that you could collect a surrogate that would be extraordinarily useful in terms of predicting venous thrombotic risk.

What is the upper limit of a point estimate of pregnancy--I think we will forget about the confidence interval at this point--that you would accept as warranting approval?

DR. GILLEN: So,I w I would actually interpret this this question more as what is the worst- case scenario I would be willing to accept. So, the idea is let's suppose that, with this lower dose, we have zero side effects. So, we know we have zero side effects with condom use, for example. So, I would want to guarantee, at minimum, that the lower limit of my confidence interval is better than the point estimate associated with condom use where I have no serious

adverse events that are coming from that is hormone-replacement related.

So, that would be my worst-case scenario if I were starting to set a threshold as I would want, at minimum, my worst case for my confidence limit to rule out my point estimates from condom use.

DR. SCOTT: I just wondered whether this a situation where the FDA could name the control preparation, in other words, like Abbey and you said, Bert.

This would be a perfect place to compare it with the progestin-only pill, or maybe two comparisons, the progestin-only pill and the 20 micrograms of estrogen, not just to come up with what your acceptance level is as far as pregnancy rate, but at least you come up with the information about this is the efficacy of this as you go down with the estrogen dose.

As far as the safety of these, I think that is almost impossible. It would take a huge amount, a huge trial, and I think that it is

unlikely that they would be more unsafe than the 20-microgram pill.

So, I think the more important criteria is what is the efficacy and that is, I think, accomplishable, at least to come up with the pregnancy rate, between those two. And there are certainly already progesterone-only pills on the market, so that should be known. There are already 20-microgram pills on the market, so that should be known.

So, this could be compared I would think. If they are going to be studied, maybe those are the comparisons that should be done.

DR. TRUSSELL: Just as a note, we have now one progestin-only pill on the market. We are down to one.

But I would say that I wouldn't accept a tradeoff of efficacy for a theoretical benefit. I would accept a tradeoff for a real benefit that has been demonstrated and then how big that tradeoff would be would, to me, depend upon what that benefit is.

For an example, I mean, suppose it doubled your fun and sex life. Well, we may be able to trade off a lot for that. But without knowing what that is, it is very difficult to say what you would trade off.

If, in fact, you can demonstrate that the tradeoff is X in terms of efficacy, I guess Y in terms of something else, you could put it on the product and let people vote with their feet.

DR. BERENSON: It was suggested earlier that the lower limit of acceptability should be the condom, and I would say it should be a diaphragm or maybe a diaphragm plus spermicide, which does have very small risks, such as a UTI. But they are very small. So, I would not want to prescribe something to my patient that had lower efficacy than that.

DR. TRUSSELL: Even though only three people use the diaphragm?

DR. BERENSON: Good point.

DR. HILLARD: Just building on what Dr. Trussell has said, I think that comparing to the condom in terms of efficacy is one comparison. On

the other hand, if a given pill had other secondary outcomes--for example, significant relief of dysmenorrhea--that is a plus that would be an advantage of that particular pill over the condom.

So, I think that that is a situation where one takes into account much more than just the efficacy per se.

DR. LOCKWOOD: We really are going to have to move on, so I am going to ask Dr. Gilliam to prepare for her presentation.

Just to summarize what I think is the sense of the Committee, that they are uncomfortable giving you a specific number, that there really seems to be a mix of attitudes in terms of the requirement for documentation of much greater safety, or other benefits beyond safety, as being required to have been demonstrated for some of the Committee members to agree to a significant increase in the upper limit of efficacy.

Others I think would accept surrogates or be a little bit more liberal, or conservative, depending on your perspective, but I don't get a

consensus from the group on this particular question and I didn't think we would.

DR. JOHNSON: It seems to me there was somewhat of a consensus that as long as women are informed and providers are informed, that there isn't really a lower limit of effectiveness as long as it is communicated to the patients within the realm of other contraceptive choices.

DR. LOCKWOOD: Very good point. I think caveat emptor was the message that everybody wanted to convey.

Topic 3 - Translation

DR. GILLIAM: I was given the task today to provide you with food for thought about introducing effectiveness into efficacy trials.

I know a number of people have said that we need to introduce these ideas and bring as much real-world data into trials. Now, this is something I am very much focused on. What I study is how do real people use contraception, but yet I am still conflicted on this.

So, I am going to just kind of give you

the universe of thought as I see it on this topic.

I will give you the background why I think this is important and give you some ideas about what we might be able to learn from other disciplines and then think about some practical approaches to adding effectiveness to clinical trials, and then give you a potential framework for doing this.

[Slide.]

So, in my mind, this is a topic that links biology, clinical world, and public health. And so the public health that we are talking about is the high rate of unintended pregnancies in this country and, even though this is earlier data from the National Survey of Family Growth, the proportions still have not changed. About 50 percent of pregnancies in this country are unintended and, of those unintended pregnancies, half will end in birth and half will end in elective abortions.

[Slide.]

We have talked a lot about the people who actually use contraception but not so much about

the ones who don't. This is again old data and now that number is more like 7 percent to 11 percent of women do not use any form of contraception. Of that percent, they account for about half of all pregnancies.

What is also important is that the other half of unintended pregnancies occur among women who are using contraception, so there is a lot of data to be had about people who are abusing methods and using them incorrectly.

[Slide.]

We have racial disparities and demographic disparities among the women who experience unintended pregnancies. Rates are highest among women who are age 15 to 24, unmarried, black, Latino, and below 200 percent of the federal poverty level.

[Slide.]

What you notice about adherence is that it really depends on method selection. The leading methods are the oral contraceptives and sterilization. But white women are more likely to

use oral contraceptives while African-American and Latino women are more likely to use sterilization. So, in my mind, that means that women are probably self-selecting for the methods that they are best able to adhere to.

Similarly, poor and low-income women are more than twice as likely than higher-income women to use the three-month injectable.

[Slide.]

The topic of Efficacy versus Effectiveness.

Archibald Cochrane, a wonderful epidemiologist, asked, "Can it work?" That is how he described an efficacy study. What we want to know is whether, in an ideal circumstance, can a method work. This is a very essential first step for testing a drug.

Then he went on to ask the second question and described an effectiveness study which says, "Does it work?" When we start to get beyond the ideal circumstances of an efficacy trial, will the contraceptive work in that setting?

[Slide.]

Effectiveness is affected by many things. It has to do with patient adherence, the personal characteristics of the patient, the patient's partner, social and cultural context for method use and aspects of the contraceptive method itself, the inherent efficacy of it, as well as a lot to do with the healthcare and delivery system, how well does the provider adhere to what we have suggested, does insurance cover a method.

[Slide.]

So, where might we introduce these ideas of effectiveness or the real world into clinical trials. I think probably a lot of this has already been done in other fields and other disciplines and I would say that the social scientists have thought about these ideas quite a bit.

Some of these I will go into in more detail, but what the social scientists have thought a lot about are issues of cultural sensitivity and cross-cultural research.

For example, if I asked a question of a

person who is Caucasian, and then ask the exact same question of a Latina, will she hear the question in the same way and will she provide a similar answer. So, these are very sensitive questions, but I think they have to be taken into account when we ask survey questions in diverse populations.

Similarly, social scientists have added theory to research. Business has also given us some tools. For example, marketing analyses; how do products--how are they preferentially uptaken by various populations. Then there are also models such as complex decision analyses; how might a person choose one surgical technique over another, what are the factors that go into that decision-making.

[Slide.]

I am just going to give you an example of how we applied social scientific theories to research, what that might look like. This is something that is useful to me in my research, which is an ecological theory of human development.

What it says is instead of that ideal person that we study in a clinical trial, what we want to think about is the individual as being sort of like the inner doll of a set of Russian dolls. So, we start to consider her family, her neighborhood, her community, and society and the way that she takes oral contraceptives or whatever contraceptive method she uses.

So, instead of just the biologic model that we started with, we are thinking of a biopsychosocial model or a bioecological model. That would be the way to start to redefine the frameworks that we would use.

[Slide.]

I think that is complicated, so some practical ideas. One is to increase the diversity among the research participants and I will talk a little bit about recruitment and retention of diverse populations. The other is to improve the measures of acceptability we use, and then the other thing I was asked to comment upon is the role of technology.

[Slide.]

So, for diverse populations, and I just give the example of if you were to try to recruit a Latina patient population, you would need a bilingual research team, Spanish language study materials.

One thought is to start to encourage companies to work through community-based agencies in which you actually befriend the staff of the agency. You have the staff participate in the research so that they can explain to the people that they are recruiting what the research experience is like. The agency or the staff members actually serve as the primary recruiters. Then there is the snowball recruitment where one woman tells another about a trial.

Engage leaders in the community in the project. Engage trusted people like mothers or partners and family members. Provide food, transportation, child care, and provide opportunities for the community to understand what the research findings are, and then also engage the

participants as experts, you will provide us research information that we may not otherwise have.

One of the reasons why people often don't want to encourage diverse populations to participate is it can be a challenge for longitudinal follow-up. We want the ideal study; you actually have the same people who start the study, finish the study. So, I thought I would just provide some potential retention strategies.

One is convenient hours of operation. This may be that you have to have early morning clinical trial sessions or late evening sessions. Convenient locations--perhaps alliances with local healthcare facilities. Working through social workers or other providers who are already trusted in the community. Offering meaningful incentives; it may not only be financial, but perhaps diapers or some other thing that is meaningful but that might have to be determined by the population that you are trying to recruit.

I think there is a really strong role for

qualitative research to understand what methods populations need, and then again disseminate the results back to the community.

[Slide.]

We have talked a little bit about the role of acceptability measures. It is important to realize that--the current methods, we are typically using surrogates, hypothetical acceptability through a survey, or we ask, does a woman actually uptake a method or does she continue to use it, and we say, well, now we understand whether that is acceptable.

But what we also know is that an acceptability study doesn't necessarily predict what will actually be used once a method is introduced into clinical practice. Similarly, if you define acceptability in a narrow population, you don't necessarily know that it will have widespread use. So, for example, the intrauterine device is highly acceptable--among people who use the intrauterine device. I won't talk about my own contraceptive method here, Dr. Trussell.

[Slide.]

The other thought is adding additional tools or measures. Again, we touched on this already. One suggestion has been to try and understand, not necessarily is the contraceptive acceptable, but to try to start to parse what the actual method characteristics are and then to rank those in studies that have looked at characteristics once they have ranked them, and then have the participant try to decide to what extent she thinks a given method represents those characteristics. It has been shown to be a more accurate way of measuring contraceptive characteristics than just is this method acceptable.

I mentioned a little bit ideas of using decision-analyses techniques, but the idea is to either use things like vignettes or look at the context and how a woman might think about whether a method is acceptable or not.

The other is to provide additional information about characteristics. So we typically

think about things like bleeding or amenorrhea. But women also care about libido and other lifestyle factors.

The final is to try and use, potentially again, vignettes to get more realistic information about what potentially the use behaviors with a method might be, so whether that includes vaginal insertion or patch application.

[Slide.]

The other question was about technology. While I think technology is wonderful in terms of the idea of getting more accurate data, and these might be monitored pill packs like we saw in the Potter trial or personal data assistance, or even use of two-way pagers in studies of teens where you actually signal them to input data, and those do have their issues.

There is also a role to even try to introduce into a study methods that might improve compliance--so that, for example, would be the two-way pagers--and then maybe the technology would actually be applicable to actual clinical practice.

But as this cartoon says, "Didn't you get my e-mail?" we have to be very cognizant of the fact that there is a technological divide and technology is not the answer to all problems.

[Slide.]

So, if you look at what Archie Cochrane originally said, he actually had a third component. He asked about the efficiency of trials, "Is it worth it?" What he was asking is saying that the third way of studying it is actually to do outcomes analysis. And so I am going to kind of twist this idea of is it worth it in a couple of different ways.

[Slide.]

A number of years ago, the Institute of Medicine published a monograph called "New Directions in Contraception" and they suggested this idea of the "Go" or "No Go" approach.

Typically, what we do in phase 1 trials is--in trials, we ask acceptability towards the end of a study. What they suggested was we could also ask this at the beginning by adding the input of

women, partners, providers, the people who actually may affect whether a drug is effective--not efficacious, but effective.

Well, you bring that in early, and then you ask "Go" or "No Go;" is this going to be a method that is acceptable to women.

[Slide.]

The other way of asking is it worth it is, is it worth it to start to bring this muddy information about what women will do in the real world into the earlier stages of a clinical trial.

I think my personal feeling is that it may very well be worth it. You may affect contraceptive access and knowledge in specific populations. It can be--it is part of a public-health commitment to medically underserved women.

It is part of a public-health and clinical commitment to getting access for minority women in particular, providing them with access to new methods through the clinical trials process, and there is the hope of development of culturally

acceptable contraceptive methods.

[Slide.]

I think we can also ask is it worth it from a cost-effectiveness trial. Actually, Paul Blumenthal, who is one of the authors on this paper, is sitting here, but this is a paper that I read very early in my career and it has just been very meaningful. It is entitled "The Boom and Bust Phenomenon: The Hopes, Dreams, and Broken Promises of the Contraceptive Revolution."

It is a very elegant paper and it brings up a lot of complex issues. But one of the fundamental ideas here is that there are so many contraceptive methods that, very early in the testing phase, show great, great promise, and as they appear and emerge on the market, they again are touted as revolutionary, they are going to be absolutely wonderful.

Then, repeatedly, what happens is they fail. And you can kind of think about method after method of methods that actually fail once they get into clinical practice because there is something

that we missed early on, or there is some perception that women have that was not anticipated.

So, I think when we ultimately talk about to what extent we should link the biologic, the clinical and the public health aspects of a contraceptive device, earlier on in the process of studying it, perhaps we can start to address these issues of the boom-and-bust phenomenon of contraception.

[Slide.]

So, if you put it all together, my thoughts would be that you would add theoretical frameworks early. You would add the theory early, perhaps considering qualitative research, better measures, and even think about the "Go" or "No Go" approach, that if something is really unacceptable to women in very early phases of development, even if it really works well, that may be a "No Go."

Think about using diverse study populations. We have talked about high BMI, but I would also say that if you start to target the

women who have the most difficulty with adherence to contraception, then we are really also talking about racial and ethnic diversity in clinical trials.

Then to really start to measure actual contraceptive use behaviors, not just the ideal, but really to understand what the pill-taking patterns might be, whether you need technology to do that, but take into consideration that there is as technological divide for some women. Then, to think about efficiency, that what we are really trying to say this is eventually going to be cost effective, because this is a method that is acceptable to women.

[Slide.]

Here are my references.

[Slide.]

And that is Chicago.

DR. LOCKWOOD: Paula, if you could give your talk, and then we will take questions on both. Then we will get to the questions and then we will have at least the presentation of the cycle-control

issues. We will probably save the questions for tomorrow.

DR. HILLARD: I was asked to talk a little bit about the real world and thinking about effectiveness and safety.

[Slide.]

I have to say initially, this is not the real world of Cincinnati, Ohio. This is not the Ohio River. It's the Liao River in China. What I would say, it is beautiful. It is very beautiful there. So is the Ohio.

However, what I would say about the real world for me as a clinician is that my patients are mostly adolescents. I was also asked to focus on adolescents as a population, as well, so I will talk a little bit about issues and effectiveness in adults compared to adolescents.

[Slide.]

We have talked earlier about what are we looking at in terms of effectiveness of a method versus efficacy, thinking about perfect use versus typical use. This is the table that we all know

from Contraceptive Technology, and really what we are thinking about is what is the difference between this column and this column, how do we get from what the effectiveness would be or the efficacy would be in perfect use versus typical use. And that is what I want to address a little bit further.

[Slide.]

We have talked previously about what are some of the differences, things that influence efficacy beyond the inherent method efficacy. And things like the user characteristics are certainly important.

The consistency and correctness of method use is going to be what I will comment a little bit more about, but keeping in mind that other factors that have already been mentioned, fecundity, frequency of intercourse, age, parity, these things are interrelated, but also impact the consistency and correctness of use, as well. So, I am going to focus on the consistency of use.

[Slide.]

For example, if we look at the failure rate of oral contraceptives in the first year, separating by age and by income, you see in the yellow bars poor and low-income women. You see in the green bars higher income women.

The question that I ask the medical students is does this mean that oral contraceptives are metabolized differently by women of low income. I don't think so. What it means is that there are real differences in women's lives. There are differences in the orderliness or disorderliness, as has been discussed earlier, for poor women versus women of higher income.

There are differences in access to care. There are lots of differences between these two populations that affect what we come out with as a bottom line in terms of the failure rate of oral contraceptives.

The other thing that is interesting to look at here is if we separate it by age, there are not huge differences by age if one looks at it in this way and I would suggest to you that

adolescents don't always do worse than adult women in terms of being effective and using a method of contraception. Some adolescents do, and there are some things that we can say about that, but it is not true--and I will defend adolescents forever--it is not true that they always do more poorly than do adult women.

In this particular way of dividing things, women in their 20s, early 20s, do a little more poorly than women in other groups.

[Slide.]

Just a word or two about the terminology, and I felt I had to say this, because many of us bridle at the term "compliance" and yet that is really what I am going to be talking about. I will find myself sort of falling back into using the term "compliance" in part because that is what has been used very frequently. But in many ways if you think about it, it is a fairly paternalistic and certainly a clinician-centered term. Carolyn Westhoff calls it "cheerful obedience."

So, the idea that my patients will do

exactly what I tell them to because I say so is certainly outmoded. It really fails to acknowledge that you are trying to establish a therapeutic alliance with the women that I am seeing as a patient. So my patients participate in the decision-making and decide and vote with their feet whether to take their pill today or not.

The term that has been proposed as an alternative is "adherence," and I think back to when I was in medical school, I have to think about platelets and platelet adherence, so I find that word a little bit difficult, as well.

Another phrase that has been suggested more recently, and I found this one used on a listserv for adolescents, the Society of Adolescent Management, is that adherence is a part of a bigger picture for illness management. And yet that doesn't fit well in this particular regard, as well, because we are not treating an illness when we are talking about contraception.

The term that has been used in thinking about contraception, and I like this one much

better, is thinking about successful use of contraception. It is certain woman centered or even couple centered if one broadens it a bit. And basically what we are talking about is women being able to meet their own family planning goals. I think that is a much more reasonable way of thinking about it.

[Slide.]

But, falling back again to use the term "compliance," what does that mean? Well, it means correct use, it means consistent use, and it means ongoing or continuing use over some period of time.

[Slide.]

It has been said that we should think about contraceptive compliance in the context of compliance with other medications and, if one looks more broadly at the compliance literature, people have trouble taking all kinds of medications. So, it is not just oral contraceptives that women have difficulty taking.

The other issue is that the potential consequences of failing to take contraceptives,

oral contraceptives, is potentially pregnancy. On the other hand, that is not an immediate consequence. If I were going to be struck by lightning if I didn't take my pill today, that is an immediate consequence.

On the other hand, the consequence is potentially nine months down the road, so that is a bit further down the road, and particularly adolescents, particularly younger and middle adolescents, are not developmentally equipped to be thinking about the consequences of their actions, particularly the consequences nine months down the road. And that is one of the reasons that we would prefer adolescents to postpone sexual activity is to get to a point when they are able to think about the consequences of their actions.

Again looking more broadly in the context of the compliance literature, there really isn't any consequence that is so severe that it assures complete compliance.

[Slide.]

To look at a comparison of adherence with

other medications, there clearly are some differences. Just think, for example, about antibiotics that one might take for an upper respiratory infection.

You take your antibiotics for an upper respiratory infection and probably I would acknowledge for myself, and I think most of you honestly, as well, if you have a medication you are supposed to take four times a day for 10 days, you probably don't take it four times a day for the full 10 days, But, at any rate, your anticipation is that you will have a decrease in your symptoms.

Again with oral contraceptives, you are avoiding pregnancy and that is a down-the-road consequence. With antibiotics, it is a positive result. For many women, or at least for some women, the consequence of avoiding pregnancy has some ambivalence associated with it, and that is true for adolescents, as well as for adult women.

Women have many choices in terms of options for contraceptives. The choice of antibiotics is usually not made by the woman

herself. Oral contraceptives need ongoing adherence to the medication.

There often are complex interactions with the partner or in the context of family or social milieu as has been discussed with the previous presentation very nicely, describing many of the things that we need to think about and the places where women are living themselves.

[Slide.]

Michael Rosenberg talks about the consequences of improper or inconsistent use of oral contraceptives and estimated that about a million unintended pregnancies a year are a result of this inconsistent use, so I think it is something that is important for us to think about.

One could quibble with that particular number but, at any rate, if the inconsistent use of oral contraceptives results in a pregnancy, then those numbers add up.

[Slide.]

Looking back at the compliance literature, it has been stated that, "The accurate measurement

of compliance is not easy; easy measurements of compliance are not accurate."

So, those who really study this as their lifetime work acknowledge that this is a challenge to do.

[Slide.]

But if one thinks about measuring pill-taking, there are a number of ways that one could do it. One could look directly, directly observed therapy, as one thinks about might happen in an inpatient psych ward with observing individuals taking their antipsychotic medications, not something that happens regularly with oral contraceptives.

Measuring biological markers in blood, again not particularly practical. So, for the most part, in thinking about clinical trials, we are using indirect methods.

For the most part, self-reports, sometimes pill counts and looking at the pill package, how many pills remain in the package. Rates of prescription refills in systems where one can keep

track of this is one way of looking at it. Assessment of the clinical response.

And the assumption that if you have a pregnancy, it therefore implies that the method was not used correctly is not true with oral contraceptives. There are method failures that occur with oral contraceptives.

We have mentioned the electronic medication monitor, and I will say just a little bit more about that. More commonly in clinical trials, the patient diaries are what are used. So these are just different ways that one could do it.

[Slide.]

I show you here, just to make a couple of points, one related to age and the other related to what sorts of things are necessary in taking oral contraceptives consistently and correctly.

This is a study by Deborah Oakley in which she looked at what she termed "micro behaviors," and the pill-taking behaviors.

Taking the pill in the same order; one would assume that that is something that ought to

happen and yet you see that it didn't happen 100 percent of the time for each of these groups, for each age group. So not all individuals in the study took the pill always in the same order or said they took the pill always in the same order.

Taking only one's own pills, again, you assume that that is the case. Here is a situation where the youngest teens, those younger than 14 perhaps shared their pills a little more, maybe with their sister or their girlfriend, and I think that says something about access to care, as well.

But some things that we really do assume happens with oral contraceptives--that is, taking a pill every day--if one looks here, the group of women over the age of 30 did best. The group in the middle still didn't do so well. About 40, 45 percent of those--only 40 to 45 percent of women took the pill every single day.

The group that did the most poorly, and I think that it is important to notice that here, is this pink group right here, those who were younger, 14 or younger, that those are the individuals who

have the most difficulty doing all of these things that we assume that they will do when we hand them a pack of pills.

But again looking at the other behaviors, the teens who were 15 and older didn't do things all that much more poorly than did older women.

[Slide.]

This is a study that we have been referring to in terms of the electronic pill packs. I like to look at it this way because, graphically, I can think about it a little more easily.

This is comparing what women said in their diaries with what the electronic pill pack said in terms of when that pill was punched out of the packet.

Several things to note here. One is that, in terms of women saying that they missed no pills, many women said that they missed no pills on their diaries, anywhere from 30 percent in Cycle 1 to really in the Cycle 3 only 20 percent said they missed no pills. In reality, it was much higher

than that.

Sixty percent or so, really, in reporting on the diary actually missed no pills. So that is reasonable. On the other hand, missing 3 or more pills here, many fewer women said that they missed 3 or more pills than the diary reported, and, as has been alluded to, as well, by the third cycle, it didn't get better. It got worse.

So, many explanations have been given for why that should be the case. One is if you get away with it once in the first cycle or in the second cycle, if you don't get pregnant, maybe it doesn't matter quite so much that you missed 3 pills or more.

I think there are some interesting things that we could take from this. You know, if you couple this in terms of technology with some sort of a reminder or an alarm that might go off, that might be helpful.

Many of the teens that I see in my practice, in thinking about and brainstorming with them, how they can take a pill consistently every

single day, many of the teens that I see set their cell phone for an alarm, and they are already using the technology.

So, while there may be some variation across different populations, at least many teenagers who have cell phones--and most teenagers have cell phones, even those who I wonder how they are affording their cell phone--if that technology can help them in taking the pill, then that is something that should be used.

[Slide.]

Looking at some other studies that have looked at pill taking, 50 percent of young women report imperfect pill use during a given cycle and about 25 percent of pill users missed two or more pills during a pill cycle. These are some studies from Potter and Oakley that suggest that, even in women who are adults, there is imperfect pill use.

[Slide.]

What is required for perfect use; just when I think I have heard every single way to take a pill pack incorrectly, I hear some other way from

my patients. And so they still have lots of difficulty in taking their pills correctly, taking them backwards or forwards or up and down in the credit card packs, taking them vertically, and then zigzagging vertically, you know, all sorts of ways that you can possibly imagine are ways that our patients are sometimes taking the pill. So, that is not always intuitively easy. The on-again, off-again use of the pill is something that we see quite frequently among adolescents.

[Slide.]

Continuing use again is something you can look at lots of different studies over time and the studies suggest that, even among adults--this is looking at 6 months--only about two-thirds of women using the pill at 6 months, adolescents do more poorly than that.

[Slide.]

Looking at self-report of missing 2 or more pills in the last 3 months, this is one study that suggested that adolescents did do more poorly in taking pills consistently. So among those who

reported missing 2 or more pills in a given month, adolescents were more likely to do so, 25 percent essentially.

[Slide.]

I think these are some of the things that contribute to our rates of adolescent pregnancy that are head and shoulders greater than in other countries. There clearly are many other factors that contribute to it, but inconsistent use of the pill is one of them.

[Slide.]

This is looking at a very recently published study out of Indianapolis looking at pill-taking behaviors as well as condom-use behaviors, and requiring, essentially acknowledging, that an individual who is taking the pill needs to consider one's birth control method to be the pill, number one, and then to take it consistently and correctly.

This study found that many young women are at risk particularly in transitions on again, off again with the pill, which happens quite frequently

among young women. And so this particular study found that that happened and happened often enough that it likely impacted the risk of pregnancy.

[Slide.]

They categorized patterns of use of adolescents as either stable over the course of a 3-month interval, or starting the pill, or stopping the pill over that interval, and individuals go in and out of pill use. So that is important to capture, and this was again in a study. But this is how my patients use the pill. They go on again and off again quite frequently.

In this particular study, episodes of 3 or more missed pills happened about twice over a 3-month interval, and you can see how that would impact the effectiveness of the pill.

[Slide.]

In another study looking at women who failed to come back to clinic at 3 months, the women who didn't come back, in this particular study in the second bullet, had all of them discontinued pill, and two-thirds of those were

continuing to be sexually active.

Among those who discontinued use over that interval of that time, they missed an average of 3 pills per month. Among those who considered themselves to be continuing users of the pill, if you asked them at the end of that 3-month period "Are you a pill user," they would say yes. Those individuals also missed about 3 pills a month, and that is in a 1-month interval. So it gives us pause in terms of effectiveness.

[Slide.]

Lots of things that I hear about the pill every day from my patients; they are concerned about rates of bleeding and irregular bleeding. The suburban teens that I see are almost universally concerned about the possibility of weight gain, so that is something that I need to address upfront in seeing the patients. But it is a question that I get quite frequently.

There still are lots of myths out there; the pill makes you sick--incidents of nausea and vomiting are relatively low and can be a problem

for some women, but the idea that many people get sick in taking the pill.

The idea that the pill makes you infertile is one that I still continue to hear. I heard it just yesterday. So, it is something that is still out there. Clearly, we all know that that is not the case, but our patients don't know that that is not the case.

Something I hear from my patients' mothers more often than from my adolescent patients is that the pill causes cancer and the mothers at least are concerned about that possibility. And recognizing that those mothers have influence on their daughters in terms of consistency of use is also important, and that boyfriends and others influence the use--girlfriends, as well--influence the use of the pill.

[Slide.]

On again, off again; the individual who is in a relationship effectively contracepting with an oral contraceptive, breaks up with her boyfriend, decides she is never again going to be sexually

active, so what does she need the pill for.

She stops the pill, and lo and behold, what happens? We, as adults, it is pretty predictable what is going to happen. She is either going to get back with that boyfriend or she is going to be in another relationship in which she may choose to be sexually active, needs contraception.

She may have heard us say, wait for your next period to start your pill, and she is waiting and she is waiting, and she may wait 9 months for that next period to come after she delivers her baby.

So, this on-again, off-again, use is something that is very common among adolescents and we need to be concerned about it.

[Slide.]

What is she going to do after she goes off the pill? If we are lucky, she will use another method of contraception. Chances are it is going to be somewhat less effective than birth control pills so that may increase her risk of pregnancy.

On the other hand, she may decide that she is going to be abstinent, whatever that means in her mind, which may mean a whole variety of other sexual behaviors that might put her at risk for STIs. But that is an alternative. On the other hand, many women do continue to be sexually active.

[Slide.]

Just very briefly, to cite a study that looked at continuing users, that individuals who had reduction in their dysmenorrhea, who got a benefit of the pill that was very noticeable to them on a monthly basis, were more likely to be ongoing users.

So, this is something that I use in my clinical practice. And what it points out to us and to the FDA, as we think about it, I think this brings up the importance of those patient-reported objective findings and our being able to assess that sort of thing, that these are the reasons that individuals may stay on the pill on an ongoing basis, relief of dysmenorrhea.

If you talk about acne to an adolescent,

that the pill will improve their acne. That is another powerful reason for adolescents to stay on the pill and perhaps stay on the pill between those relationships if she has those other benefits and recognizes those other benefits, as well.

[Slide.]

And then, finally, to just point out, as has been pointed out by the previous speaker, we have talked about patient-related issues.

There are provider issues that are barriers and sometimes clinicians and providers don't always have knowledge about--for example, formulary and coverage, those sorts of things. But there are lots of things in the healthcare system that impact our patients' abilities to be successful in using contraceptives.

The formulary issues, the issues of whether or not she has any health coverage, whether her health insurance actually covers contraception is important; how much did the pill cost; what are our office hours; can she get a refill of her pills when she needs it; does she have to get a new--does

she have to go to the pharmacy once a month or can she get multiple pill packs.

[Slide.]

This is a recently published study from California. Providing more than one pill pack was beneficial to individuals in continuing use of the pills. So, providing a full year's worth of oral contraceptives was helpful in helping women to continue to use their method of contraception.

These are the sorts of things that clinicians need to be aware of, and certainly our healthcare system impacts.

[Slide.]

So, overall, this is the real world that I am looking at, and thank you all.

DR. LOCKWOOD: Thank you.

We are going to, first of all, take any questions on the presentations. We are going to address three questions on translation of clinical findings in the real world, and we will finish with Dr. Trussell's presentation on cycle control, if he will be back by then, and then we will discuss

cycle control tomorrow where we have a little bit more latitude in time.

So, questions about the presentation?

[No response.]

DR. LOCKWOOD: Crystal-clear? I think some of your points will lead right into the first question, which is Question 17: Can trial design be modified so as to provide results that are more reflective of actual effectiveness in the real world?

DR. BERENSON: Is that first word supposed to be "Should"?

DR. LOCKWOOD: Well, you can read it any way you like. It looks like "Can" to me, but "Should" may be equally appropriate. I think "Should" we answered this morning, though. I think now the question is the nuts and bolts of what practical things can we recommend.

I mean, I think we covered some of these in the morning, expanding the age of entry, in fact, not having any specific cutoffs certainly in the younger group. We talked about expanding

indications with BMI and not limiting it to a specific class.

I think I was not entirely comfortable with expanding it into women with a family history of venous thrombotic events, but certainly, you know, if a sponsor believes that their agent is particularly safe in that context, that might be a reasonable expansion of an indication, as well.

Other thoughts? Paula.

DR. HILLARD: One thing that I would like to say is that if one looks at adolescents and their pill-taking behaviors, one conclusion that I would hope would not be the case from my presentation is that, because adolescents may not do as well in taking the pill consistently and correctly, then that might argue not to include adolescents in clinical trials.

I would certainly say that adolescents are using oral contraceptives to try to prevent pregnancy. And so, if one wants to look at and more closely approximate the real world, I think it is important to include adolescents, as well.

DR. LOCKWOOD: Dr. Trussell.

DR. TRUSSELL: I thought several of Melissa's suggestions were really quite good--I, mean making sure that the clinical trial site can be accessed on nights and weekends when otherwise people who work can't get there.

I mean, it is doing all the things that you would do in family-planning clinics to try to better serve clients, you know, making the locations near bus routes or other public transportation. I mean, there are just a whole long litany of these things that have been suggested for clinics and they would follow directly over to clinical trials.

DR. LOCKWOOD: I think that a number of Melissa's points are going to be very important in trying to expand subpopulations, particularly being able to reach into the Latina and African-American community. I think getting community leaders to support clinical trials, particularly in New Haven, Connecticut, is a very important thing to try to do.

Other points? Dr. Johnson.

DR. JOHNSON: I think that most of the members of the Committee agree that it would be much better to have a real-world group involved in these studies. But are there any potential barriers for the younger age group? Is there going to be any opposition to that in the United States?

Then with the BMI, higher BMI, I absolutely agree that we should include women with more than a BMI of 35. But should we put an upper limit there? Is that going to be a problem in terms of any concerns if we let the BMI be as high as it could possibly be?

So, are there any limitations or any barriers to making it more of a real-world study?

DR. LOCKWOOD: Well, should there be, I think the answer is no. Are there, probably the answer might be politically yes. But, in fact, I don't believe that there are any specific legislative or regulatory constraints on any of the things we have talked about. Is that correct?

DR. SCOTT: We have to get consent for

under 18.

DR. BUSTILLO: Yes; what about informed consent for the adolescent? Is that a problem?

DR. LOCKWOOD: Not presumably. It is needed, but that shouldn't be a hurdle.

DR. BERENSON: That is a barrier because many times--well, let me rephrase it. In certain states, the parents are not required to accompany the adolescent to the clinic and so those adolescents obviously would not be eligible. That would not really be a problem except you do get into an issue of generalizability when you leave those adolescents out of your trials.

DR. HILLARD: Can I add to that in terms of informed consent? That can be addressed by individual institutions' IRB in terms of consent, that there can be exceptions in situations where adolescents are allowed otherwise to consent to getting contraception, that they may consent, as well, to participate in clinical trials. But that is something that needs to be addressed on an institution-by-institution basis.

DR. ESPEY: And many institutions, the IRB will not allow you not to get a consent from the adolescent even if the adolescent can get contraceptive without parental consent.

DR. HILLARD: There are some guidelines from the Society for Adolescent Medicine that address that and changing the rules from the IRB. I agree. It's not easy.

DR. LOCKWOOD: Dr. Tobert.

DR. TOBERT: I think it is now, if the FDA accepts the panel's unanimous decision about active controls, that it would be much--the incentive that companies hitherto had to drive that Pearl Index down as low as they could will go away and, therefore, it will be much easier to have trials that do reflect more the kind of real-world issues that the two excellent speakers we just heard referred to.

DR. LOCKWOOD: Abbey.

DR. BERENSON: The reason I asked at the beginning if the question should be "Should" on Question 17 is because we can just answer that yes

and move on; of course, you could redesign the studies that are more reflective of actual effectiveness.

In fact, if you wanted to get very stringent about it, you could require them to break down the groups into subgroups that reflect certain characteristics in the U.S. population, but then you are getting into very large studies when you start to have subgroups.

DR. LOCKWOOD: And expensive.

DR. BERENSON: And very expensive.

On the adolescent issue, one point for consideration is that adolescents are physiologically very similar to adults since we do not give birth control to anyone that has not gone through menarche so it is rare that you have anyone on contraception that is younger than 12. And many medications that the FDA approves are not tested in children and are given to children less than 12.

So, if they aren't that different, according to Paula's slides, from the adult population in behavior, then would that need to be

required if there are other issues that make it difficult to include them in the trials?

DR. LOCKWOOD: So what is the lower limit of age, is that what you are--

DR. BERENSON: I don't think I have ever had an adolescent less than 12 on oral contraceptives.

DR. LOCKWOOD: I don't think I want to go there.

Dr. Peterson.

DR. PETERSON: I think it is straightforward that you want to have the study population selected from the population that you want to generalize the findings to and that there are ways to do that. So, I think, as Abbey said, 17 and 18, the answer is yes.

Something that I think is implicit in the comments that have been made that may be helpful to be explicit about is when we were looking at BMI and smoking and family history, is that,even if you include those in the trial design, is that you will not be able to answer the important question that

is almost implicit in trying to put them in there and that is that does the safety or effectiveness for those subpopulations differ.

That is going to be, as has been mentioned, a post-marketing--presumably a post-marketing assessment.

DR. LOCKWOOD: Right. Unless, of course, there is such dramatic differences that they would appear but the power won't be there.

DR. PETERSON: But you would have to plan for those and power it accordingly.

DR. BERENSON: Clarification before there is a riot in the room. Patients that get oral contraceptives between 12 and 15 are usually getting them for cycle control or for acne. So, while you can include them in these trials, you are not going to get your data that you need on efficacy.

DR. LOCKWOOD: I think that one of the entry criteria would be that they--one of the entry criteria, I think that is universal is that they would be at risk for pregnancy. So they would--

DR. BERENSON: And you are never going to get enough numbers to address the question.

DR. LOCKWOOD: They wouldn't be candidates. I think that they--we will leave it at that.

DR. LOCKWOOD: Dr. Trussell.

DR. TRUSSELL: In particular, with respect to BMI, one advantage of having more people in the trial is that, over time, the FDA can pool the data from several trials to see if there is any indication of an effect of weight on OC efficacy.

DR. LOCKWOOD: Very good point.

Question 19 was: Should clinical trials investigate new technologies that may facilitate compliance in real-world use? I would like to spend the last couple minutes talking about that.

Specifically, I think we addressed some of the issues in the context of using electronic devices, web-based diaries, and so forth. I think that Paula raised the issue with pill kits that record the time of the removal of the pill at least.

I think one of the dilemmas there is that the better the technology is in recording data, the more likely it is probably to providing clues for compliance or adherence, and that is a bit of a conundrum.

Any comments about that?

DR. BLUMENTHAL: The way the question is phrased again, if we are dealing with clinical trials that are designed to assess the efficacy of a new contraceptive, then it gets risky to me to also incorporate, or nested in there, an investigation of a new technology.

Maybe we should say we should, a priori, investigate technologies that may facilitate compliance, and once validated, these technologies should be incorporated or can be recommended to be incorporated into trials. But first validate the technologies and then incorporate them. Don't run sort of two nested studies while you are doing an efficacy trial.

DR. TRUSSELL: And a corollary there would be I don't think it would be at all helpful to

include technologies that may work, like directly observed therapy, in a clinical trial if it is not going to be used in the real world because then you are just going to get an inflated view of efficacy.

DR. LOCKWOOD: We are going to hold off the discussion on cycle control until tomorrow. It will give us something to think about as we rest our heads on our pillows tonight. I think we have covered the translation of clinical findings to real world.

Topic 4 - Cycle Control

[Slide.]

DR. TRUSSELL: I am reporting today on a pair of papers that was published in the January issue of Contraception that reflects the recommendations of a consensus group that met to recommend standardized data collection and analysis procedures for bleeding.

That group included all the people listed here, but particularly Anita Nelson, who is in the back of the room.

[Slide.]

We all know that decreases in doses of estrogen and progestin have occurred since the pill was first introduced, which has resulted in increased incidence of unscheduled bleeding and spotting. Diverse approaches have been used to assess cycle control in 12 clinical studies that we identified going back in time. And standardization of methods for collecting and analyzing such date are long, long overdue in our view.

[Slide.]

There have been diverse approaches that have been used in these 12 studies that we analyzed. All of them required subjects to keep a daily diary of bleeding and spotting but, in most cases, the diary content was not described and sample pages were not provided.

Little information regarding data collection or patient instruction for completing the diaries was available. They were mostly paper diaries collected every three months, which means that they probably were filled out every three months and no information about the validation of

methods at all.

[Slide.]

Interactive Voice Response was utilized to confirm contraceptive method adherence in one study of the NuvaRing, and active inquiry regarding the incidence of bleeding and spotting and daily electronic data capture with time and date stamping of all entries was utilized in one study, that of Seasonale.

[Slide.]

Now, most of these studies utilized the WHO Belsey criteria which were developed quite a long time ago and are as follows; that vaginal blood loss requiring sanitary protection is classified as bleeding, and vaginal blood loss not requiring sanitary protection is classified as spotting.

Now, there are some exceptions in the studies we analyzed. Several studies classified bleeding as requiring more than one pad or tampon, and one study asked women to classify bleeding as light, normal, or heavy.

[Slide.]

In the meanwhile, current common use of mini-pads and pantyliners further clouds the interpretation of bleeding and spotting data. This was not an issue in the earlier research, but it certainly is an issue that must be addressed with rules about what to classify today.

None of these studies addressed the impact of these products when collecting data in bleeding diaries.

[Slide.]

The criteria for inclusion of a cycle in the analysis of bleeding and spotting is rarely described and varies significantly among products. So, for example, if a contraceptive method was not used for three or more consecutive days, trials of one OC and a contraceptive vaginal ring excluded those cycles from bleeding and spotting analysis. Most studies did not specify the number of cycles excluded from bleeding analysis or delineate the reasons of why they were excluded.

[Slide.]

Inconsistent criteria were used to calculate rates of unscheduled bleeding or spotting. Bleeding or spotting that occurred during the last week of active hormone administration within a cycle was not always counted as unscheduled, but instead as "early withdrawal bleeding." Bleeding that was reported on days 1 to 4 of active hormone administration was not consistently considered unscheduled, but instead as scheduled.

[Slide.]

So, it is possible for only bleeding reported on days 5 to 17 of the 21-day active pill cycle to be defined as unscheduled. And bleeding that occurred in the other eight days of the cycle may be excluded from calculations which allows, of course, for significant underreporting compared to analyses that don't use that method.

[Slide.]

Cycle control or bleeding profile of hormonal contraceptives usually is presented as an incidence, but the definition of incidence varied

from the proportion within a population, the incidence within a specified time frame ranging from a single cycle to a year, the percentage of patients achieving an intended bleeding pattern.

So, you get the idea here there is a lot of variability in how these are done.

[Slide.]

Amenorrhea was variably defined as the absence of withdrawal bleeding, two consecutive cycles without bleeding or spotting, or no bleeding or spotting throughout a 90-day reference period.

[Slide.]

So, one medical reviewer plaintively stated in his review that, based upon the same raw data, the percent of cycles in which unscheduled bleeding or spotting occurred in patients who had taken the product ranged from 19 to 29 percent in the first cycle and 13 to 19 percent in later cycles when evaluated using varying definitions employed in prior regulatory reviews of other combined oral contraceptive products.

So, this is clearly a problem that the

medical reviewers have noted.

[Slide.]

Now, in addition, they had problems with the Belsey criteria themselves. They are not particularly useful for the reporting of cyclic bleeding in women using combined hormonal contraceptives without appropriate modification.

They recommend use of a predefined reference period, most commonly 90 days, but they don't differentiate bleeding occurring during active hormone therapy from that occurring during the placebo interval and, therefore, they cannot identify unscheduled bleeding.

[Slide.]

Regardless of the formulation, method of delivery, or cycle length, unscheduled bleeding and spotting episodes are more frequent in women who do not use the contraceptive method consistently, in first time users compared with long-term users and during initial cycles of use.

So, from our review of the literature, we concluded those three things, and that is all.

[Slide.]

Beyond these findings, data from existing studies are not adequately consistent to permit meaningful comparisons of unscheduled bleeding or spotting or to provide clinicians useful information to guide their practices.

[Slide.]

So, we set about making recommendations for study design. There is a whole long list of them, and I will just run through them. A minimum duration of six months for studies of cyclic hormonal contraceptives and a longer duration for studies of extended regimens. Duration of the reference period for cycle control should correspond to the longest cycle evaluated in the study.

In a controlled comparison of 28-day regimens, the reference period should be 28 days. In studies that include an extended regimen, the reference period should be as long as the complete cycle, for example, 49 days or 91 days or 364 days.

[Slide.]

The bleeding analysis should include all women eligible for combined hormonal contraceptives without restriction to body weight--so again the same recommendation that we just made about getting into an efficacy trial--but subjects at risk for untreated Chlamydia should be screened because chlamydial cervicitis often causes abnormal bleeding and spotting.

[Slide.]

Now, as for terminology, we suggest the following:

Bleeding is evidence of blood loss that requires the use of a tampon, pad, or pantyliner.

Spotting is evidence of blood loss not requiring new use of sanitary protection including pantyliners.

And an episode of bleeding or spotting is bleeding or spotting days bounded on either end by two days of no bleeding or spotting.

[Slide.]

We recommend abandonment of the use of "period" or "menses" with regard to combined

hormonal contraceptive use and replace it with "scheduled" or "withdrawal" bleeding. Any bleeding or spotting that occurs during the hormone-free intervals, regardless of the duration of the regimen, should be counted as bleeding and it may continue into days 1 to 4 of the subsequent cycle. The term "scheduled bleeding" emphasizes that withdrawal bleeding is not the same as menstruation at all.

[Slide.]

Abandon the use of "breakthrough" bleeding or spotting and replace with "unscheduled" bleeding or spotting. It is any bleeding or spotting that occurs while taking the active hormones with two exceptions; bleeding or spotting that begins during the hormone-free interval and continues to days 1 to 4 of the next active cycle not considered "unscheduled and bleeding or spotting that is reported on days 1 to 7 of the first cycle of any study medication not be considered as "unscheduled."

[Slide.]

Abandon the use of the term "amenorrhea" and replace with "absence of all bleeding and spotting."

[Slide.]

Regarding data collection, we suggest asking subjects to document use of combined hormonal contraceptives and incidents of bleeding and spotting in a consistent manner every 24 hours, and to encourage recording data at the same time within each 24-hour period.

[Slide.]

We recommend daily real-time electronic collection. It could be a daily phone call initiated by the woman to just call in and report, electronic diaries, text messaging, other validated systems. None of these systems have been validated, I must add. So, they first need to be validated.

And prospective comparative studies are needed to assess the accuracy of electronic data collection versus traditional paper diaries. We did feel that they are likely to do better, but we

admit that we do not know.

[Slide.]

What about data analysis? We recommend presenting observed bleeding patterns within a reference period as total days of bleeding, unscheduled days, scheduled days, and, for bleeding and spotting, bleeding only and spotting only. Present the incidence, the percentage of subjects, in the absence of bleeding or spotting.

[Slide.]

Structure the trials to allow analysis of cycle control stratified according to body-mass weight index, weight, age, parity, smoking, hormonal contraceptive-use history, untreated Chlamydia infection.

[Slide.]

To evaluate bleeding patterns over time, it is important to analyze data from subjects who complete the trial because what happens in many trials is you get, say, the proportion of women with breakthrough bleeding in Cycle 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, and it goes down. And you

say, aha, it goes down with time. But it may not go down with time. It may go down because people exited the trial due to a poor bleeding profile. So, if you are going to examine the question of whether it actually goes down over time, you can look at it only among people who use it for a pretty good while.

[Slide.]

And then, finally, we recommend that you analyze the incidence of unscheduled bleeding and spotting on a daily basis and present it in this graphical form contributed by Carolyn Westhoff.

That is the end of my presentation and we can go home 40 minutes early.

DR. LOCKWOOD: Thank you.

We will allow some questions about the presentation, but we won't get into the questions on cycle control.

DR. SCOTT: Do that again, Charlie. What does it mean?

DR. LOCKWOOD: We will allow questions specifically to this presentation, but we will save

the FDA's questions for tomorrow.

DR. SCOTT: You mentioned several times that the survey instruments haven't been validated. How would you actually validate it? Would you have to observe the bleeding and spotting?

I mean, when you say it hasn't been validated, like a lot of instruments that are used--for example, databases-- you can at least look at the charts and review the charts and see that they agree, but how would you validate these?

DR. TRUSSELL: I don't know that we had an idea for how to validate them except--the idea here is that you don't want to fool yourself into thinking you are collecting something when, in fact, it is just random reports. So I don't know how to validate it.

DR. LOCKWOOD: Abbey.

DR. BERENSON: Actually, there is a study that did that. I believe it was in England, and they required the women to bring in the pads.

DR. JOHNSON: I was going to say the same thing, and I know that currently there is a

contraceptive that is using that for confirmation that it prevents menorrhagia, but basically bringing in all the pads, all the tampons. And there are other ways of measuring blood loss, as well, that can be done with pads. So, that would be the only truly scientific way of proving the effectiveness and that could be something that someone--perhaps they are willing to undergo that, but, really, it is testing the diary and seeing what is most effective and I am not sure any company would be willing to do that.

DR. STADEL: In the context of the discussion earlier today, I wonder if you could comment now. or, if not, perhaps tomorrow, on what you see as sort of the sample size needs in a comparative trial for these more common events of bleeding and perhaps other discontinue--what sort of numbers would you need to get a good comparison between two products in a controlled trial?

DR. TRUSSELL: Well, I mean our idea primarily was that this would be embedded in a clinical trial that you are already going to do for

efficacy, so you are going to have much more power because these are much more frequent events.

And we weren't recommending that you do separate trials, but to do it all at once. If you are already going to be collecting data in the efficacy part on whether they use the product and stuff like that, then you can also collect the bleeding data.

DR. WESTNEY: I had a question regarding--with respect to bleeding, whether there is a need for quantification of bleeding. I mean, just extrapolating from the urinary incontinence data, it is clear that the factors that drive protection usage are variable from person to person, and also can be economically driven, whether they are at home--you know, there are any number of factors. So the end result is that we have to do 24-hour pad tests to validate any anti-incontinence therapy whether it is medication or drugs. So, if the quantity is important, then you are better off weighing it.

DR. TRUSSELL: I don't believe that we

thought that you needed to go that far in every clinical trial, but we did think and reject the notion that, in one of these studies where women had been asked whether their bleeding was none, light, medium, or heavy, it didn't really mean anything and so that didn't seem like a very good idea.

DR. WATKINS: Dr. Gillen.

DR. GILLEN: I just had a quick question about the longitudinal analysis of bleeding patterns over time where you suggest removing patients that prematurely stop the study. Certainly, you could bias your results one way or the other; right?

So you have talked about the concept of data being missing at random where you are allowing patients the trajectory out and maybe that is not representative of what would have happened had they stayed in the trial. But the same could be true, as well; right? By removing those patients you could be eliminating a portion of your population that is either--in some sense has different

trajectories from every else or some random variability; right?

DR. TRUSSELL: Well, you can do it both ways. But I would count as convincing evidence that unscheduled bleeding goes down over time as a result of the body getting used to the hormone. That is the story that we all hear. That is what everybody tells all their patients. It is convincing only if you are looking at women in whom it went down over time, and those have got to be the same women.

DR. GILLEN: But if your data is missing at random, then likely their base methods could still pick that up; right? I mean, their past trajectory should indicate something about their future trajectory if it is--

DR. TRUSSELL: I doubt that it is missing at random. I think that--but I don't know. But just because, in the sample of women who happen to be using--in Cycle 1, unscheduled bleeding was 10 percent and it goes down to 5 percent in Cycle 3, doesn't necessarily mean that it goes down for

individual women over time.

DR. GILLEN: No; I absolutely agree with that. I am just saying that if you are comparing on the aggregate level, and you are using something that is likely basically feeding off the trajectory of the information you have had on individuals in the past, unless you have non-informative missingness, it is likely that estimates should be consistent.

DR. TRUSSELL: And we suggested doing it both ways.

DR. JOHNSON: Just to ask one more question related to that. I am obviously not a statistician, but if you delete the women who don't complete the study, then aren't you taking the group who had the least bleeding; therefore, they stay in the study as the ones that showed the changes over time?

You would think those would be the people who would have the least bleeding from the very start and therefore not the real world.

DR. TRUSSELL: That is precisely what you

are looking at. And if you do not look at it that way, then you cannot tell whether it is actually going down over time or not. There are people who had spotting and unscheduled bleeding in Cycle 1, who did not have it in Cycle 3 or 4.

DR. JOHNSON: But aren't you selecting a certain population that has less bleeding by deleting the women who had bleeding at the beginning and then stopped?

DR. TRUSSELL: I think that you look at it both ways. But you cannot answer the question about whether there is a physiological decrease in unscheduled bleeding without looking at the same people.

DR. LOCKWOOD: I think there are sort of two questions. One is what is the rate of attrition because of the bleeding and so that is likely to occur early. Certainly with long-term progestin-only contraceptives, it tends to occur early. And then you sort of are left with the residual patients who weren't disturbed by their bleeding or their bleeding wasn't that bad, and

they persist with the agent.

So, you may select out the worst bleeders, if you will, initially, and the ones that are happy and content with the side effects are the ones that tend to persist.

So, I think you do need to look at it both ways.

DR. TRUSSELL: That is why we said look at it both ways. On the other hand, if the heavy bleeders are selected out early, what are you willing to assume about what would have happened to them had they stayed in the trial? If you do not look at people who actually stayed in the trial, then you have no information about what went--

DR. LOCKWOOD: Then you would have to design such a trial, that you paid them enough to continue for a year and observed the amount of bleeding they did.

I think that, in general, though, when you look at their endometria, they progressively became more atrophic and there is literally less surface area to bleed, and so the bleeding keeps dropping.

DR. TRUSSELL: Yes.

DR. LOCKWOOD: Thank you all very much. See you tomorrow morning.

(Meeting recessed at 5:30 p.m., to reconvene Wednesday, January 24, 2007.)

PAPER MILL REPORTING
Email: atoigo1@verizon.net
(301) 495-5831