at

AT

DEPARTMENT OF HEALTH AND HUMAN SERVICES

FOOD AND DRUG ADMINISTRATION

CENTER FOR DRUG EVALUATION AND RESEARCH

ADVISORY COMMITTEE FOR REPRODUCTIVE HEALTH DRUGS

Monday, September 29, 2003

8:35 a.m.

Hilton Hotel
The Ballrooms
620 Perry Parkway
Gaithersburg, Maryland

MILLER REPORTING COMPANY, INC.
735 8th Street, S.E.
Washington, D.C. 20003-2802
(202) 546-6666

at

PARTICIPANTS

Linda C, Giudice, M.D., Ph.D., Chairman
Shalini Jain, PA-C, M.B.A., Executive Secretary

MEMBERS

Susan Crockett, M.D.
W. David Hager, M.D.
Nancy W. Dickey, M.D.
George A. Macones, M.D.
Joseph B. Stanford, M.D.
Scott S. Emerson, M.D., Ph.D.
Vivian Lewis, M.D.
Larry Lipshultz, M.D.
Valerie Montgomery Rice, M.D.

SPECIAL GOVERNMENT EMPLOYEE VOTING CONSULTANTS

Robert G. Brzyski, M.D., Ph.D.
Adelina M. Emmi, M.D.
David L. Keefe, M.D.
Lawrence C. Layman, M.D.
James H. Liu, M.D.
James P. Toner, Ph.D.

ACTING CONSUMER REPRESENTATIVE

Lorraine J. Tulman, D.NSc.

FDA

Daniel Shames, M.D.
Shelley R. Slaughter, M.D., Ph.D.
Audrey Gassman, M.D.

at

C O N T E N T S

P R O C E E D I N G S

Call to Order and Opening Remarks

DR. GIUDICE: Good morning. I would like to begin our meeting this morning. This is the FDA Advisory Committee for Reproductive Health Drugs. I am Linda Giudice from Stanford University and I am the Chair of the Committee. Today, we will have a general discussion and tomorrow we will have a product-specific discussion with Sorono.

Some housekeeping issues before we have introductions of the committee members. Please, if you would turn your cell phones and beepers off or at least put them on "Silent" so that the proceedings are not disturbed. Rest rooms are down towards the main desk.

So I would like to begin with introduction of the committee members and perhaps we can start on this side with Dr. Hager.

DR. HAGER: David Hager, University of Kentucky.

DR. CROCKETT: I am Susan Crockett and I am from Christus Santa Rosa in San Antonio, Texas.

DR. MACONES: George Macones from the University of Pennsylvania in Philadelphia.

DR. LEWIS: Vivian Lewis from the University of Rochester.

DR. TULMAN: Lorraine Tulman, University of Pennsylvania, Consumer Representative.

DR. LIPSHULTZ: I am Larry Lipshultz from Baylor College of Medicine in Houston.

DR. KEEFE: David Keefe from Women and Infants Hospital and Brown in Providence, Rhode Island.

DR. DICKEY: Nancy Dickey from Texas A&M Health Science Center in College Station.

MS. JAIN: Shalini Jain, Executive Secretary to the Advisory Committee for Reproductive Health Drugs.

DR. EMERSON: Scott Emerson from the University of Washington in Seattle.

DR. EMMI: Adelina Emmi from Medical College of Georgia.

DR. STANFORD: Joseph Stanford from the University of Utah, Salt Lake City.

DR. BRZYSKI: Robert Brzyski from U.T. Health Science Center, San Antonio.

DR. TONER: Jim Toner, Atlanta Center For Reproductive Medicine.

DR. RICE: Valerie Montgomery Rice,

Meharry Medical College.

DR. GASSMAN: Audrey Gassman from the FDA.

DR. SLAUGHTER: Shelley Slaughter from the FDA.

DR. SHAMES: Dan Shames, FDA.

DR. GIUDICE: Thank you very much. Dr. Layman just walked in.

DR. LAYMAN: Hi. Larry Layman, Medical College of Georgia.

DR. GIUDICE: Thank you.

I would like to introduce now Shalini Jain who will talk about the conflict-of-interest statement.

Conflict of Interest Statement

MS. JAIN: Thank you, everyone, for participating today. I would like to read the conflict-of-interest statement for the Advisory Committee for Reproductive Health Drugs for September 29, 2003.

"The following announcement addresses the issue of conflict of interest with respect to this meeting and is made a part of the record to preclude even the appearance of such at this meeting. The committee will discuss issues relevant to the conduct of clinical trials and

outcome measures for consideration of approval of drug products for the indications of induction of ovulation and pregnancy in anovulatory infertile women and development of multiple follicles, and pregnancy and ovulation women participating in assisted reproductive technology, or ART, programs.

The topic of today's meeting is an issue of broad applicability. Unlike issues before a committee in which a particular product is discussed, issues of broader applicability involve many industry sponsors and academic institutions.

All special government employees have been screened for their financial interests as they may apply to the general topics at hand. Because they have reported interests in pharmaceutical companies, the Food and Drug Administration has granted general-matters waivers to the following SGEs which permits them to participate in today's discussions; Dr. Scott Emerson, Dr. W. David Hager, Dr. Larry Lipshultz, Dr. Valerie Montgomery Rice, Dr. Susan Crockett, Dr. Adelina Marie Emmi and Dr. James Liu.

A copy of the waiver statements may be obtained by submitting a written request to the agency's Freedom of Information Office, Room 12A-30

of the Parklawn Building. Because general topics impact so many institutions, it is not prudent to recite all potential conflicts of interest as they apply to each member and consultant.

FDA acknowledges that there may be potential conflicts of interest but, because of the general nature of the discussion before the committee, these potential conflicts are mitigated. In the event that 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.

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

Thank you.

DR. GIUDICE: Thank you.

I would now like to introduce our first speaker, Dr. Daniel Shames, who is the Director of the Division of Reproductive and Urologic Drug Products at the FDA.

Opening Remarks

DR. SHAMES: Thank you. I would first like to welcome everybody this morning and I would like thank Dr. Giudice and all our advisors and consultants and speakers for taking time out of their busy schedules to educate us about the issues surrounding the use of drugs for the treatment of female infertility.

I would also like to take this opportunity to thank the two members of our division that were most responsible for assembling and producing the elements of what I believe will be a very exciting forum. Both are fully trained and certified reproductive endocrinologists and the division is fortunate to have them. Thank you to Dr. Shelly Slaughter and Audrey Gassman.

In our division, we regulate a diverse group of drug products for such indications as advanced prostate cancer, sexual dysfunction, post-menopausal therapy, incontinence, BPH, among others. I must say that, among all the indications that we deal with, I find female infertility the most challenging from a clinical-trial-design, scientific and regulatory perspective.

As you know, FDA approves for marketing

safe and effective drugs. Safety and effectiveness are demonstrated by adequate and well-controlled clinical investigations. Science and drug development have considerably advanced using randomized, blinded, controlled-trial design. The randomized clinical trial has revolutionized our view of what are effective treatments for many diseases such as cancer therapy and heart disease.

For female infertility, we want to make sure we provide the public with drugs that are safe and effective. This is accomplished by the provision of evidence from properly designed and conducted trials. The complex nature of clinical treatment protocols and the rapidly changing technologies and pharmacotherapy makes it a challenge to establish standards for insuring that clinical trials of drugs for female infertility are appropriate.

There is an overriding charge for the assembled experts attending today and tomorrow. This charge is to inform the division regarding elements that must be incorporated into clinical trials for pharmacologic therapy of infertile women so that these trials will provide the level of evidence needed to conclude that these therapies

are, indeed, safe and effective.

With this information and input from all appropriate sources including the pharmaceutical industry, the division will write a guidance for the clinical evaluation of drugs in this area resulting in more rapid and efficient development of pharmacologic therapy for female infertility.

Thank you.

DR. GIUDICE: Thank you very much.

We now have a guest speaker, Dr. David Keefe, who will talk to us--who is the Director of the Reproductive Medicine and Infertility Unit at Women's and Infant's Hospital in Rhode Island. He will talk to us on Ovulation Induction and Assisted Reproductive Technology, Background and State of the Art.

Dr. Keefe.

Ovulation and Assisted Reproductive Technology

Background and State of Practice

DR. KEEFE: Great.

[Slide.]

Thank you, Dr. Giudice and thank you for the opportunity to come here today and share in this very important forum.

[Slide.]

There are a number of aspects to my presentation. I am going to give an overall introduction to the ART procedures and, in particular, in vitro fertilization and comment in particular how they relate to the issues of study design, particularly the study population, study design and then, hopefully, give you a sense of what I think is the future of ART, particularly the issue of ART outcomes.

[Slide.]

Assisted Reproductive Technologies actually encompass a variety of different techniques. The one most commonly practiced in the United States is in vitro fertilization with embryo transfer. It is a little bit of a misnomer to call it in vitro fertilization because probably the most crucial part is the actual implantation which follows. But convention calls it in vitro fertilization.

In addition, gamete intrafallopian transfer is a variation on this theme in which the gametes are placed in the fallopian tubes or placed after the fertilization has actually occurred, zygote intrafallopian transfer or tubal-embryo transfer.

These three are almost of historic interest because they are not widely practiced. There are rare indications and I am not going to discuss controlled ovarian hyperstimulation with intrauterine inseminations which some would put under the rubric of assisted reproductive technologies. Rather I am going to focus most of the discussion this morning on in vitro fertilization, the most widely practiced of the assisted reproductive technologies in the United States.

[Slide.]

The first step is to downregulate the ovaries, typically done with oral contraceptive pills or progesterone with gonadotropin-releasing hormone agonists or antagonists, to first try to synchronize the follicular cohort since the idea is to superovulate the woman. The controlled ovarian hyperstimulation step is typically done with gonadotropins and monitored with a number of ultrasound and estradiol levels.

Next, after the ultrasounds and estradiols have identified a size of follicle and an estradiol level consistent with oocyte maturation within the follicle, then hCG human chorionic gonadotropin is

used to trigger maturation largely because hCG serves as a kind of a surrogate for luteinizing hormone, the physiologic trigger of oocyte maturation. Just practically hCG has a much longer half-life and is more practical, at least under current technologies.

Retrieval is done under general sedation and then fertilization is typically performed transvaginally. Fertilization is effected either with an incubation with the gametes in the test tube or through direct injection of the sperm through intracytoplasmic sperm injection.

Next, the embryos are cultured for a number of days, between two and six days, depending on protocols, particularly under the indications of the patient. The embryos, then, are transferred through a very non-invasive procedure in which a very small flexible catheter is used to place the embryos up into the uterus, typically without but occasionally with a hatching procedure in which the shell, the zona pellucida, around the embryos is first thinned or breached.

Next, and finally, the luteal phase, that second half of the menstrual cycle, is supplemented since the retrieval had aspirated a number of the

follicular cells and the gonadotropin-releasing-hormone agonists or antagonists may have shut down their ability to produce progesterone adequately initially. So progesterone is administered either vaginally or intramuscularly for a number of days.

[Slide.]

So just an overview here. You have the follicle stimulation. I think this gives you a sense of the time frame. This is a very drawn-out procedure which may go up to six weeks for an individual single cycle in which these follicles are stimulated with gonadotropins after downregulation and then the mature eggs are removed with a 15 or 20-minute aspiration under general sedation.

Then the eggs and sperm are joined together in the fertilization step and the embryo is cultured. So it gives you a sense of the drawn-out nature of this procedure which is, I think, particularly challenging for the couples going through--when the woman often is working or has a life outside the IVF cycle.

[Slide.]

The in vitro fertilization step also

involves a number of laboratory procedures, and each of them provides a potential for an outcome; the number of eggs retrieved. The next step is that the eggs are stripped and they have to be equilibrated in the culture media. And then fertilization, how effectively was that performed?

Then the incubation provides at least up to six days when each of the steps of embryo development can be monitored in vitro. So these processes can be broken down and a number of metrics can be applied to each, useful for outcome studies if one is evaluating the efficacy of the various medications that are used.

[Slide.]

Controlled ovarian hyperstimulation is going to be dealt with by the next speaker in much more detail, but you can see at least these two commonly used protocols, the GnRH agonists protocol, first.

This is the one down below, the lower in the panel here, in which this agonist is administered in the luteal phase of the preceding cycle that we are actually going to move forward on the aspiration. Because there is an initial stimulatory effect, the luteal phase is the most

refractory phase of the menstrual cycle to gonadotropins. This flare effect is masked by the high levels of progesterone in the luteal phase of the antecedent cycle and then allowed to move, five to seven days later, into the suppressive phase when, then, the stimulation can begin and then, finally, the hCG trigger.

The antagonists are direct inhibitors of the GnRH receptor and, therefore, do not need to be preceded by--or administered in the preceding luteal phase. Instead, the gonadotropins are initiated at the beginning of the cycle directly and then the antagonist is added later towards the mid-portion to later portion of the sort of mock follicular phase finally hCG is triggering. So you can see at each point along the way, a number of potential markers.

[Slide.]

With regard to how we evaluate studies now, the study populations that are used, it is really important to note that there are--each of these procedures done, in vitro fertilization, intracytoplasmic sperm injury in donor egg, are very, very different. These are different procedures but they are administered in very

different patients. So it is very problematic to group data from each of these.

[Slide.]

They really reflect quite different pathology. The diseases that are being treated with these oftentimes are not even reflected in the diagnosis that may appear in the chart. Intracytoplasmic sperm injection implies, therefore--you know, it is a very significant sperm problem and assumes, oftentimes, that there is a normal egg complement although that also may be a concomitant problem.

Donor egg, at least when it is administered on the recipient side, means presumably the eggs are actually quite healthy. So there is a major difference in egg dysfunction with in vitro fertilization having a higher likelihood of poor egg reserve and function, more than intracytoplasmic sperm injection, and egg donation, of course, would have the least likelihood of having dysfunctional eggs.

Indeed, egg dysfunction, also known as ovarian reserve and very crudely estimated by the chronological age of the woman, is the best predictor in virtually every study that has ever

been published on the outcome of in vitro fertilization, more important than the diagnosis, more important, oftentimes, than the chronologic age, itself.

Indeed, there are log-order differences in pregnancy rates among groups of patients from a woman who is in her mid-40s with a diagnosis, say, of tubal disease as opposed to a woman in 20s. And there may be women who are matched by age but exhibit markedly different reproductive-aging markers.

So egg dysfunction underlies much of the outcome and has to be very carefully controlled in any studies, either through inclusion-exclusion criteria, case control or stratification.

[Slide.]

There are a number of issues regarding study design, particularly efficacy measures and the question of how we should define success and what are safety endpoints are also quite important in evaluating in vitro fertilization and the status of IVF today.

[Slide.]

Of course, deliveries, the number of babies that actually come out of a study, the

cycles that were initiated, is the gold standard in any study. But this is, as you can imagine, costly and the power needed to show differences when you have something that is happening 20percent to 50 percent of the time is quite large.

So a number of surrogate clinical outcomes, as well as surrogate biological outcomes, are frequently employed in studies. Ongoing viable pregnancies and clinical pregnancy rates, which is essentially a sac with a fetal heart detectable, are widely used as is biochemical pregnancy rate which is the very earliest marker of a pregnancy even before the sac or the fetal heart is detectable, typically with just the beta hCG rising appropriately, are widely used.

The surrogate biological outcomes that have been employed in studies include the number of follicles, the peak estradiol, the number of eggs that are aspirated and the fertilization rate as well as the embryo cleavage and morphology rates.

[Slide.]

The deliveries per initiated cycle, of course, is the gold standard but you need huge power to show differences between a study that--in a study where the expected outcome is 30 percent.

This makes is very expensive. It is also difficult to measure, at times, but it is important because there may be huge patient-specific differences in groups that are oftentimes subtle and not obviously reflected in simple things like the FSH or the age.

[Slide.]

The surrogate clinical outcomes are closer to the gold standard than the purely biological ones. You need much less power because of this, but they are clinically important outcomes as well. There are differences, though, in miscarriage rates according to protocols as well as with patients that may make it less than desirable, less than the optimal.

They also may be contaminated heavily by clinical practices. For example, clinics vary significantly in their level of cancellation, the criteria they will use to not allow a patient to go to transfer or not even to start a patient or, once they are started, to cancel them. So there can be a lot of confounders in any studies and these are ones that clinicians have been very careful to evaluate themselves when they look at the study.

[Slide.]

Looking at biological outcomes, they are

quite distant from the gold standard of the pregnancies per cycle started but much more sensitive. They may not reflect really clinically important outcomes. For example, there are many women that have a low response to controlled ovarian hyperstimulation but they still have excellent outcomes.

So if you were using just peak estradiol or the number of eggs in a 25-year-old, that is something very, very different than if she is a 45-year-old. There are subtle differences, as well, in drug potencies on egg yield in estradiol that may be interesting and significant statistically but not so important clinically because you just manage that by changing the dosing.

[Slide.]

So how should success be defined with assisted-reproductive-technology studies? It has been proposed that there should be placebo controls. It is a little hard to do that in a situation in which, say, somebody is going through in vitro fertilization because they have blocked tubes.

[Slide.]

Rather, I think success should be defined according to the pregnancy rate but also according to a number of other issues which include things like convenience and discomfort level. The pregnancy rate is important but there are a number of other factors.

You saw, at the beginning, when I discussed the IVF proposals, how complex and time consuming they are. So a study which showed equivalence but convenience, as one of the outcomes, was superior in one of the arms could actually be a superior product. So convenience and discomfort are very important.

[Slide.]

The other issue is that, with regard to the importance of accepting sort of noninferior or equivalent drugs, is that we really need competition in this area. There are also a number of patient-specific preferences. Options in allowing patients to choose would be enhanced. There would be an enhanced sort of customer satisfaction.

For example, some patients prefer vaginal-route progesterone over intramuscular administration of progesterone, or vice-versa. So

it is important to have variation and different options available for administration routes and other factors.

An example of this would be, as I mentioned, the vaginal progesterone versus intramuscular and, of course, the differences in the number of days that the agonist is administered versus the antagonist. I showed you at the beginning the very long cycle that the agonist requires and the antagonist shortens that.

So, just look at pregnancy as the only outcome would really miss important factors because of the time cost and the convenience cost and the comfort cost that couples accrue, particularly the woman, as they proceed through IVF.

[Slide.]

Safety endpoints conventionally include the ovarian hyperstimulation syndrome, miscarriage, multiple pregnancy rate and ectopic pregnancy rate. Because ectopic are so rare with IVF, I won't even discuss them.

[Slide.]

But, obviously, ovarian hyperstimulation--this is a life-threatening condition, potentially life-threatening condition,

in which some factor emitted from the ovaries, secreted from the ovaries, confers a vascular permeability throughout the body of the woman and can lead to ascites, large accumulation of fluid in the abdominal cavity, pleural effusions, instabilities in the hemodynamic system as well as increased coagulation. There have actually have been a number of deaths, strokes, loss of limbs from clotting and vessels.

So this really sets the upper limit on the controlled ovarian hyperstimulation that the woman is going through for in vitro fertilization. There is a pretty good correlation between the amount of stimulation, the number of eggs, the peak estradiol, sort of the number of lottery tickets we buy each time we put the woman through IVF and this sets the upper limit.

It constrains how high you can go. The risk may be modified by lowering those peak estradiols. A number of examples include recent introduction of aromatase inhibitors and luteinizing hormone which may alter the peak estradiol and alter some of the estradiol-related molecules, vascular endothelial growth factor, and so on, that may be mediating the risk for control

of ovarian hyperstimulation.

So it may alter those, but not alter the success rate. And that would be a big win.

[Slide.]

Miscarriage is a very common occurrence following assisted reproduction as well as following any pregnancy. It is quite common, even in young women going through IVF. Up to 15 percent of them undergo a miscarriage after conception. In older women, women who are in their early 40s, up to 70 percent of their pregnancies will end as a miscarriage.

These rates are highly affected by patient-specific factors--for example, the age of the woman, her ovarian reserve--but also could be influenced by a number of stages of the assisted reproductive technology that she is undergoing. The stimulation regimens theoretically could influence the risk of miscarriage. Particularly overstimulation with luteinizing hormone has been shown to disrupt developmental potential of the embryos, disruption of the normal luteal-phase support as well as a number of laboratory-related processes that we are not discussing this morning, the culture media.

So this is a very important endpoint.

[Slide.]

Multiple gestations are major risks for IVF. It is very common, between 15 and 50 percent, depending on the aggressiveness of the center. There are, of course, major obstetric pediatric and public-health concerns from multiple gestations including prematurity. Cerebral palsy risk is increased threefold with twins and twelvefold with triplets, Caesarian section rate from almost 100 percent with triplets to 40 to 50 percent with twins, preeclampsia, gestational diabetes. The list goes on. This is a major risk which has actually reached public-health proportions.

It is affected by patient-specific factors; for example, her age and her ovarian reserve. It is affected by rather illusive clinician practices, the number of viable embryos transferred. It is obvious that the doctor is very committed to optimizing the success rate, not only for the patient but for his or her own center, and it is a very tricky thing to figure out exactly how many embryos to balance the needs of the couple going through the procedure versus the bigger concern about preventing multiple gestations.

Of note, monozygotic twinning is also increased significantly after all forms of assisted reproductive technology, not just after IVF, not just after blastocyst transfer, but even after clomiphene citrate ovulation and gonadotropin ovulation and induction. This is a real significant problem as well. While it is not anywhere near as common, it approaches 3 to 4 percent in some IVF practices including blastocyst transfer as opposed to the 1 in 800 baseline.

It is a significant cause of morbidity because of the twin-twin transfusion syndrome that can result from a monozygotic twin; that is, where a single zygote or embryo later has split. It is, by the way, more difficult to control through modifying the number of embryos transferred because you could put in one embryo and still end up with monozygotic twins.

A number of other developmental anomalies are just sort of appearing on the horizon. It is more of a question of whether some of these should be included as potential risk factors or safety concerns with assisted reproductive technology. I think it really just requires more studies before we conclude about their importance.

But imprinting abnormalities, and particularly the Beckwith-Wiedemann syndrome and Angelmann syndromes have been implicated in complications of IVF. Pregnancy-induced hypertension is increased even when you control for multiparity following in vitro fertilization and Baha Subai and others have argued that one of the etiologies of pregnancy-induced hypertension is an imprinting abnormality.

So I think we just need more data before we know. They tend to be quite rare to begin with so it is a little hard to know to what extent what we are seeing is a detection bias.

[Slide.]

Finally, just to look into the future of where IVF is going and, particularly, assessment of IVF, we really do need more randomized clinical trials.

[Slide.]

We need multicenter networks doing these randomized clinical trials. I think those of us who deal with menopause patients are just overwhelmed by the impact of the Women's Health Initiative on our clinical practice and why can't we get this far in fertile patients. They deserve

it. There should be more randomized clinical trials to answer these questions.

In addition, we need more racial and ethnic diversity in our clinical studies to ensure generalizability, particularly as mandates for in vitro fertilization coverage spread across the country. I work in two states, Massachusetts and Rhode Island, where we have a mandate and it is really gratifying to see couples from all walks of life coming through our practice, not just very wealthy investment bankers, doctors and lawyers.

[Slide.]

In addition, we need to improve biological surrogate markers. I think is going to come before the other two. In particular, aneuploidy is ubiquitous and it is related to assisted reproductive failure. There is increased embryo apoptosis or cell death. There is implantation failure and miscarriage from aneuploidy that is abnormal chromosome number.

This could, then, provide a meaningful biological surrogate for outcomes. In addition, a number of safety problems in in vitro fertilization stem from attempts to overcome egg dysfunction and its core aneuploidy through controlled ovarian

hyperstimulation. Essentially, what we are doing is we are pushing harder and harder to get more and more eggs in the hopes that we will find one or two, and then we will put more and more in until we finally find the right one.

But if we could figure out which of those embryos are the developmentally competent ones up front, we would avoid much of that. There is some evidence that controlled ovarian hyperstimulation, itself, may predispose to aneuploidy by shortcutting sort of the normal selection process of follicles and by altering the follicular environment.

There are a number of new technologies on the horizon to diagnose aneuploidy which will make it practical. Just to give you an example, many of the studies we read in the literature will use high-quality embryos, or healthy-appearing embryos or viable embryos as a marker, an outcome measure.

[Slide.]

This just shows you a number of photomicrographs of a healthy-appearing embryo that, on Day 3, was diagnosed with trisomy 21 by preimplantation genetic diagnosis with normal development of blastocyst. Indeed, there is some

evidence that certain trisomies are more likely to reach blastocyst than others suggesting that blastocyst development culture in vitro to this later stage of Day 5 or 6 is not the answer and, as we know, it also may introduce other complications through enhanced stress on the embryo.

[Slide.]

Preimplantation genetic diagnosis is probably going to improve the implantation rate. This is a study from Gianaroli showing a doubling of the implantation rate when he prescreened them with a set of nine probes. It seemed to predict the outcome.

[Slide.]

Here patients had failed IVF three times and were, then, submitted to preimplantation genetic diagnosis and were either found to have no one or a greater than one normal embryo with these limited number of probes.

[Slide.]

You can see, at each point, an increase, a significant increase, in the birth per patient suggesting, again, that this is going to be key. Tony Pellicer in Valencia showed that you could also reduce the pregnancy loss in IVF patients that

he had gone through.

[Slide.]

Gonadotropins are key to this. This is where evaluating gonadotropins and how well they do, we know that it is the gonadotropins that, of course, when we stimulate the immature follicles, develop and it is these spindles within the eggs that are teasing apart the chromosomes here at metaphase 1 lined on the metaphase plate. This is when the aneuploidy first appears.

[Slide.]

You can see that there are abnormal spindles. This is a shameless self-promotion of my own research here for a moment, if you will--you can see abnormal spindles by Battaglia in these eggs from abnormal spindles.

[Slide.]

This just shows that we can actually image this noninvasively using some pole technology that we were involved with and, by doing this, demonstrate improved development and improved pregnancy rate when we do this noninvasive investigation of the egg quality.

[Slide.]

In addition, there is a new area that we

are working on in which we look at some of the chromosome structure and its propensity to aneuploidy and we have shown a number of factors that predispose to chromosome abnormalities in embryos that can be predicted by this biological marker.

So, in summary, then, in vitro fertilization is improving. We have got still a very complex proposal. We need it to be simplified. We need studies that show which of the better drugs--that use randomized clinical trials. In the meantime, we need to continue to improve these shorter-term biological markers as surrogates.

Thank you.

DR. GIUDICE: Thank you, Dr. Keefe.

Questions from the Committee

The committee now has several minutes to ask questions of Dr. Keefe, if there are any. While people are gathering their thoughts, David, I would like to ask you if you could please comment upon the issues and challenges of placebo control in ART treatment and also the issue of blinding.

DR. KEEFE: Okay. The issue of placebo control is problematic in ART for a number of

reasons. In particular, a number of the treatments that we now use in in vitro fertilization have already been demonstrated to be better than nothing. For example, the Canadian study has shown that untreated infertility, unexplained infertility, over the course of five years while Canadians were waiting to get into the Canadian health system had an expected fecundity of around 2 percent per month, which is significantly less than what we experience through most of these treatments.

In addition, a number of diagnoses for which we now use in vitro fertilization would be expected to have a zero percent pregnancy rate, especially completely blocked, occluded fallopian tubes. Finally, those with severe male-factor infertility in which there is no sperm, azoospermia, in which the sperm has to be extracted from the testicle and has essentially zero motility, or very low motility, and is only going to get into the egg through a direct injection route, I think it would be unethical to use a placebo.

I think in other areas, where there is questionable value, you can imagine the utility of

a placebo. Unexplained infertility when everything else seems to be working but there is no pregnancy is one example, a 2.5 percent pregnancy rate per month. I think that is one area where there might be some use of it, but, for the most part, especially for in vitro fertilization, I think it would be better to use existing technology as the control and then add in the additional.

And then the second was a placebo and then blinding? Blinding is difficult for in vitro fertilization compared to placebo, obviously, but, for in vitro fertilization with Treatment A as opposed to Treatment B, it is very reasonable. Blinding, of course, is a fundamental study design and is always desirable.

I don't think that there is much impact of blinding on the patient side. I don't think there is much of a placebo effect here but, certainly, on the clinician side where you can see the complexity of the treatment regimens, where there is a great potential for confounders to be introduced in terms of the way the cycle is handled, it is a very valuable strategy to be able to control for those potential confounders in which the doctors may be treating differently the treatment versus the

control group. So I think blinding would be very valuable, particularly with medications in which there is a potential to treat according to what you think might be the best or the more desirable way.

So I think both blinding on the patient and the doctor side was desirable although, obviously, on the doctor side of it is more important.

DR. GIUDICE: I think we all realize--

DR. KEEFE: There is a question. Do you have a question?

DR. GIUDICE: I think we also all realize, though, that in doing an ovulation induction cycle or ovulation enhancement for ART, if there is a placebo involved, that blinding is nearly impossible because of the follicular response. In monitoring a cycle with the placebo, if there is no follicle response, it is very clear on ultrasound and so that is an issue that I think is important as we look forward to the subsequent discussions that we will be having.

There are other questions. Yes?

DR. EMERSON: On one of your slides, you remarked about the superiority of being--whether it was necessary or not and you included in there

superiority against a placebo would potentially have--would not be a necessity and due to convenience of what? I guess I am not understanding where you would not require superiority against a placebo.

DR. KEEFE: The emphasis in that slide was that superiority does not need to be demonstrated to show superiority--if the outcome is pregnancy, it is not necessary because there may be superiority in other outcome variables such as convenience, pain, and so on. So the emphasis there was on the importance of taking a broad view of the outcomes, not just pregnancy rate, or not just ovulation, but also user-friendliness, the intrusion on the person's life.

So the placebo issue, as I mentioned, I think there is a limited role for placebo treatment in most IVF studies at this point, although I could imagine for ovulation induction, certain treatments are unexplained in fertility, there could be a limited role. So there, the importance of that slide, or the point was that it isn't just pregnancy outcome but also convenience and pain would also be important outcomes.

DR. EMERSON: Another question. When you

were listing your potential for improved biological surrogate markers, it wasn't immediately clear to me whether this aneuploidy, as a predictive value, could just be used to improve the efficiency of the whole procedure rather than be a surrogate endpoint, as itself. I mean, is there enough evidence, really, to suggest that it, as a surrogate, would be indicating whether the treatments were successful or is that just a means whereby we can improve the whole IVF process and it is not really indicating the drug response.

DR. KEEFE: Those are good questions. We don't have enough evidence at all. I mean, there are only limited studies from Gianaroli's group, from Santimine's group, that are suggesting its potential value. I mean, we really need to do more studies, larger studies.

But that really was looking to the future. As we look forward, I think we will see new technologies that allow us to look at all of the chromosomes. Once you have a helpful predictor of outcome, that, I believe, will become the most useful marker of the outcome.

So, if you have something that is predicting with high fidelity the implantation rate

of a given embryo, then that will become a useful marker of how you are doing with that, and one of those factors will be the treatment. It is true, though, that, as even with IVF, itself, probably the most important thing is not the drugs that are used or the way they are used. It is the patient. There is a huge patient-specific or patient-dependent parameter that is difficult to get our minds around and to measure.

I suspect that a lot of that will be aneuploidy. Then, once you have that nailed down, then you will be able to look at the potential effects of drugs or stimulation regimens on that. Until we do that, you know, we use things like age or FSH. These are explaining about 10 percent of the outcome in logistic regression equations. They are almost noise.

They are the best we have but we don't have a good understanding of the factors that drive the outcome. My bet, and, again, this is pure speculation at this point, that a lot of this will be aneuploidy. It will be the propensity towards aneuploidy. PDG is just the tip of the iceberg. I suspect that SKY or comparative hybridization will allow us to really get the handle on that and then,

once you have that level of determinism, then you can look realistically at potential impacts of drug-stimulation regimens on the outcome meaningfully.

Of course, as it is now, you can just randomize everything and it will all fall out. But you have a lot of noise in the system and it makes it very hard to do studies in a practical way that shows anything.

DR. GIUDICE: Dr. Macones, Dr. Rice and then Dr. Stanford.

DR. MACONES: Dr. Keefe, you had a couple of slides about ovarian hyperstimulation. I was wondering if you could give me some information about how predictable it is based on estradiol levels and, I believe, either follicle size or number.

DR. KEEFE: So ovarian hyperstimulation syndrome is a really, really adverse outcome that is very hard to predict. So, for example, there is a review by Mary Lau for Infertility and Sterility about six or seven years ago in which he sort of put it all together. Most of us use the cutoff that he used which is 3500 picograms per ml of estradiol at the peak to block the trigger, to stop

us from triggering. But that gives a predicted risk of hyperstimulation of about 5 percent.

So we are acting very conservatively because of the risk, the low risk, of a severe outcome. Just as PGD, preimplantation genetic diagnosis, should enable us to better understand the things that are driving outcome, the increased understanding of the pathophysiology of ovarian hyperstimulation, I suspect, also, will allow us to get around that outcome, adverse outcome--growing evidence that vasculoendothelial growth factor is one of the drivers of this vascular permeability may allow us to use that as a maker.

But, in the meantime, most of us use a very conservative cutoff to avoid the risk of the severe sequelae of hyperstimulation even though it is a quite rare outcome.

DR. GIUDICE: Dr. Rice?

DR. RICE: Dr. Keefe, you briefly mentioned the SART database which I guess now is sort of the SART-CDC database. I am wondering, your comments on that database and how we can effectively use it as to help us address some of the outcomes or address any differences in product.

I was struck by the fact that I sat on a

thesis committee for one of our colleagues who was getting his Masters about 1998, 1999 and he used that database as the basis for his thesis and really did show that African-American women who participate in ART procedures, even when you control for Day 3 FSH or age, had lower pregnancy rates and this was never really brought out in a real public manner.

So I am just wondering if we could better utilize that database to help us address some of these questions.

DR. KEEFE: That is a really good question. I think the SART database is a beginning. We have begun to answer some interesting questions. There is a tension within the community about the SART database. On the one hand, some view it as sort of a threat, that they are being exposed, the real size of their pregnancy rates is being revealed before everyone else.

They will often argue that, gee, the patients don't really want to see all these numbers. It is very complex. On the other hand, there is a huge potential. I think what we should use--get a lot more information from that on the patient side. There has been an argument that

there is too much on there already. Patients go to it and they get just overwhelmed by so much information in front of them.

I think that what we could do is we could use, like, hyperlinks. If you want to know more about the inclusion and exclusion criteria about a specific clinic, that should be put in there. One of the problems is there is a lot of variability among clinics and who actually is going through. That should be posted. It should be "buyer beware." In other words, all information available for those who want it and then you could have hyperlinks which you click and that brings you into inclusion and exclusion criteria for a given clinic.

If the inclusion and exclusion criteria are not published there, the patient should have full rights to go through it, whatever particular situation they have.

On the research side, I agree with you. There are a lot of interesting questions that could be asked. One of the concerns that I have had is I am not sure how available that database is to those who have questions. I have talked to a couple of colleagues who have approached them and discussed

projects, and they kind of go into a queue.

I am actually just newly elected to the Registry Committee for SART so that will be one of the issues I will be bringing up is accessibility. I think that the data should be transparent. It should be accessible. It should be available. People that have interesting questions like the one you are raising should have a shot at it. Only then, you know, we will be able to really take advantage of it.

As we used to say, they have all the answers but we have all the questions.

DR. GIUDICE: Dr. Stanford?

DR. STANFORD: Dr. Keefe, I wonder if you could comment generally on the issue that these ART protocols involve multiple complex regimens, as you mentioned, multiple drugs with multiple different scheduling of them. When it comes to investigating and trying to establish the efficacy of a new drug, you have all these other variables, all these other drugs, drugs that you are using with it that may or may not have been approved for that protocol or that indication as well as the patient factors.

This may be something that we discuss later as well and I would be interested in FDA

staff input, but I am just interested in your--what perspective you would have on that if you are trying to nail down, is this drug effective, when you have all these other variables.

DR. KEEFE: That is a great point. I think it was Alfred North Whitehead that said that science is the art of the answerable. This is a lot of art and not too much answerable. I agree with you. It is such a complex moving target. But that is why, in a blinding, at least with regard to drugs, it would be very helpful because if all that goes out in the wash, then it is less important. But if those other things are being tweaked in response to the biases on the part of the investigators about the control versus the new drug, you can imagine a lot of mess that would create.

But I think if you can let it all fall out in the wash, you could still find useful information out of it. It is just the numbers are just enormous when you have so much noise, so much error, in the equation.

You are right. It is a very complex experiment we are doing.

DR. GIUDICE: Dr. Toner?

DR. TONER: Just a point of information regarding the question of the SART database. It is, at this point, a joint effort with the CDC and, to use those data, you need permission from the CDC. That is doable. I have done it on a couple of occasions.

But probably more pertinent to the gonadotropin question is the fact that, at this time, the database simply collects how many units of FSH were used in that cycle. So you have nothing about the brand, nothing about how much LH was or wasn't part of it, or any concomitant medications that certainly would influence outcome.

So I doubt that that database, in its current form, will be of much use for this topic.

DR. GIUDICE: Dr. Crockett?

DR. CROCKETT: Yes. I am particularly interested in the research that you presented about the abnormal spindle formation and the aneuploidy rates. It struck me, as I was listening to your talk, when we talk about using control groups and placebos and double-blinded studies, there is very little discussion about comparing these outcomes to what are normal, healthy, fertile, maximum fecundability is as humans.

Particularly regarding the aneuploidy and the miscarriage rate, could you just discuss the ramifications of using our healthy human population as a comparison?

DR. KEEFE: That is a good question. Battaglia's research on spindles actually used normal volunteers. He had an NIH grant and normal volunteers were monitored and then their eggs were retrieved under the spontaneous surge, hence the low numbers in the study It is hard to get a single egg.

That is one of the few studies. Others have used natural populations, natural cycling populations, to look at aneuploidy rates. Placheau has found 25 percent of eggs are aneuploid in normal situations when they are aspirated. Also, Pat Hunt has found similar rates with using karyotypes.

There are no head-to-head comparisons, but if you look at assisted reproductive technology cycles, the rates go much higher, particularly as you get into women who are in their late 30s and early 40s. Using comparative genomic hybridization, and SKY, the group in St. Barnabas is finding the majority of eggs in women with

infertility have aneuploidy, so more than 50 percent, one study up to 70 percent, of the eggs could be identified.

There is also a false-positive rate with that. Hypoploidy is actually a missing chromosome that washed off the slide. But, still, there is a very, very high rate of aneuploidy in humans. One wonders why any of us are walking around talking like this because we are all survivors of this massive struggle for existence.

Some have argued that, in fact, the human has evolved as a mono-ovulator with subfecundity because it is such as social species and there is such an important role for socialization. You can see the same pattern of reproduction in other long-lived social species like whales and elephants and primates.

So there are two parts to it. One is that we are mono-ovulators and we sort of get programmed in a very limited fecundability and, on top of that, we live long. So you have sort of two hits. You have a defective egg, or a low number of eggs that are likely to be defective at any given time plus a long period of time where wear and tear sort of damages what minimal reserve we started with.

DR. CROCKETT: I guess what I am asking is when we look at these reproductive technologies, is our goal to overcome that or to match it.

DR. KEEFE: That is a good question. A lot of people argue, well, gee, disease--aging is not a disease. Aging is a natural process. But I think that draws a distinction that doesn't exist in the rest of medicine. For example, I have a bad hip. No one has ever questioned that, as a 49-year-old walking into an orthopedic surgeon, that I shouldn't expect to have a bad hip when I work out every day and so on. Yet, no one questions whether that should be covered by insurance.

I work in a state where there is an IVF mandate but almost every day I have to battle with the insurers that a 38-year-old, 39-year-old, 40-year-old going through infertility with a high FSH is going through a natural process.

Nature has never been thought to be benign. I mean, nature causes diseases. This is sort of the equivalent of a balanced polymorphism. It is like sickle cell in certain populations that lived in environments where that was advantageous. I think reproductive failure in women was

advantageous at a certain time in history. Currently, it is a disease. It gets in the way of women who make great sacrifices in their lives to do things for other people and then, at their stage of life, are now prepared to complete a family and nobody told them that, by the way, this is a natural process, you should have done this instead of go to law school or med school.

So this is a disease like any other disease. You know, sure; it is a natural process but it is a natural process which wreaks havoc. It is a misadventure for our current society. It is the other side of the contraceptive evolution and one that we should, I think, attack with equal vigor.

DR. GIUDICE: Dr. Brzyski?

DR. BRZYSKI: I would like to ask, thinking about the different types of patients, you mentioned a little bit about the different types of patients that are treated with fertility medications. Do you think that that should be taken into account in terms of the safety and efficacy outcome variables that we look at?

For instance, a specific example I was thinking about is women with polycystic ovaries

have a high miscarriage rate. So you could envision an intervention that might not be very efficacious in terms of ovulation rate but may significantly reduce miscarriage rate whereas that might not be an outcome that might be of relevance to someone, say, with hypogonadotropic hypogonadism.

So I just wanted your opinion about whether the outcomes that you look at should vary from patient population to patient population or should it be standardized across the whole drug group?

DR. KEEFE: I agree with you there. There are important diagnostic distinctions to be made. The one you described is probably one of the most important as well as the male-factor patients that--there is a big difference between hyper-responders, particularly those with a polycystic-ovary-type dynamic in unexplained infertility patients or patients with tubal disease and what is going on inside their ovaries. So that is a useful distinction.

I think less useful distinctions are minimal endometriosis, mild male factor, unexplained infertility. Those diagnostic

categories are much less meaningful. So it might be useful to set up a series--sort of revise our diagnostic categories as a field. We have been very hazy in our thinking.

You wouldn't see this with cancer or heart disease. You have different stages. You have very clear-cut diagnostic categories. We tend to use these textbooks that we all read back in, you know, the first year of medical school about the causes of infertility. Most of them are not really that useful.

I think the one you mentioned is, polycystic ovary syndrome, severe male factor. Those really mean something. Complete tubal disease, tubal obstruction, those are meaningful. And then the rest, sort of the rest, of the diagnoses have much less meaning in terms of outcome. They are not anchored as much in--they are not as valid.

DR. GIUDICE: Dr. Rice?

DR. RICE: I think Dr. Toner made a good point when he captured really the essence of what I was getting to that we probably don't ask all the right questions of the SART database in order to do the right research.

I want to ask you your opinion on this. You are in a mandated state and I know the insurance carrier there requires that you get a certain amount of information on a patient before starting her on an IVF cycle. Are you finding a level of consistency by these insurance carriers requesting similar types of information when you are in mandated state versus when we are in a non-mandated state where people don't have maybe the same criteria for starting a patient on IVF? Are you seeing any inconsistency?

DR. KEEFE: There is a lot of inconsistency. It is a huge problem. We spend a lot of time haggling over whether somebody, indeed, merits IVF treatment or not. We have convened a group, sort of working group, of the different IVF programs to come up with a consistent series of inclusion and exclusion criteria that we can all agree on.

The problem there is that, even within our group, there is a disagreement, particularly as you get into the low-prognosis patients. Are those the ones that should first be treated because they have the least probability of conception on their own? Or should they be the last that are treated? So

there is a lot of inconsistency.

DR. GIUDICE: Are there any other questions from the committee? Yes; Dr. Shames?

DR. SHAMES: I just wanted to make a comment sort of to prod the committee, and this is early in the process, to start thinking a little about the problems that we have as regulators. Dr. Keefe brought up one point which makes me want to use as an example. Is the point of doing a noninferiority trial, say, for a drug in a situation where the procedure or drug would have an advantage of convenience or discomfort, et cetera, et cetera, which we all certainly understand.

The problem for us, of course, is how to measure the convenience and discomfort, how to incorporate it actually into a statistical analysis plan, et cetera, et cetera, on top of a noninferiority trial. I am not asking for details. I am just saying ultimately we want to sort of standardize things which we will believe will help move this process faster.

We need to standardize to be sort of fair so when we discuss things like this, any of these issues, we should have in our minds how we actually are going to translate this into real trial designs

and analysis plans, things like that. That is my comment.

DR. GIUDICE: That is an important comment. We will get to Dr. Lewis in just one moment. Along the same lines, in addition to outcomes, there is also the issue of time for getting to a particular outcome which, in infertility therapy, some women respond very quickly to gonadotropin stimulation. Others take a very long time. Those are measurable outcomes.

On the patient side, there is also the issue of getting to the clinic or getting to the office, time away from work. So, somehow, I think we also probably should keep that in mind.

Dr. Lewis?

DR. LEWIS: I agree with the points you just made. the other thing I wanted to just mention briefly about the SART database, while it is useful to look at, it is all retrospective data and it is not controlled. It is particularly biased by the specific clinic and what their criteria or prescribing habits might be as well as their selection criteria which vary hugely.

So I don't think that it can substitute for a well-designed, prospective trial.

DR. GIUDICE: Thank you for your comments. Are there any additional questions from the committee? Yes?

DR. EMERSON: I would just like to visit the question of the gold standard that you listed which was the number of live births per cycle of therapy. I guess, you know, there are some worse in terms of needing larger sample sizes and you could be looking at the time until you had successful live birth, so measuring the number of cycles needed until someone got that. These have some problems as well.

I guess what I am worried about most in all of these is the ability of bias to creep in for one treatment versus the other in terms of the clinicians response to cancelled cycles or to not trying further cycles with a patient. Could you comment on that?

DR. KEEFE: That is a huge factor. In clinical practice, it is an enormous factor, enormous, because now, all of a sudden, any patient can go to the web and pull down the statistics which are a reflection, largely, I think, today, with standardized media, with standardized training for most of the embryologists, with standardized

protocols for clinical, a lot of the variability is arising from the patient mix.

There is a growing practice of explicit or implicit exclusion of the sickest patients which I think is immoral. On the other hand, some would argue that physicians are increasingly charged with allocating scarce resources. But those are tricky and difficult to balance. I don't think, from the standpoint of the traditional mission of the physician, it was our major commitment.

There is a growing practice of what I call the Lake Wobegon effect where everybody is above average in my practice so we are above average. That is a problem.

There was an article in The New York Times Magazine about healthcare about two or three months ago. The only other area of medicine that has a registry that is so visible as that for in vitro fertilization is cardiovascular surgery. That is exactly what is going on with cardiovascular surgery.

There was an article by a cardiologist in New York City about how difficult it is to get a bypass on a patient who really needs it in New York City today because the "best" surgeons don't want

to touch the poor guy. So that is a problem for all of us in medicine.

I am the first one to believe in consumerism. There is sort of an avalanche of consumerism. That is great. But my response to that wouldn't be to stop the data from coming but to increase the flow. So, for the registry, I think we would mandate inclusion and exclusion criteria be attached through a hyperlink to those clinic-specific SART rates so that, if a patient goes there and they get cancelled, and they go on the web and say, whoa; wait a minute. This isn't on your exclusion criteria for who can to through the next cycle, then they have a right to go through again.

I think the AMA and ACOG, ASRM, all have, in our ethical sort of statements, the right of every physician to not treat a given patient as long as it is standardized and mutually agreed upon and it is explicit.

So I think the data should be there so that couples can decide, up front, or women can decide up front, whether they want to go to this center that is extremely exclusionary versus another which isn't and so they can go the whole

ride if they need with a given center. That is the only way I think we can deal with it.

We can't have less data, less information. We need more information. The argument that patients don't want to see all this. They don't have to. They can have a hyperlink.

DR. EMERSON: I guess one of the points I was at was a more technical one. It is the concept that, for instance, there can be differential actions by treatment arm according to deciding to cancel a particular cycle and not go forward with the stimulation or whatever.

DR. KEEFE: That is a good point. In terms of the outcome, if you were randomized, again, that should fall out in the wash. You might want to balance both arms of the treatment within a given center because they tend to be center-specific criteria. You also might want to make explicit up-front the clinic's criteria before they start to enroll patients.

DR. EMERSON: Although, it won't pick that up if you are using as your denominator the number of cycles. There is a lot of room to play with number of cycles that you go through, both in terms of sort of the--I always look at things from the

"intent-to-cheat" perspective and say, how can people cheat on a particular thing to make something look good.

If you set out to do it, what would you do? And I say, well, gee, if I want my treatment to look good, what I will do is, as soon as I hit a patient that looks pretty hopeless from, for instance, aneuploidy or whatever, that they do that, if they are on the arm that I don't like, I will just keep on sending them through cycles and, if they are on the arm that I like, they won't.

Similarly--

DR. KEEFE: Can't you just deal with that by, once they get included in the study, then they get randomized after inclusion?

DR. EMERSON: The idea is--the problem is using the number of cycles as a denominator whereas, if you just used time, and said, we are going to measure everyone as the time until you achieved a live birth, and people who you cancelled permanently after that, and you just refused to do more cycles on them, well that is at infinity.

DR. KEEFE: The other way to do that is just to include first cycles in the study which I think, in terms of the way IVF works, is probably

not a bad idea because there is that huge source of confounder which is whether they get to go through it again, according to the policies.

But there is also a difference, not so much first and second, but first and third, first and fourth, in terms of the mix of the patients. They are different people.

DR. GIUDICE: Yes; Dr. Stanford?

DR. STANFORD: I think there is another potential concern. You talk about a gold standard being live births per cycle initiated. But then there may be--ART, obviously, there are certain categories that you mentioned where it may be the only way for a pregnancy, azoospermia, et cetera. But there are other categories where it is becoming more broadly used and there is debate about should you go with controlled ovarian hyperstimulation versus ART or metformin for certain categories of PCO and all that kind of stuff.

I don't think you can make a fair comparison on the per-cycle pregnancy basis. I think, in those cases, you have to come up with some kind of measure that is an overall--given a course of treatment, whatever that is, which may take longer on a non-ART side or a non--whether or

not you call controlled ovarian hyperstimulation part of ART, but on a non-IVF side, you have account for another course of treatment that may be cheaper, may have less risk in some ways but may take longer.

Could you comment on how you could compare those in a randomized way?

DR. KEEFE: One study design that I think has been underutilized is a crossover study. Because the majority of our patients are not going to get pregnant in the first cycle, and the first and second cycles are not that different in terms of the patient mix, one way to do it is to do a crossover where each patient is their own control.

Of course, there are problems with that, obviously, but one of the huge advantages is that, as I mentioned, and I think everybody would agree in this room who takes care of these patients, that the biggest confounder is the patient, herself, her eggs, her aneuploidy, whatever is going on inside there.

There are fertile people and there are infertile people. So if you randomize the patient up front to Treatment A in the first cycle, or B, and then flip it around on the other side, then I

think that is a very sensitive way to tease apart some of these.

Of course, it gets around the issue of not allowing them to go through because they are stuck. I mean, they have got to do at least two cycles if they don't get pregnant the first. If you get pregnant with the first, obviously that creates a problem; they don't get to the second, and how you treat those.

But, given the problems with the other study methods, you can imagine a way that that could be a very powerful way to be able to put together all these very complex factors that are influencing the outcome.

DR. GIUDICE: Yes?

DR. TULMAN: I have a question about the gold standard as well. You mentioned the gold standard being a live birth. Could you comment on adding the adjective of "healthy" live birth to that?

DR. KEEFE: Okay. The issue of anomalies with assisted reproduction is really important. I put them in safety concerns. I think I would still leave them as a distinction. I have, in my practice, my office, last week a woman who

came--she is a physician, actually, an OB-GYN, who brought her twins from IVF. Both have cerebral palsy and she was coming back for some more.

I said, "How can you do this? You must have a lot of stuff on your plate." She said, "They are my babies. They are my kids. People say they are abnormal. I love them. They are beautiful." And she wants more.

Now, that is not everybody. I think there is enormous value for somebody who has never had children to have a baby. Not everybody receives their cerebral palsy the way she did, but I would still put down the birth of a child as a benefit. I would put down the others as side effects, as potential complications, and try to eliminate them or reduce them.

We don't know all the mechanisms. We know that the biggest component is the multiple gestation. That is the biggest driver. But there may be other abnormalities as well that arise from imprinting abnormalities, problems with polarity. This monozygotic twin thing is huge. It is not just the zona hardening. There is a lot going on there. They tend, overall, to be so rare that you have to use a study like the Australian with 10,000

patients following in a registry over huge periods of time before you find them.

So I would hold fast that we still are looking at babies. I think the definition of a healthy baby--it is important to put that in as a complication, but, for the vast majority of patients, and I think a number of studies have looked at that, they are very happy with that.

DR. TULMAN: Just a follow up, if I may. Is there any move within the FDA or CDC to establish some sort of registry within the United States of the children who were born as a result of the reproductive technologies to follow them for not only the problems that may be apparent at birth or soon thereafter but in terms of a longer range, in terms of other health problems?

DR. GIUDICE: Dr. Brzyski?

DR. BRZYSKI: I will comment on that. I am the President of the Society for ART and so work very closely with the CDC on the registry. That discussion has been, at several levels. CDC has certainly recognized in their publications the issues of, say, adverse outcomes and the problems that are associated with trying to identify adverse outcomes such as birth defects, congenital

anomalies. That has been discussed in some of their publications.

From the CDC's standpoint, well, let me just say that from the consumer standpoint, that is certainly an issue that I have discussed with patient advocacy groups. There are sort of two opinions about that. There is a group that believes it is extremely important to collect this information and another cohort that is concerned about privacy issues and stigmatizing the children born from ART technologies by sort of identifying them or categorizing them in that way.

So it is definitely an issue that needs to be discussed more, both from sort of a philosophical standpoint and from a practical standpoint in terms of the cost of generating such an effort.

DR. GIUDICE: I am wondering, Bob, if you could comment on the European experience of keeping long-term outcomes compared to in the United States because that may address some of the question.

DR. BRZYSKI: I don't have direct experience with this, but there have been discussions at the SART Executive Council level of the experience in Belgium with ICSI outcomes. My

understanding is that it has been an extremely expensive and time-consuming effort to collect those data on ICSI cycles, even in a relatively homogenous environment of a small country like Belgium. Well, I guess it is not really that small because there has been some--those patients have probably traveled from various places.

But it has been a noteworthy effort to collect those data but at a significant price.

DR. KEEFE: I think the European experience has been an excellent one in terms of generating very useful data for the rest of us to use. We started on a small project that was funded by the Rhode Island Foundation to look at outcomes of IVF babies born in Rhode Island because we are sort of a Mayo Clinic, the only IVF program for this million-and-a-half people.

So we had a lot of potential there. About three months into the study, we got a letter from a very irate former patient who had a beautiful baby from IVF but was, herself, a lawyer for the American Civil Liberties Union. She was absolutely livid that this letter arrived at her house, which had been approved by our IRB, that her mother-in-law had inadvertently spotted and

mistakenly opened because they have similar names, threatening to sue us for revealing--the problem is we are in the frontier here and we still have a lot of sort of notions of individuality and freedom and so on that it makes it a little harder.

They shut down our study. That one example was enough to scare the IRB and the hospital to shut it down and say, "No more." So it is tricky, in America, where there is so much concern of privacy and individuals' rights. I think it is valuable information, though.

DR. GIUDICE: I think it is also important to point out that there is no organized effort for keeping track of babies born from gonadotropin stimulation cycles as opposed to ART.

Dr. Layman, did you have a comment?

DR. LAYMAN: Yes; I had a comment. I think it is important to try to collect the data. I know it is difficult, especially with HIPAA. We know that, in general, congenital anomalies are in 2 to 4 percent of all couples and that IVF increases slightly sex-chromosome abnormalities and some other aneuploidies, but I think, as far as the issue of the imprinting disorders, if we don't start getting a database, we will never know for

sure whether this is a real issue and whether it is due to IVF or ICSI or both, or whatever.

So I think it is important to address this, at least future SART collection data.

DR. GIUDICE: Thank you. Are there any further questions? Yes, Dr. Emmi?

DR. EMMI: I just wanted to ask a question, or if you would just comment. You had said that different infertility diagnoses are not quite so important. Would you just comment about different institutions using different policies for diagnosing conditions in in vitro, even, as far as unexplained or male factor and how that might affect the studies.

DR. KEEFE: The rigor of the diagnostic categories used for IVF is very low. So you will see frequently minimal endometriosis scattered throughout. That is probably a much better marker of whether the patient had a laparoscopy as part of her workup than it is about her disease or her prognosis.

Similarly, with unexplained, it depends on how extensively you look for other factors. So, as I mentioned, I think the most important diagnoses are severe male factor, complete tubal occlusion

and polycystic ovary syndrome with insulin resistance. These are the sort of diagnostic categories which I believe, in the end, will hold their own, at the end of the day, will still be standing after critical review.

We, as a field, I think, should revamp our diagnostic categories. I think we should also have a dual classification for infertility, differing staging for the level of infertility rather than--age is a very, very poor predictor of anything. Chronologic age is not a good predictor of biological age.

Just look at anybody's clinic when you look at patients who had an anastomosis failed or who had a tubal ligation and they go through IVF. They have a very good prognosis independent of age. If you look at the registry, these studies that use sperm donation clinics to look at the effects of age on fertility, 45-year-old women still have sort of a 30 percent pregnancy rate, 25 percent pregnancy rate.

The fertility rate doesn't drop that precipitously in the general population as it does in the IVF placebo because the IVF population is that subset of the general population that really

has it bad. I mean, clearly, age affects fertility but the relative impact of age on a given--or senescence, I should say, or aging rather than chronologic age, is highly variable. That is the key factor.

We should develop way more rigorous work. I know Jim Toner's very important work on FSH is a beginning, but there is a need for increasingly precise measures of a specific patient's individual senescence level or aging factor.

DR. GIUDICE: Dr. Shames?

DR. SHAMES: I just want to make a comment that the FDA can and does ask sponsors, as part of the approval process, to do some pregnancy registries. However, this is voluntary and we cannot compel them to do it.

The other problem, of course, is, in this area, we are dealing with multiple drugs and procedures, for that matter, so it is difficult, of course, to tease out the cause of some particular abnormality. But we can, in the process, ask sponsors to do that.

DR. GIUDICE: Thank you. Dr. Rice?

DR. RICE: Why does the FDA have limited authority in requiring that part of the

registration of a drug, approval of a drug, not lead to a pregnancy registry?

DR. SHAMES: That is sort of a legal question. There are certain things that we can and cannot do by regulation. I mean, I would have to go back and ask that, why we do not have regulatory authority. It would have to be created as a regulation or a rule which has not yet been done. Whether we would even have the right to do that, is a whole legal issue which I cannot speak to. But we can certainly investigate that situation.

DR. GIUDICE: Thank you. Any further questions? If not, I would like to thank Dr. Keefe again and the committee members. Let's take a break and return at 10:15. Thank you.

[Break.]

DR. GIUDICE: For the second half of this morning's session, I would like to introduce Dr. James Toner who is Director of the Atlanta Center for Reproductive Medicine in Woodstock, Georgia. He will tell us about Gonadotropins in ART.

Gonadotropins in ART

DR. TONER: Thank you, Dr. Giudice and thanks for the invitation to attend this panel.

[Slide.]

I hope it leads to more streamlined ways for us to understand what kind of efficacy endpoints are expected of us in evaluating new drugs. I was asked to talk about a few different components of this and we will start, really, by talking about in vitro in this country, kind of how we have come in the first twenty years of use and then tie in the approach we have taken on the clinical basis to achieve these improved outcomes, especially as regards to gonadotropin usage patterns.

[Slide.]

As most of you in the room know, IVF has been successful for almost twenty-five years now, the first birth being in England and a birth shortly thereafter in the United States in 1981 in Norfolk, Virginia.

[Slide.]

The story within this country, ever since that beginning, has really been one of improving success rates, reduction of multiple pregnancy rates, introduction of new therapies. It also, I think, is a story of the advantage, in a way, of the American system, the flexibility we have had as clinicians in this country to incorporate new

treatments. To adjust the number of embryos going back according to age and other things has led to higher success rates in this country than certainly in Europe.

[Slide.]

The simpler treatments include things such as the ability to remove eggs transvaginally. It used to always require a laparoscopy and a transvaginal approach under a light sedation is certainly an improvement and, also, the essential abandonment of replacement of embryos into the tubes, also through laparoscopy.

In the early days, most of us also asked the women to come daily or nearly daily through the whole stimulation process to evaluate their response. That is not necessary anymore. Also, in the beginning, essentially all the medications that we needed to give were given by the intramuscular route. Now, with the possible exception of progesterone, almost none of them are.

So those have all been improvements in the treatment simplicity.

[Slide.]

There has also been introduction of new therapies. Initially, it was designed as a process

to deal with the problem of blocked tubes but it has since turned out to be quite effective when sperm numbers or quality are quite diminished, when eggs, themselves, are diminished by really substitution therapy, using donor eggs. When the uterus has an intractable problem, a carrier can be brought into the mix and that is a very successful strategy.

Initially, surplus embryos were a problem but cryopreservation is now quite an effective use of those extra embryos. As Dr. Keefe alluded to, genetic problems, either that come with age or might have been there all along, in particular, couples can be screened for with preimplantation genetic diagnosis. None of these things were there at the start.

[Slide.]

Now, the innovations over time have really proved to be a fairly steady progression. Cryopreservation in the early 80s, donor egg in the mid-80s and then the attempts at putting the embryos or eggs back into the tube, themselves, where they normally would have been by GIFT and ZIFT were widely practiced for a time but, as you can see, are not common anymore. Co-culturing of

embryos, dissecting away the zona to some extent in hopes of permitting sperm easier entry or simply putting the sperm under the zona were early efforts to achieve fertilization when sperm were a problem, again, approaches that have since vanished for all intents and purposes.

Those have been replaced by ICSI, the direct injection of sperm into eggs. Hatching was introduced in the mid 1990s, preimplantation genetic diagnosis, again in the mid-90s. In efforts to get around problems of diminished egg quality, attempts have been made to either transfer cytoplasm from healthy eggs into aging eggs or to move the nucleus across to a healthy egg.

So, in the field, there really have been steady innovations. The three that I have put Xs next to are no longer permitted to be done in this country until appropriate, I guess, safety studies have been performed.

[Slide.]

Now, the focus of this meeting really is gonadotropins. I wanted to illustrate, first, a time line that illustrates when these gonadotropins and the things that affect gonadotropins have been introduced.

As many of you know, the urinary products came first. hMG was available even years before in vitro came into existence and was really the only gonadotropin available for use in the earliest years of IVF. In the mid-80s, a more purified version, still urinary but with a dominance of FSH, was released.

In the late 80s, Lupron, which is a GnRH agonist was released and widely incorporated into practice shortly thereafter. More and more purified forms of FSH were also developed in the mid-90s, again urinary at first and then recombinant shortly thereafter.

The most recent development has been the development of GnRH antagonists. It took us actually years, on a clinical level, to figure out how best to use the GnRH agonist and we are still, I think, on a learning curve of the same sort when it comes to understanding how best to use the antagonists. But they have become adopted into practice pretty widely.

[Slide.]

Now, kind of with that as a background, let me describe to you what I think has been the general trend of use of these medications over time

for the purposes of in vitro. The very earliest success with in vitro was, in fact, in natural cycles. That is how the cycles went in England, initially. But the Jones, who tried to get it going in the States, had a very poor experience with that and, instead, moved quickly to Pergonal, an hMG product and used, as you can see, at very low dosages, 2 amps a day, typically.

It wasn't long thereafter that the advantage of even stronger stimulation was recognized because, other things being equal, within a certain range of hMG use, the more you give, the more eggs you can allow to grow or cause to grow. So 2-amp-a-day protocols became very uncommon and more typical dosages were 4 or even 6 amps of medication per day.

They were often blended with FSH in those years although, in the early years, there was typically always some use of hMG which contains not only the FSH but the LH product, the LH hormone, I should say.

When Lupron became available, it was first employed as a suppressive drug starting in the preceding luteal phase and seemed to very clearly allow more eggs to grow, at least more synchronous

growth of eggs, so that, in most patients, simply by adding this sort of a pre-treatment, you were apt to get more high-quality eggs than if you went without such a product, even with the same dose of stimulation.

A few years later, it was discovered that you could take advantage of the fact that the agonists first cause a flare of gonadotropins and that flare was then used to augment the stimulation and was begun typically at the early part of a menstrual cycle along with the other stimulatory drugs to cause an even stronger response.

For a time, there was a lot of advocacy of the use of pure FSH rather than the hMG-style product through the mid-1990s but many clinicians have since gone back to a blended protocol where you are taking at least some LH in many cycles.

Birth-control pills were introduced in the mid- to late 90s as a way to control the cycle start time, especially in a case where you are not going to use a suppression protocol of this sort. Without PCO pretreatment, you really had very little control of when the cycle would actually start and the pills have been useful in that arena.

Then the most recent change has been the

introduction of the antagonists into clinical practice which have largely been used simply late in the stimulation to prevent the premature LH surge.

[Slide.]

Now, with those medications being used in the way that they have been used, there have been certain trends over time. Clinics; we have, obviously, many more treatment clinics in the country now than we did in the early years which, in turn, has led to many more cycles of treatment, now over 100,000 such on an annual basis and many more deliveries.

[Slide.]

But really, more importantly, is the fact that the success rate per treatment cycle has steadily climbed no matter which major kind of treatment you want to consider. In vitro is here in the middle, in green. In the early days, 15 percent of cycles produced pregnancy per transfer. Now, 40 percent of them do.

The orange at the time is donor-egg success rates which have doubled from 25 to 50 percent per try. And freezing, which didn't work well at the beginning, in yellow here, now adds

about 20 percent or 25 percent advantage to the cycle if you have had extras and can freeze them.

The other two little things in here are the ZIFT and GIFT which were, at first, more successful than standard IVF, at least on a national basis, but, as the IVF success rate rose, that advantage has disappeared and, in a sense, killed off that therapy altogether. This shows the proportion of cycles over time that were GIFT and were ZIFT. You can see these things peaked in the late 80s but, as the success rate between them and standard IVF has increased, their usefulness has gone down to the point where only 2 or 3 percent of cycles are done by that method.

[Slide.]

One real success story has been--my computer has frozen. I will tell you what that slide said. I don't know why it wasn't working. In the early days of in vitro, when sperm were not normal, simply putting sperm in the eggs, sperm in the same dish as the eggs, produced some pregnancies but nowhere near the normal rate.

The short story is that this direct injection of individual sperm into eggs, ICSI, has taken a very low poor outcome that was happening

before the technique was introduced and erased that problem altogether to the point where now, as long as there is sort of one living sperm for every living egg, it is as if there is no male-factor problem at all in terms of pregnancy rates.

[Slide.]

This shows the incidence of multiple pregnancies over the years as well which have been highest in donor egg and lowest in frozen embryos and mid-range, about 35 percent, in standard IVF which is higher than we would like but I did want to emphasize the problem of triplets and quadruplets, at a minimum, has certainly been decreasing. The last year for which we have data is already four years ago but, even then, you could see that the triplet rate was declining and, also, and dramatically so, the quadruplet or worse rate.

[Slide.]

One factor in IVF success that continues to be a thorn in all of our sides is this very strong effect of a woman's age. Apart from substitution therapy with donor eggs, we have not really been able to lick this one with in vitro fertilization. These are the published results on the CDC's website from 2000, the top line being

pregnancy rate and the bottom line being delivery rate.

You can see that, by and large, the pregnancy rate holds together pretty well until the mid-30s and then declines with each year, almost being a zero percent pregnancy rate at 45.

[Slide.]

The flip side of that is miscarriage which increases with those same years. Again, this is not something that we have been able to work around using in vitro fertilization.

[Slide.]

In those women with older age, the only solution that has proved rather effective is to become a recipient of donor eggs. This is an example of a study actually done with this SART CDC dataset over a three-year swath in which you can see that the clinical and ongoing or delivered rate among women using donor eggs stays good and high until the late 40s. So this is a solution for women who are running into difficulties with their own eggs, not necessarily one that appeals to everyone but it does work and it does tell us that the problem of being 45 or 46 is not the uterus or the body, in general. It is the eggs, themselves,

because, with substitution therapy, that probably disappears.

[Slide.]

I also mention that in the United States, we have had better success than in Europe with our therapies. Again, the last year for which we have got comparable data is '98. In that year, IVF worked 10 percent better here than in Europe, donor egg, 9 percent better and frozen cycles, 10 percent better.

There is a cost to that, in part because we, in the States, are typically putting back an extra embryo or so and the cost is reflected here in the multiple-pregnancy rates. In the U.S., as in Europe, most pregnancies are still singletons and most of the rest are certainly just twins. But, in that year, we had a 6 percent triplet rate and a 0.2 percent quadruplet rate.

[Slide.]

So that is kind of where we have come in the States with success improving, getting a handle on the problem of multiples, newer therapies and simpler therapies. Now, I wanted to kind of move our focus to considering the ovarian stimulation component.

Obviously, IVF is a treatment process for which there are many, many influences in terms of ultimate success rate. Ovarian stimulation is one of the things we have to do to make in vitro work but it isn't the only thing that influences whether or not in vitro works.

For the purposes of this meeting, I think we want to focus on the ovarian response, the thing that the gonadotropins can influence. For the next several slides, that is what I am going to talk about.

We need to understand, really, before going any further, that most women, given the right gonadotropin stimulus, can be made to produce multiple eggs in a cycle. But there is quite a bit of individual variation here. Some women can make a lot of eggs. Some women can make very, very few eggs. Those differences we consider to be differences in ovarian reserve.

Since there is a wide range here, we have to adjust to that range in terms of how we would manage the stimulation, and you will see that as the slides play out. All of this has implications for how one might assess the efficacy of gonadotropins which, again, is the point of this.

Let me use this as sort of an orienting slide. I think we all know that women are born with all the eggs they will ever have in their life. Over their life span, those eggs are doled out. Now, most of them are never ovulated. Most of them never grow much at all. But, nonetheless, they are lost. The rate of loss over time is fairly constant but it is logarithmic so that things go down by an order of magnitude at each even interval.

Within all of the eggs that are available, a very small proportion of those eggs are available for recruitment at any one point in time. But that percentage that is available is pretty much fixed at all points in one's life and seems to be, as I said, a very, very small number, a thousandth of the percent, perhaps.

What that means is that, since a woman at 25 is apt to have still a very large number of eggs, let's say, 100,000 altogether, whereas a woman ten years later will have a tenth of that number and a woman ten years beyond that, a tenth of that number.

If you apply the expectation that a fixed proportion are recruitable, what you see is that a

woman at 25, given a full stimulus to egg growth, could make 100 eggs whereas you can give a woman at 35 the very same stimulus and get a tenth as many eggs and, ten years later, the very same stimulus, and get one egg.

So we need to understand that. I think that is a very important thing before we decide how studies would be done here to understand that there really is a ceiling within a particular cycle of ovarian reserve. There is a potential number of eggs but, no matter how much drug you give, you can't ever get past that number. You can't make a 45-year-old make 100 eggs. There is no way to do it, given, at least, current technology.

[Slide.]

Now, with that in mind, clinicians try to estimate what kind of an individual is this. Is this an individual who has 100 eggs, potentially, or an individual that just has ten or five or one because then we are going to adjust the stimulation strength to that expectation.

So, if a woman, in fact, could be arrayed along this dimension of how many eggs are potentially available, we are going to take a different approach depending on where she happens

to lie along this dimension. For the purposes of IVF, we generally would say that our goal would be to get 10 to 20 eggs. We don't want 30 or 40 because of the risk of hyperstimulation and because it seems that the eggs, then, aren't even as good as they could be if you had fewer.

So we have a target for the number of eggs that we bring into development, 10 to 20 for IVF, maybe three to six for ovulation induction. To reach these targets, we are going to do things differently according to where we think the individual woman is.

Perhaps the easiest case to understand is this one. If we have a woman who could make 100 eggs, we want to be very, very careful and not overdo it because, to get 100 eggs, will not only make her sick but we won't be able to do a fresh transfer. So, in women with a high ovarian reserve, we are going to use very low dosages of drug.

On the other hand, if we think we have got a woman who has just two or three eggs, max, per cycle, we are going to throw all the drug at her we can because we really believe there is no danger in overdoing it. We might as well get every blasted

egg that there is to get.

There are those in the middle where we take kind of a middle course. Women in the middle can surprise you. They can underdo it and they can overdo it. So, sometimes, we guess wrong about what is going to actually work out. But these predictions of response do certainly influence our choice of therapy.

One of the things that you have to notice here is that there turns out to be an inverse correlation between the dose employed and the response observed. So it is kind of counterintuitive. You are giving hardly any drug here and you are still probably going to get 20 eggs or 30 eggs. You give four times as much medicine over here and you are excited if you get three, again, because of the underlying physiology, that there is a limit in the terms of the number of eggs that a cycle can produce.

Sometimes that limit is so low, we are concerned that we may have to cancel. Sometimes it is so high that we have to be very, very careful not to overdo it.

[Slide.]

Again, clinically, how do we try to get a

handle on that? It tends to be two things that we use clinically to assess what category of reserve do we think this particular woman is. In those with what we think is going to be low reserve, we typically see very, very few follicles if we do an ultrasound. We can hardly see any of these small antral follicles.

When we look at the ratio between FSH and LH in these women, it will typically be high. And when we see this pattern, we think this is probably the actual operating range of number of eggs that we might get and, for that reason, use a very strong stimulation protocol understanding that, even if it works, we are not going to get 20 eggs.

At the other end of the extreme, if we think a woman is apt to be one with many, many eggs, 30 to 80, for instance, in a particular cycle, the ultrasound might show some of those, typically not all of them. But, on the ultrasound, we see a very different pattern of many, many little follicles. Here the LH tends to exceed the FSH and, with that pattern, we are going to be very gentle and hope that we don't overdo it.

Again, our goal is 10 to 20 in this case but sometimes we shoot a little too high and find

ourselves getting 30 or 40 and would, for instance, cancel if we have really guessed wrong and overdone it.

Then the average patient, we see some follicles. We see more typical FSH being around the same level as LH and would use a middle-ground stimulation in hopes of getting that 10 to 20.

[Slide.]

A consequence of these sorts of principles is that, yes, dose does directly affect ovarian response. No matter what category of reserve you are talking about, if you go up in the dose within that reserve, yes, you will get more of a response. But the dose you typically would choose to use is inversely correlated to what you judge the reserve to be; namely, in the low responder, you are going to use a lot of drug. In the high responder, you are going to use hardly any.

So, unless, in a study, we control for this, we are going to get it completely backwards.

[Slide.]

Now, the process of IVF is illustrated on this slide and it really revolves around this ovarian stimulation. That is the heart of all of these efforts to get eggs. It typically involves,

for sure, some FSH and either LH as an injection or LH from endogenous contributions to allow us the appropriate level of steroidogenesis to go on.

These other things that are typically done along with the heart of the stimulation are applied to certain kinds of patients in hopes of either better controlling the stimulation here or augmenting it, as you will see in the slides that follow.

[Slide.]

What I have just flashed up there in these little red lines are supposed to be illustrative of the change in the FSH levels that occur when you do these things. So, when you give FSH injections, for instance, the FSH level goes up. That is what the point of it is.

[Slide.]

If you give luteal Lupron, you get this initial little flare in endogenous FSH which lasts a few days but then turns into a suppression of LH below baseline and a suppression of FSH below baseline. That is used in one particular patient type I will discuss in a minute.

If, instead, you use this so-called microdose flare, a very small dose of Lupron, you

get this initial flare but it stays up forever. It really never turns into suppression and so can be used to augment the stimulation throughout.

[Slide.]

Then the antagonist of GnRH shuts off without any kind of a flare effect the FSH and the LH secretion instantly and is typically used simply to obliterate the potential for a premature LH surge.

[Slide.]

So I am going to give you kind of an example of how these are applied to patients of different response types as we would use them in our clinic and I think many clinics would do something like this, if not the same thing, but certainly for the same kinds of reasons.

Among low responders, what we are trying to do is stimulate very hard, not only with the exogenous use of gonadotropins but by getting the body, itself, to contribute extra FSH by using this dilute Lupron or microdose flare approach because we can't really have too much FSH stimulation.

In the average patient, we have now switched to antagonists because it minimizes shots. You often reduce three weeks worth of shots by

making this shift. We still rely on the gonadotropins as the primary vehicle for inducing follicle growth and simply add the antagonist at the tale end once the risk of a surge starts entering into the picture.

And then the high responders are ones that we really want to dampen down quite a bit and would typically put them on pills in the month before and also on Lupron which will downregulate their endogenous secretion so that the only stimulation they see, once they get underway, is that from the gonadotropins. The dose of that gonadotropin is apt to be a low dose as well.

If all goes well, we hopefully will see, on ultrasound, a good number of follicles growing and, upon retrieval, get eggs from most and we would expect the majority, we hope, to be mature and with placement with sperm, or injecting sperm into those mature eggs, we would hope that most of them would fertilize normally.

Among those that fertilize normally, some, but not all, we would expect to divide normally and, amongst those dividing normally, we are going to put back some at the point of transfer. Now, you can see there is a range here. The reason for

the range is that an embryo from someone who is 30 is much more apt to implant than an embryo from someone who is 40 even when they look exactly the same.

So there is a choice point here that hinges a lot on not egg production, per se, but rather on embryo quality which can have almost nothing to do with egg production. If there are extras, they can be frozen. Everyone with me on that point?

[Slide.]

But I alluded to the fact the while we use the gonadotropins to achieve a certain quantitative response, namely 10 to 20 eggs, if it works out well, there are other very important factors relating to quality that come into play typically past the point that we can readily observe.

Let me tell you what I mean. Consider a 32-year-old going through IVF and a 42-year-old going through IVF. Unless this older woman is really in dire straights, we still might be able to get the same number of eggs from her as a 32-year-old if we get our stimulation approaches correct.

So we get a certain number of eggs in both

situations and about 85 percent of them will typically be mature and 70 percent of them will fertilize and 60 percent of them will turn into pretty good embryos three days out, and half of those might go on to blastocysts. Those percentages are apt to be exactly the same whether or not we are talking about an egg that came from someone 32 or 42.

But, beyond the point that we have got these blastocysts, the age matters a lot. In the 32-year-old, putting two blastocysts back will give a very high pregnancy rate and a relatively low miscarriage rate whereas, at 42, the story is very different, low pregnancy rate and high miscarriage rate. Even though we started with the same quantitative material, we ended up with a very different outcome, at least as judged at the point of pregnancy.

[Slide.]

Here is sort of a picture of the same thing. What this shows is potentially two different women. Let's say this is the 32-year-old on the top row and this is the 42-year-old on the bottom row.

Again, odds are we can probably get the

same number of eggs from both women. As the days go along, the behavior of those eggs in the lab may be indistinguishable. A good proportion of the eggs will fertilize. A good proportion of those that fertilize will divide and divide. So we may end up in the very same position come the morning of transfer. But, all along, there has been stuff that is important that we have never been able to observe; namely, which of these eggs are normal.

The younger woman is, statistically speaking, much more likely to have a normal egg than an older woman. Those differences are invisible to us in the laboratory, by and large. But, again, let's, just for argument sake, say this 32-year-old who made these twelve eggs really had three good ones and nine klunkers. Of these three, two fertilized and carried on and turned into two blasts.

If we put these back, she is apt to get pregnancy and she may even have twins. But the 42-year-old may only have had one good egg and it didn't fertilize. It didn't interact with the sperm appropriately and turn into a normal karyotype.

So, even though we saw development, past

that first day, we really had no chance for pregnancy left in that cohort of developing embryos. So, again, it is important to understand that there are really two categories of phenomena going on here. One is quantity and the other is quality.

We can control quantity to some extent by our dosages but we can do nothing to the quality. It is just part of the process that is largely invisible to us at least clinically. Now, Dr. Keefe alluded to some technologies that may be brought to bear in upcoming years such as biopsying and embryo like this and testing its chromosomal constitution.

If you could do such a thing reliably and relatively inexpensively, then you might know that there was really never any chance here and to dissuade a woman like this from trying again. But, at this point, it is not a widely practiced tool.

[Slide.]

The best predictor of the quality of an egg or an embryo seems to be the age of the woman. In in vitro, that quality decline is depicted here out of a study of a number of years ago. Again, it is important to understand that, and I may not have

said it clearly before, that, within the laboratory evaluation process, a 42-year-old embryo typically looks no different than a 32-year-old embryo. So those age differences that are important in terms of pregnancy outcomes are largely invisible to us in the lab. The fertilization rate is the same in older eggs. The development rate is the same amongst older eggs.

But we know that the probability of one of them taking is very, very different according to age.

[Slide.]

So we have got quantitative and qualitative effects that are strong influences on pregnancy rate. The best predictor of the quality seems to be the maternal age. The best predictor of the quantity seems to be the ovarian reserve and how you run your stimulation.

[Slide.]

In the end, you can see, in this slide, that both factors are very, very important. These are, again, some SART CDC data based on a study that has been submitted for publication looking at two-years worth of IVF data in this country broken out not only by age but by an estimate of ovarian

reserve.

[Slide.]

We have already talked about ovarian reserve. It turns out that, for most clinics, the most convenient marker of ovarian reserve is FSH and the higher the FSH is, the worse the reserve is. FSH normally is under 10 always unless you get into some egg problems and, in menopause, it is going to be a 100 or more. So, in these normal women trying to get pregnant, you always hope to find the FSH to be low because then you hope that you are really in a situation with high ovarian reserve.

But, as that FSH goes higher and higher and higher, the reserve goes lower and lower and lower. You can see how important this is. Not only within every level of FSH is age an issue, but within every age, reserve is an issue. Okay? These trends occur even though clinically we are trying our darndest to counteract them.

The ones with high age or high FSH are the ones getting the high stimulations, the very, very strong dosages of gonadotropins and they are also the ones in whom we are putting back every embryo we can get our hands on. And yet, pregnancy rates,

even so, decline.

So these two factors of age and ovarian reserve are really fundamental influences on pregnancy.

[Slide.]

A general strategy clinically; we typically would adjust the stimulation strength to the predicted ovarian reserve in hopes of getting into that sweet spot of 10 to 20 eggs. That is what gonadotropins do for us.

But we also, then, at a later step, a subsequent step, would adjust the number of embryos to be transferred according to our estimate of their quality. If they either are not good looking or are from a woman of advancing age, we are apt to put back more to try to compensate.

But, in a way, these are different components of an overall treatment strategy. We deal with this thing first, how much drug to give in hopes of getting a certain number of eggs and then, subsequently, how many embryos to return.

So I would argue that, in terms of assessing gonadotropin efficacy, we have to understand that both the FSH and the LH are playing critical and complementary roles. I will go into

some of this in a minute. We certainly need FSH to get the follicles growing, the eggs to develop. This is the part that the clinicians are worried about when they pick a protocol and adjust the dose and strength of stimulation.

But FSH doesn't have a whole lot to do with the production of the necessary hormones of estradiol and progesterone. It can only help once LH is in the mix. LH, on the other hand, is necessary for estrogen production. Without appropriate estrogen production and appropriate LH tone, even though you can get follicles to grow, you won't get pregnancies out of it.

So both are needed. Maybe, since both do different things, we need different measures of efficacy for these different hormones.

There has been an ever-increasing understanding, I believe, at least clinically, that LH does have a role to play and that there certainly can be such a thing as too little LH to foster appropriate follicle growth or healthy follicle growth and there may, in fact, be too much as well as is typical for very bad PCO patients.

So, for LH, we are typically, in a clinical setting, trying to get kind of a

permissive dose of LH to allow everything else to run the way it is supposed to and then titrate the FSH dose up and down to get the number of eggs that we would like.

[Slide.]

In considering this a little bit further, I wanted to highlight a few of the things that affect egg production. FSH, obviously, is sort of the driver. But, as I just alluded to, other things matter as well; LH tone during the stimulation. That means a few things. How much of the analogue are you using? Did you use birth-control pills in the prior cycle? Are you using hMG as your stimulating drug or pure FSH and are you using metformin.

All of those clinical tools are affecting LH tone and LH tone, in turn, affects what good comes of the FSH you have been given. The use of hCG as a trigger is also important and all of us in clinical practice have probably had the unfortunate situation where a woman goes through all the trouble of taking stimulatory drugs for ten days but doesn't take her hCG and we get zero eggs.

It also turns out that, at least within our practice, the doc doing the retrieval will

affect how many eggs you get out of that ovary. That is another thing to keep in mind as studies are being designed.

[Slide.]

Let me talk for a moment about this LH phenomenon. I know this is a very busy slide but I think it helps us understand the ways in which both FSH and LH are important to successful treatment. This was a study in which the Ganirelix in a preclinical trial was given at varying doses to figure out how much of this GnRH antagonist was necessary or desirable to achieve orderly folliculogenesis.

These were cycles in which the stimulus was pure FSH and there was no downregulation. So, endogenously, there was a little bit of LH in the mix. The Ganirelix was used to titrate that endogenous LH. With a very little dose of Ganirelix, your endogenous LH was still pretty high, 3.6. But, as your Ganirelix dose as increased ever higher, the effect on endogenous LH secretion was ever stronger to the point where he practically turned it entirely off. Okay?

Again, a fixed dose of FSH in all these cases. Now, because you are no longer making LH at

these high levels, and LH is critical to steroidogenesis, when you got deep into the stimulation, and normally would be making a good level of estradiol, instead, with high doses of Ganirelix, you were making very little estradiol because you had hobbled the system from being able to produce estradiol at high efficiency.

Now, that influence had nothing to do with your ability to, nonetheless, grow eggs. At all of these doses of Ganirelix, the FSH still caused the same number of eggs to grow. And, in fact, the fertilization in the lab and the development of those embryos in the lab was also unaffected. So, for all you knew, it didn't matter that your estrogen was high or low, or your LH was high or how, because you ended up with the same number of eggs and embryos at all of these doses.

But it did matter. It did matter, at least to pregnancy outcome, and that is shown in the graphing part here at the top. Let me first talk about the dark-blue bars, the pregnancy rates. You can see that they are pretty high up to the dose of 0.25 milligrams but low thereafter. You can also see, in the light-blue bars, that the implantation rate or probability of an individual

embryo taking really paralleled the pregnancy rate and was highest with this 0.25 dose.

Lastly, in the red, you can see that miscarriage was a very common event when these very high doses had been used. What does this tell us? I think it tells us a few things. One of them is, again, evidence that FSH is the driver for folliculogenesis. No matter what your LH tone is, you can probably still force egg growth with FSH alone. But, in terms of clinical success, LH has a huge role to play and there really may be a sweet spot where you can have too much but, certainly, too little LH to permit the eggs to be of any use.

[Slide.]

Another factor that can get in the way of seeing clearly an FSH effect is metformin. This is, perhaps, a good thing because we have known for years and years that PCO patients in IVF may make lots of eggs but still are less likely than many others to get pregnant.

So Laurel Stadtmauer did this study a couple of years ago in which a group of PCO patients were randomized to go through standard IVF either without any metformin or with metformin. Metformin is an insulin-sensitizer that has now

been commonly applied to PCO patients and, in many, will allow spontaneous ovulation.

But it also had dramatic effects on what was seen in the IVF setting. First of all, those on metformin had a dampened response to follicle stimulation. Many fewer eggs were seen to be growing. But this reduction had mostly to do with the smallest of the eggs because, when you looked at how many big eggs, big follicles, there were, that difference was much smaller than the overall difference.

At the point of egg retrieval, there was no difference and even more favorable with respect to metformin is the fact that, even though you typically got the same number of eggs with or without metformin, the number of mature eggs was a much higher proportion if metformin had been used and the fertilization rate was higher if metformin had been used and the pregnancy rate was more than twice as high if metformin had been used, again, with the same exact FSH stimulation. So this is a very significant modulator of response as well that we need to track if we are going to be doing a study asking about is this gonadotropin efficacious.

[Slide.]

The last example in this area that I would give is that, at least in our own practice, depending on who the doctor doing the egg retrieval is, you can end up with different egg counts.

This is us over a couple of years. While the number of mature eggs in the end was hardly different and certainly isn't statistically different at all, this guy gets lots of immature eggs that the other doctors probably leave behind. So how would you control for this in a study? Perhaps by focusing on mature eggs rather than eggs altogether or hoping that the randomization will take care of it, that Dr. C won't happen to do a disproportionate amount of the retrievals on your old drug or something.

But these are influences that kind of get in the way of understanding efficacy.

[Slide.]

A couple of other things and then I will be wrapping up. I think it is understandable to everybody in the room that many factors, in fact, affect outcomes here. We do things in terms of stimulation to get eggs. Eggs become embryos. We hope the embryos become pregnancy and pregnancy

becomes a delivery.

But the gonadotropins that are the topic of discussion today here really are tightly linked to eggs and less clearly linked to anything downstream of that event, I believe. Again, the gonadotropin response is affected by the LH tone as driven by metformin use and OC pretreatment and analogues and whether you gave the hCG. If you don't, you are not going to get any eggs. So there are a lot of simultaneous considerations here that have to be accounted for or controlled for, stratified for, whatever in terms of the quantitative character of eggs.

But we also have to remember that there is this parallel quality factor going on, that not all eggs, even though they may look the same, are the same in terms of their potential. Consequently, age enters into the progression here at many points past the egg retrieval. A 42-year-old with the same number of embryos going back as a 32-year-old is not as likely to get pregnant and is substantially more likely to not deliver than that other gal who had the same number of eggs to start with.

And then there are a bunch of doctor

variables in here, too. Is it a good lab? Are the culture conditions up to date? What day did they do the transfer? Was it technically a good transfer? How many were put back? What kind of luteal support was given, et cetera, et cetera.

So, because of all of these sort of downstream events, it can be very hard to clearly link a medication event to one that is very far downstream. And these are some of those other downstream influences that I just alluded to.

Again, going back to this, one might have the same starting point in terms of quantity but a very different ending point in terms of quality. I am unaware, frankly, that gonadotropin use actually drives this difference which would be a pertinent question if it did. But I am unaware that, to get twelve eggs with a certain dose here and twelve eggs with a different dose here is the driver for what is normal and what is not.

[Slide.]

This is embryos that, in many clinics, are put back, eight cells, three days along. But if you wait a couple of more days, half of those that looked pretty good on Day 3 would have not progressed to Day 5 and would, therefore, have been

much more likely to implant and turn into a pregnancy. So, when you put them back matters.

[Slide.]

How normal the cavity is matters, what kind of an evaluation was done of the cavity. If it is just a plain old ultrasound, this big polyp in the cavity may have been overlooked. And, again, the fundamental ones that drive this more than anything else is our ability to estimate ovarian reserve and fertility. They, more than anything, can give you a heads-up about pregnancy events, downstream events, past the point of a certain number of eggs.

[Slide.]

As an example of how difficult these things are to predict, I put this slide up. This is old work but shows that--and Dr. Keefe alluded to this--that, as you go from proximal events in the process such as eggs to downstream events such as delivery, your ability to predict who it is that going to get enough eggs, embryos, pregnancies or be delivered goes down and down and down.

[Slide.]

There is very, very weak predictive ability of either, in a sense, age or FSH about who

is going to be successful; much more predictive ability about eggs. In part, this is because of the difference in the nature of these datapoints. For instance, if you are able to pick an endpoint which can be characterized as means, such as how many eggs you got, as opposed to proportion such as who got pregnant, your sample size requirements are apt to be much, much lower when means are used, even with the same difference to be detected.

[Slide.]

You could take a 30 percent bump in the number of eggs that you were able to retrieve because of a new gonadotropin but if you were looking, instead, for that same 30 percent change in pregnancy rate, if it was even there, you would have to study almost three times as many patients to see it.

[Slide.]

Now, another very important consideration, though, is the fact that this benefit of high ovarian response is often not even something you can see in the fresh cycle. I don't know how to overemphasize this point and I will just describe the slide first and then try to say it a couple of different ways.

In IVF, we typically try to get a lot of eggs. But we are not going to put a lot of embryos back. So we often are in a position where we, in fact, have a few extras that we opt not to transfer. Okay? Again, an old study, but really not different in any of the newer studies, in the old days of IVF, when the pregnancy rates generally were lower, it really didn't matter whether we had a few eggs or a lot of eggs to who got pregnant on that occasion because we were only going to put three of them back anyway, and we had three from here and three from here and three from here.

So far as we could tell, everybody got equally pregnant. The advantage of those extra eggs really came, at least in this era, from the fact that they could be frozen and transferred back in later cycles. So it is clearly better to be in this greater-then-ten category. Clearly, many, many more people ultimately were pregnant.

But, if you had picked as your primary endpoint fresh pregnancy rate, you would make a false conclusion which is that it didn't matter how many eggs you got when, in fact, it does. This phenomenon has really been seen in all of the comparative gonadotropin trials where you might get

an extra egg or two here or one fewer egg here on antagonist.

It makes no difference in the short run because you will always have enough embryos to go back in the short run but it may hold a benefit that appears a few months down the road. So this is another reason to be a little concerned that if the initial pregnancy rate becomes the primary endpoint that you will lose some of the effect that really may be going on with a new kind of a gonadotropin in producing extra eggs or extra eggs of good quality.

So, I think that is all I want to say. Well, let me say this. I guess I would argue that, at least as one of the endpoints, therefore, one might want to look for endpoints that almost are observable before you even get to the retrieval. If the doctor can have an effect on how many eggs enter into the laboratory, that is a problem. Maybe you want to look for things that you can see on ultrasound or measure in the circulation that takes some of the operator dependence out of the picture.

We can talk about what those might be over the next couple of days. I think that's it. So,

again, my take is a little different, I think, than Dr. Keefe's. Certainly, the goal of the treatment called IVF is pregnancy. But the treatment, itself, is multifaceted. There are different components of that treatment.

One of the components is ovulation induction. That component certainly does affect what happens downstream but it isn't the only important thing that affects who is going to get pregnant. There are many other equally important things.

To the extent that gonadotropins, their efficacy is under consideration, it strikes me as most logical to try to link it to what is really happening with eggs being grown rather than those other more distal events which are subject to many other important influences.

Thanks.

Questions from the Committee

DR. GIUDICE: I would like to open this up to the committee. Dr. Hager and then Dr. Crockett.

DR. HAGER: Jim, that's an excellent presentation. Thank you. I have three questions and I will ask them individually and let you respond. You have indicated the benefit of

cryopreservation and freezing embryos. With the current technology and work being done related to cryopreservation of oocytes, can you kind of update us on where that may take us also regarding ethical considerations and so forth?

DR. TONER: Well, it would certainly be nice to be able to freeze eggs. For women who don't have a partner, for women undergoing cancer treatments that may erode their ovarian reserve, it would be a wonderful technology.

It could also be used even among couples who are high-response type in which you may say, okay, well, I am only going to inseminate eight eggs and work with those for this transfer but then freeze everything else. If you don't get pregnant, we will thaw out a few, fertilize a few. But there are ethical things that come these days from the fact that we don't have an effective way of freezing eggs.

You know, divorce happens. Embryos are then stuck between an impossible situation.

DR. HAGER: Do you see a future for that technology?

DR. TONER: Yes; uh-huh.

DR. HAGER: Okay. Second, regarding age

and retrievable eggs, if we could enhance the number of retrievable eggs in an older population of patients based on the information we have regarding aneuploidy, would that truly be a service and where do you see the future going regarding the ability to retrieve better quality eggs in the older patient?

DR. TONER: We don't have an ability nowadays. Nowadays, all we can do is push with our drugs and accept what eggs are developed and hope they are not aneuploid. We could, as a first step, screen them for aneuploidy and transfer only the normal ones. That seems to be a short-term tool.

But, in the long term, obviously, it won't be a good solution. We don't understand what it is that makes an egg recruitable. At this point, we just have to accept it for what it is. But there may be some cocktail of factors that could take the very limited supply of eggs from an older woman and make a higher proportion of them available for recruitment. But we don't know how that would work at this time.

You know, take an ovarian biopsy and inducing these primordials to grow in the lab is something you can dream about but it isn't there

now and it really wouldn't get around the aneuploidy problem. We think the aneuploidy problem is sort of preexisting, absent something like cytoplasmic transfer or nuclear transfer.

There is some question that the aneuploidy may or may not be preexisting. Certainly, if it is already an aneuploid egg, you are probably not going to be able to do anything with it. But it is known that eggs from older women don't have near as many healthy mitochondria so they are kind of underpowered. And it may be that some of the aneuploidies developed in the growth process that weren't really there at the beginning. So if you could give them adequate energy, maybe it wouldn't have gone awry at near the same rate.

That was the thought behind cytoplasmic transfer and nuclear transfer. But that is not permitted. We can't do that work at the current time.

DR. HAGER: Finally, with the rate of multiples decreasing, probably because of limitations on transfer or self-imposed limitations, is that information transmitted to patients in a similar manner, do you think, by all centers. Do you understand what I am saying?

DR. TONER: Is every center working to reduce the number transferred or--

DR. HAGER: No. Is that information being transmitted to the patients based on that risk? Do you think that everyone is making that information available, that risk?

DR. TONER: The risk of multiples within that practice?

DR. HAGER: Not only within that practice but based on the number transferred.

DR. TONER: Yeah; I think so. I don't know that every center has a handout describing it, but most centers that I am aware of, if not a handout, have a discussion on the day of the intended transfer about what happens if we put back three, what happens if we put back two.

DR. HAGER: Is SART monitoring that?

DR. GIUDICE: Dr. Brzyski, do you want to respond?

DR. BRZYSKI: No. Actually, I wanted to ask another question. But I would say that, you know, one of the--speaking about monitoring. Everyone's program reports their multiple pregnancy rate to the national--you know, that is published. Also SART members, when a center undergoes

validation, which is a process whereby there is a random sampling where cycles are examined by the validation committee from SART. This is partly supported by CDC to get a handle on data quality so the records at the Center for that cycle are examined and compared to the data that are submitted to the CDC to test the accuracy.

Part of that visit for SART members includes exploration of practice issues including, again, a random sampling of cycles, were the number of embryos transferred in that cycle consistent with the SART ASRM guidelines for transfer and, if not, was there some documentation of why there was a variation from that practice.

That is one of the criteria that are used to determine ongoing membership in SART. So that is something that is done. Now, each year, about 10 percent of IVF programs are visited by validation committee members. So there is a random sampling of centers, also.

Does that clarify? Now, can I ask my question? We talk about ovarian reserve clinically as something palpable. But I wanted to ask you your comments regarding the ability to determine ovarian reserve in terms of the way that we can

determine--if someone has blocked tubes or no sperm, it is a very black-and-white issue.

But what do you think our positive predictive value or negative predictive value of identifying problems with ovarian reserve is with current technology? Is there a best test? Should we be critical of current technologies in determining that, what is in the future?

DR. TONER: Well, the best test is not one that is used most times and really would be giving everybody a lot of drug and seeing how many eggs they do grow because that is what we are really trying to get a handle on. But it is impractical and expensive and it will make some women sick.

So, instead, we use other things. The basal FSH level on Day 3 of the cycle is the most widely used. We actually have also looked at LH because the ratio, as I alluded to, seems to be very predictive of response. We have known for years that women with PCO have an LH, an exaggerated level of LH, with respect to their FSH. The people with low reserve are the opposite.

Another convenient metric for most of us is simply the ultrasound. If we take a look at a gal early in a cycle, how many of those little

follicles we see is also highly predictive of how many eggs she can be made to grow. So those are the primary clinical tools, I believe.

DR. GIUDICE: Dr. Rice and then we will come to this side of the table.

DR. RICE: I enjoyed your presentation but I do think there are some contradictory statements that I would kind of like for you to help my clarify. On one hand, you say that the number of oocytes retrieved is important because it will impact pregnancy rates, maybe not in that first cycle but in subsequent frozen cycles if there are enough embryos left over.

But then, on the other hand, you recognize that regardless of the number of eggs retrieved for an individual woman based on age, ovarian reserve, et cetera, that there may be only one or two eggs in that cohort that are ever going to lead to a pregnancy.

So, when we look at gonadotropins, do we only assess them for the number of oocytes they produce or should we really be looking downstream, looking at the fertilization rate of those oocytes that are produced by that specific gonadotropin and then, finally, looking at the pregnancy rate

because I think that is the big question of how we should be judged, so when we set up these clinical trials, we are asking the right question of that gonadotropin.

DR. TONER: Yeah; I think that is why we are all here. My own view is that the thing the gonadotropins do is induce follicles to grow. The potential follicle is kind of on the launching pad, will be permitted to grow if they see enough FSH. But the FSH that is used doesn't determine whether or not they are aneuploid. That is determined by other things, age being a convenient marker, typically.

So, while obviously this is all being done for the purpose of pregnancy, my fear is that if you use pregnancy endpoints, then you would--in terms of evaluating gonadotropin efficacy, you would probably have to restrict the range or at least stratify by those other dimensions such as age and reserve that we know also matter and might predict the aneuploidy piece.

DR. GIUDICE: Going up here. Dr. Crockett, please?

DR. CROCKETT: Thank you. I also enjoyed your presentation. I want to talk a little bit

more about FSH as a determinant of ovarian reserve particularly from the standpoint of when you look at the graphs about FSH declining with women's age increasing, it looks like a nice linear graph. But we know that each woman within that graph is, in essence, an n within herself and that FSH and ovarian reserve can fluctuate within a woman. For instance, when a woman starts to go through menopause, ovaries don't just decline gradually. There are some cycles where they ovulate and some cycles when they don't.

It is also my understanding that there can be sort of transient ovarian failure in even younger women where it appears like they are not ovulating or their FSH may be elevated for a period of time and then it goes back up to a normal level and they are able to conceive.

Long way of asking a question, but my question to you is how do you take these into account in dosing the gonadotropins or should be we taking that into account when we look at the safety and efficacy of these medications?

DR. TONER: I think most of the clinicians in the room would agree that there is some month-to-month difference in ovarian responsiveness

and in measures of ovarian reserve. You might have a high FSH this month not followed by a high FSH next month. This cycle, with a certain dose of stimulation, you might get ten eggs. Next month, same stimulation, you might get 14 eggs. So there is a bit of difficulty in this sort of one-to-one mapping. Every cycle isn't the same.

But, at the same time, I think you would get argument from a lot of the clinicians that people move wildly from one category to another. A woman who is given a strong stimulus one month and makes four eggs is never going to make 20 eggs, never, no matter what her FSH is next month or next year.

At the same time, a woman who makes 40 eggs is never going to make two within that year unless you hardly give her any FSH. So, in terms of the full range of ovarian reserve, people tend to stay where they are. Over time, they tend to run down hill.

The problem of fluctuating predictors--FSH, for instance, can be high one month, low another--has led, again, most clinicians to adopt the view that the highest one you ever had is the real one because that tends to be the best

predictor or responsiveness. It has been done three or four times in different studies.

If you have a high FSH of 14 this time, and you say, well, that is a bad cycle, obviously I will wait until it is low. And you wait until it is low and they still don't make a normal number of eggs. I am not sure that the ultrasound assessment is subject to that level of variance, though. I think the basal antral follicle count that a lot of programs would do would be subject to much less of this noisiness that FSH can have.

DR. GIUDICE: Thank you. Dr. Macones?

DR. MACONES: I think you just partially answered my question. You presented a great slide looking at retrieval rate by physicians suggesting that looking at retrieval rate probably isn't a great idea if we look at these gonadotropins.

Again, you suggest using ultrasound as a primary measure, of course, assuming that ultrasound is reliable and has good intra- and inter-observer reliability. Is that the case? I assume it is, but--

DR. TONER: I think it is until you get to lots of follicles, lots and lots of follicles, because then you will find some clinicians say, I

can't measure 45 follicles. I lost track of the last three anyway. But I think, within the normal operating range of zero to 20, if the question is how many follicles are there and are they bigger than 14, there would be a lot of reliability in that kind of an assessment.

DR. GIUDICE: Dr. Lewis? Please go to the microphone.

DR. LEWIS: Thank you. I enjoyed your presentation. I think you raised a lot of important issues one of which, of course, has to do with aneuploidy screening. Aneuploid eggs, I am sure, are very important but, at this point, we don't have any way of assessing what baseline aneuploidic rates are.

We believe they would increase with age but we just don't know. So, to use that as a measure of how effective a gonadotropin is I don't think would be something we could practically do at this point in time.

I do want to just comment on your slide showing that having some frozen embryos would give us another measure of efficacy of a drug, but I think that is subject to tremendous inter-laboratory variation. We did hear Dr. Keefe

say that if you had healthy embryos, they would be frozen. That is something that is done in a lot of centers and that is going to vary according to what kind of quality to laboratory has, what kind of culture system they are using, whether they are freezing at the one-cell or blastocyst stage. So I don't think that that would be something that would be practically useful.

I wonder if you and, perhaps, people from the FDA could comment on what kind of trial designs are used in European centers where many of these drugs are approved for use in ART cycles as they may not be in the United States. What endpoints do they use there?

Then, finally, just two small comments. If you could also comment on assay variation, LH and FSH. A lot of centers are using different kits which may not give the same exact values. Lastly, co-culture, you commented on as not being done because it is not safe. It is really an aside, but I do think that there haven't been safety concerns raised about autologous co-culture of embryos and there are a few centers that are doing that.

DR. TONER: I am not sure that there is any evidence that it is unsafe. I think the

judgement was that it is not known and, until we know, we won't proceed with the nonautologous.

The European trials; I have not been inside the regulatory systems there. Typically, what we see are the published studies which, as you can imagine, typically report all the endpoints. They are typically comparative trials with blinding and multicenter and that kind of deal. What you typically see in them is similar efficacy, again, at least with respect to fresh pregnancy rates because everyone is getting the same number of eggs back, embryos back, even if they got very different numbers of eggs initially and that you would see, for instance, that antagonist trials have typically shown two fewer eggs than in the other approach and one fewer embryo than the other approach. But it didn't matter to the pregnancy rate in the short run.

The frozen rates, I wasn't arguing as something that you probably should include for the reasons you alluded to plus the reasons of years. I mean, when are they going to get around to getting them back--you never get an answer--but, rather to show that the differences that may, in the long run, be meaningful to the patient are

inapparent in the short run, if you use pregnancy as your endpoint.

DR. GIUDICE: Dr. Slaughter, will you be addressing any of the clinical-trial endpoints in your talk today?

DR. SLAUGHTER: I will be addressing just some of the applications that have come through to the FDA and the actual endpoints that were used in that trial, in those trials.

DR. GIUDICE: Okay. Thank you. Dr. Layman?

DR. LAYMAN: I had a comment on the ovarian reserve. According to John Collins, at least, who has done a lot of the work on the positive predictive value, at least what I heard him say six months ago was that Cycle Day 2 or Day 3 FSH was the best predictor, the estradiol wasn't good and HIPN-B wasn't good and I can't remember if the clomiphene challenge was as good as FSH, but I kind of think it was either not quite as good or, because it is less involved, was preferred.

But the other thing to remember, of course, that really the only part that is predictive is that, if the FSH is high, you predict a poor outcome. For some of the members of the

audience, if the FSH is normal, that certainly doesn't guarantee a good stimulation. It is only that if the FSH is elevated, there is a high predictive value with a poor response.

DR. GIUDICE: Thank you for your comments. Dr. Lipshultz?

DR. LIPSHULTZ: As the only urologist here, I think I kind of keyed in on your statement that with ICSI, if there are ten sperm available, then, basically, there is no male factor. That, unfortunately, is a general consensus and, unfortunately, it is not true.

A man who produces ten sperm, obviously, has a disease and deserves the same amount of evaluation as the female. The question becomes one of, given these patients who do have a disease process, we know now that these normal-looking sperm have a very high rate of aneuploidy and we are learning each year of the increased genetic problems that these men have that, in fact, probably are one of the reasons why they are not producing sperm.

So, your end result, then, will be embryos with increased abnormalities and, perhaps, children with increased genetic problems which goes back to

the question about how to look at this outcome in terms of ICSI and IVF.

DR. TONER: Yeah, I grant you that I oversimplified the system. I was trying to make the point that, before IVF, before ICSI, was available, men with few sperm were not well helped by IVF at all but that ICSI has at least permitted fertilization and pregnancy to be established.

While there are some concerns, admittedly, that there are higher sex aneuploidy rates and other things with the pregnancy outcomes, themselves, still the biggest study is Van Steerigum's and the rates of problems are not astronomical. They are higher than the background, higher than the reference population, but not "no go" kind of rates, in my opinion.

DR. GIUDICE: First Dr. Emerson and then Dr. Keefe.

DR. EMERSON: I guess my questions are, again, relating to this question of the egg production as being the endpoint. Do we actually have evidence that says that this FSH regimen can't affect the aneuploidy rates in these? Do we have any baseline rates of what that should be and how it has changed by the FSH regimen?

DR. TONER: Not that I am aware of. But I would say that the rates that are observed line up fairly nicely with the incidence of no pregnancy or failed pregnancy in natural populations.

DR. EMERSON: Yes. But we are also addressing the idea with new treatments. This is a problem with any surrogate endpoint is that the validity of the surrogate endpoint is an interaction between the treatment and the disease. So the idea is that, just because you have shown that something is a perfectly good surrogate in one treatment does not mean it would necessarily be in another treatment and that we may have one regime that doesn't cause aneuploidy and another one does and that would be the pregnancy rate.

DR. TONER: Absolutely.

DR. EMERSON: I will note that, of course, your evidence that you are saying that we need to consider the effects of cryopreservation is, in fact, based on the idea that it affects the pregnancy rate, not the number of eggs produced.

But I guess my other concern, as we are trying to talk in generalities here, is where there are differences in the FSH-LH combination and that we are talking about that, and where you have

suggested that the LH level has an impact, and I am presuming--I am not too knowledgeable on the nonstatistical aspects here--but the idea of the uterine environment, the uterine effects, ought to be able to be affected somewhat by the FSH and LH and do we have very much data on how the subsequent, the implantation of cryopreserved, might different from the fresh cycle and whether any element of differences there can be the regimen that goes in before the--for the harvesting of the eggs relative to not having that cycle just before the implantation.

DR. TONER: That model has not been very instructive because two things are changing. I mean, yes, you have a more natural endocrine environment in a frozen cycle but you have also got an embryo that was frozen. So the two things may cancel one another out.

DR. EMERSON: A surrogacy.

DR. TONER: Yeah. Another model that has been done to try to get the same answer involves, for instance, egg donors who might take a few embryos back, themselves, and give others, also fresh, to another woman whose uterus has been prepared in a more natural way.

There, the evidence, in small studies, has been pretty contro--not controversial but there is a group of three or four studies that I know of that showed no benefit of them going back into a more natural cycle and, two, showing a trend favoring the more natural cycle as the better environment.

You can also look at those who happen to make a lot of estrogen and those who happen to make a little and that will modulate the uterine milieu. Again, that doesn't seem to be very instructive because, even though the high levels might, other things being equal, be averse, they typically come from the women who are younger, making more eggs, et cetera. So it is an important question but it has been a tough one to answer.

DR. GIUDICE: Thank you. Dr. Keefe, you had a comment?

DR. KEEFE: Sorry to ask this with my back turned to you but a comment and then a question. The comment is that you kind of circled around the issue of nuclear transfer and cytoplasmic transfer, especially in the context of your introduction where you emphasized the freedom that ART developed in the U.S. and the advantage of that.

I think it is important to put it on the table. There is not a shred of evidence, not a shred of clinical evidence whatsoever, that either of those procedures ever made a single difference in anybody's life. They were totally uncontrolled studies. In the published study in Lancet there were 17 patients in their mid-30s who had an average of something like 18 embryos. These women did not have egg dysfunction the way we are talking about egg dysfunction.

The whole story is completely based on a biologic rationale which is frail, very frail. I mean, the energy theory is completely lacking in evidence. Microtubules are stored with energy. They don't need ATP from the mitochondria. The mitochondria are quiescent. They have no cristae, just almost zero oxygen consumption in eggs at the time of fertilization in humans and mice.

There is no evidence whatsoever that there is a positive ATP driving aneuploidy in any mammalian egg that has been credibly deduced.

Conversely, you can imagine a number of biologic rationales that would make this a very dangerous procedure because it is very clear that mitochondria are involved in apoptosis and killing

bad eggs that are predisposed towards aneuploidy. In recent published studies from--like, Nature Genetics had a paper two months ago that polymorphisms in mitochondria determine intelligence in mice.

I don't know exactly how they determine intelligence in mice. I guess Mickey Mouse would have starred in their intelligence testing, but it is very important in brain development. Most mitochondrial disorders, should they be passed forward through this process, we, and others, have shown small levels of mitochondrial DNA mutations in eggs from infertile women would only appear later.

So there is no way that this could be considered safe. It is germ-line gene therapy. It is very clearly germ-line gene therapy and somebody should stop it, whether it is the federal government, whether it is ourselves as a profession. Since we didn't do it, I think it is very important that somebody did it. Anyway, that is an editorial comment.

DR. TONER: And that may all be true. I agree that it is premature to be doing those things. What I was trying to highlight, though,

really, is the fact that, for reasons that are still mysterious, a 42-year-old egg, although it looks like a 32-year-old doesn't behave like a 32-year-old egg. So all we have now for those women is substitution therapy.

We don't have a way to remedy the egg and those procedures that you allude to were conceived in hopes that that would be a remedy. That's all.

DR. KEEFE: Yeah. I think we agree. The question is should we consider frozen embryos a benefit exclusively or should there also be considered a side effect. Consider there are a quarter of a million embryos that are in freezers in the United States and I think up to 5 to 10 percent of those are abandoned. There is a double-edged sword to the excess embryos.

I agree with you that it is very valuable but, to sit down with a couple who has their twins and now are figuring out what to do with these embryos that are excess is kind of a double-edged--how do you see that in terms of trials?

DR. TONER: Kind of the same way but I think it may be a bridge technology. Again, if we could freeze the eggs, themselves, even among

married couples, that would really be the preferred avenue so as to avoid these conundrums that we find ourselves in.

But I think, again, my point was that if the primary effect of a gonadotropin FSH is egg production and one is stronger than another and produces more eggs but it gets buried in the fact that you are never going to put as many embryos back anyway as to show the benefit, then you wouldn't want to use the fresh pregnancy rate as your only endpoint, or you would miss the benefit.

DR. GIUDICE: Dr. Emmi?

DR. EMMI: I had a couple of questions. My first is you had spoken about ovarian reserve as being the best predictor or how people will respond. And then, later, you talked about PCO and the fact that metformin will actually change the response in a PCO patient and we all know that they tend to be super-responders.

How would you factor that into a protocol since they do tend to respond differently. Most people treat them differently and start them on different levels of gonadotropins.

DR. TONER: You might either require it being used in all people who meet the criteria for

PCO or balance for it, you know, stratify by it, so that you don't end up with one treatment having an disproportionate number of people who happen to get that adjunctive treatment.

DR. EMMI: You used as your main protocol called the antagonist, you said, for the average patient. How many programs in the country do you think are probably--I mean, just average? Do you have any idea how many are using that now?

DR. TONER: I would guess half or more are using it some way or other.

DR. EMMI: But I mean in their average patient population, I guess, is my question.

DR. TONER: Don't know.

DR. EMMI: Okay. Thank you.

DR. GIUDICE: Dr. Stanford and then Dr. Rice.

DR. STANFORD: I guess I would just like to amplify again on the comments of Dr. Rice and Dr. Emerson in that I was a little bit confused by--I thought I heard you say that we can't really affect the egg quality by the gonadotropin stimulus.

But then you had a clear study where the amount of LH clearly did affect ovum quality,

whether the LH was in the optimum range. So, to me, it seems that, at least in that way, gonadotropin stimulation protocol does affect egg quality. So, that raises, in my mind, questions for using just pure number of eggs as an outcome.

DR. TONER: I don't think we know in that Ganirelix trial whether the lower pregnancy, higher miscarriage, was actually an egg effect or an endometrial effect. So it could be that it did nothing at all to the eggs that would have otherwise grown as evidenced by embryo development in the lab and had everything to do with the fact that the endometrial development was deranged.

But I agree that, given the nature of the structure of the study, there is no way to know which it is because it could also be eggs that were adversely affected.

DR. GIUDICE: Dr. Rice?

DR. RICE: I guess this is more of a comment. I am sort of brought back to my early work when we used our mouse model and were looking at PMS and hCG and we were excited about the number of oocytes that we could get and then we got purified FSH and we were still excited about the number of oocytes we could get.

Then we did a study where we looked at FSH and then we looked at FSH as the trigger for ovulation instead of HCGM. We were still excited about the number of oocytes we got but our fertilization rate dropped off significantly. So you can't just stop at the number of oocytes that you get. You have got to look downstream and then account for the other variables that affect the downstream outcome.

But just looking at the number of oocytes is not enough when it looks at differences between gonadotropins. You have got to look at those other variables. Now, the question is how far downstream do we look and then how can we account for those or control for those confounders.

DR. GIUDICE: Did you want to comment, Dr. Toner?

DR. TONER: I am not sure, in my own mind, that it has to be downstream events. I think, if you have the appropriate LH tone and the appropriate embryology support, then you might see a very clear FSH effect holding all those other things constant that has nothing to do with anything but now many eggs you get.

I think, in the studies where we have seen

problems, we have monkeyed with a part of the system that didn't have anything to do with FSH, really, in the first place. If you take away all the LH and then have a bad outcome, it isn't because the FSH was a problem. It is because you mucked up with the LH side of the equation.

So I think it depends what kind of a--what gonadotropin you are trying to examine. If it is FSH, and you put parameters around what else needs to be there for me to ask the question fairly of FSH, that would include adequate LH. It would include metformin for PCO, perhaps. It would include adequate luteal support.

But, with that frame, then I think you could potentially get a very clean answer about what FSH is doing or not doing with endpoints not past egg retrieval.

DR. KEEFE: I have a question.

DR. GIUDICE: Yes; one last question.

DR. KEEFE: Just a comment. It is hard to do these studies looking at the effects of gonadotropins on egg quality in humans but there is a great study by Barry Bavister in the hamster which has kind of a subfertility that approaches that of the human. He took PMSG-stimulated hamster

donors and then he took naturally cycling donors, and they were both fertilized in vivo. He then transferred the zygotes into the horns, the right and left horn, of the hamster and found that the development of the naturally ovulated embryos was severalfold that of--the implantation rate was severalfold that of the PMSG-generated embryos suggesting that there could be some detrimental effects on development.

It is just harder to do those studies in humans without the controls. But there is evidence that, in normal donors undergoing controlled ovarian hyperstimulation for egg donation who are not split-cycle donors, they don't have, themselves, infertility--they are normal unselected donors--that they have about 40 to 50 percent of their embryos that are aneuploid. This comes from Tony Pellicer in Valencia where they have done PGD in donors.

Frank Barnes who goes around the country as the itinerant PGD biopsy--this is a sort of business he has--has found similar experience working with Santimine who has, again, it is a reference lab for PGD--there is a growing evidence that normal donors, unselected, have high rates.

What isn't known is if you just check their eggs on unstimulated cycles, would they approach that level, although Placheau had found about 25 percent of oocytes. So there may well be an effect. It is just hard to get that into a way we can study it carefully.

DR. GIUDICE: Any further questions from the committee? Okay. I want to thank Dr. Toner. We will now adjourn. The committee will be having lunch at the hotel restaurant. For the others, you can also have lunch at the hotel restaurant and also there are restaurant names available from the front desk.

We will reconvene at 1 o'clock. For the committee members, we have just distributed questions from the FDA that we have been asked to review for the afternoon session. Thank you.

[Whereupon, at 11:55 a.m., the proceedings were recessed to be resumed at 1:00 p.m.]

at

A F T E R N O O N P R O C E E D I N G S

[1:00 p.m.]

DR. GIUDICE: We are now starting. I would like to introduce Dr. Shelley Slaughter who is the Medical Officer Team Leader at the Division of Reproductive and Urologic Drug Products. She will be talking on Human Gonadotropins and the Regulatory History.

Dr. Slaughter. Where is Dr. Slaughter? She will talking about that very soon. While we are waiting for Dr. Slaughter to get to the podium, for the committee members, you have a list of questions. We have a session later this afternoon at 3 o'clock, a presentation of questions and committee discussion. We will have 13 questions with various subparts and we have about an hour and a half to go through those.

Some of them are--well, actually, the division has asked that we come up with recommendations for each of these questions. So we have a task ahead.

Dr. Slaughter, we have already introduced you in absentia so welcome to the podium.

Human Gonadotropins--Regulatory History

DR. SLAUGHTER: Thank you.

[Slide.]

This afternoon, I am going to present the FDA's regulatory perspective on clinical trials of human gonadotropin drug products. As Dr. Shames said this morning, the purpose in convening this meeting is for us to get some information from you and some recommendations from you in order to craft a Guidance Document for Industry.

[Slide.]

The Guidance Document for Industry represents the agency's current thinking on a particular subject. It does not create or confer any rights for or on any person and does not operate to bind the FDA or the public. An alternative approach may be used if such approach satisfies the requirements of the applicable statutes, regulations or both.

[Slide.]

First up, I would like to thank Drs. Keefe and Toner for their excellent presentations this morning. I really could stand up here and just throw the questions out, but I will spend just a little bit more time. But thank you, Drs. Keefe and Toner.

[Slide.]

Briefly, I will review some gonadotropin drug products, give an overview of the clinical studies for selected approved gonadotropin drug products and, hopefully, all of the discussions that you have heard today will lead to a very productive discussion by the committee and we all take your suggestions into consideration, as I said, to craft our guidance document.

[Slide.]

There are two types of gonadotropin drug products that are marketed, urinary-derived gonadotropins and recombinant gonadotropins. This is a somewhat busy slide, just gives a picture of those approved products under both of those categories.

[Slide.]

The approved indications are ovulation induction in chronic anovulatory women and some of the labels will actually say ovulation induction and pregnancy in chronic anovulatory women and the second indication, multiple follicle development in ovulatory women for ART.

[Slide.]

The goal.

[Slide.]

In the 30 years since the FDA approved the drug Pergonal, the technology use in the treatment of infertility and the resulting clinical pregnancy rates have improved. I think Dr. Keefe brought this point out very well this morning. With that, it is time for the FDA to reexamine the clinical studies for gonadotropin drug products.

[Slide.]

The purpose of this overview that I will present will not be to reexamine the data for efficacy to make any reassessment of that data but rather to give an historical perspective on the study design, the efficacy surrogate endpoint and analysis, and safety endpoints.

[Slide.]

I probably don't have to have this slide, but I will just read it. "A surrogate endpoint is the laboratory or physical sign that is used in therapeutic trials as a substitute for a clinically meaningful endpoint that is a direct measure of how a patient feels, functions or survives and that is expected to predict the effect of the therapy."

[Slide.]

The first of the gonadotropin drug products to be approved was Pergonal. It was

approved on June 23, 1970 for induction of ovulation. The second indication was added on March 1, 1988 for development of multiple follicles in ovulatory patients participating in an IVF program.

[Slide.]

The data to support the efficacy and safety of Pergonal came from literature reports of IVF data representing the clinical experience with 192 patients at the Jones Institute over the period of 1981 through 1984 and IVF data from Australia and New Zealand from 1979 through 1984.

[Slide.]

The endpoint that was evaluated as this literature was assessed was the mean number of oocytes retrieved at time of laparoscopy.

[Slide.]

Metrodin, a FSH product, urinary FSH product, was approved on September 18, 1986.

[Slide.]

The data to support efficacy and safety for Metrodin was also a literature review of retrospective data from five open-label, noncomparative, clinical studies of ovulation induction in 80 patients. There were observational

reports of ovulation and pregnancy. The medical officer felt that Metrodin should not be approved. However, this issue was taken before an advisory committee in 1985 and the advisory committee determined that, even though the data to support Metrodin was sparse that, indeed, Metrodin should be approved.

[Slide.]

The next drug the I will talk about is Gonal-f. It is actually the first gonadotropin drug product for which the actual data was submitted to the FDA for review. However, we did not see the protocols prior to submission for the NDA, so we actually had no input into the study design and conduct of those trials.

[Slide.]

The efficacy and safety data was from four controlled studies. Two ovulation studies were open-label, active comparator to the drug Metrodin, phase III noninferiority studies in chronic anovulatory women. The primary endpoint was a cumulative ovulation rate, cumulative over three cycles of treatment. It was defined by a serum progesterone level greater than or equal to 10 nanograms per ml.

The IVF indication was supported by two randomized, again open-label, active comparator, phase III noninferiority studies in normal ovulatory women. The primary endpoint was follicles on ultrasound greater than 14 millimeters.

[Slide.]

Both of these trials--this is ovulation induction trials, now. One was conducted in the U.S. and one was conducted in Europe. As you can see, these, as I stated before, are noninferiority comparative trials. Both studies show that Gonal-f was no worse than Metrodin by 16 or 21 percent, so about 20 percent difference. The lower bound of the 95 confidence interval suggests that it was no worse than about 20 percent than Metrodin.

[Slide.]

There were, as I said, two IVF studies. One was conducted in the U.S. and one was conducted in Europe. This goes back to one of the questions that was asked earlier this morning. This was mean number of follicles. There was not much difference in the results between the U.S. and the European study. Again, the Gonal-f was no worse than Metrodin by about three follicles.

[Slide.]

The next drug was Follistim. It came in a little later than Gonal-f but received its regulatory review approximately about the same time. It was approved on September 29, 1997. Again, actual data was submitted for review to the FDA but we were not involved in the protocol in terms of study design endpoints.

[Slide.]

Efficacy and safety to support Follistim was from four controlled studies. One study supported ovulation. It was a single or assessor-blind study, active comparator to the drug Metrodin, phase III, noninferiority trial in chronic anovulatory women.

[Slide.]

The drug Follistim was actually better or superior to its active comparator in terms of ovulation rate with about 85 percent ovulation rate.

[Slide.]

Three randomized assessor-blind, active comparator to either Humegon or Metrodin, phase III noninferiority studies supported the IVF indication in normal anovulatory infertile women. The

endpoint that was evaluated was the mean number of total oocytes retrieved.

[Slide.]

These three studies were all conducted outside of the United States. Most of these were European trials. The largest of the trials I have indicated as IVF Study No. 1 actually showed Follistim to be superior to its active comparator in the total number of oocytes retrieved. The point to bring out on this one is that this was a trial that was, indeed, powered to show a difference in pregnancy rate and it did, indeed, show that.

I would just have you look at the numbers because that is a point that we will be coming back to.

Now, the lower bound of the 95 percent confidence interval, as stated, showed that Follistim was either better or no worse than its comparators by about one oocyte.

Now, at this point, the point I would like you just to take home is that we have been using various surrogate endpoints for pregnancy. Our sponsors and their advisors indicated to us that trials could not be powered to show pregnancy or

would be difficult to power, and you do see what the magnitude of the sample size that would be needed to actually show pregnancy.

We, then, as I said, looked a surrogate endpoints. Our question has always been what difference would be clinically meaningful in these surrogate endpoints; in other words, how do we make the interpretation of the difference between the drug and its active comparator.

[Slide.]

The safety evaluation on all the four examples of drug products that I gave you, the review looks strongly at ovarian-hyperstimulation rate and multiple-birth rate.

[Slide.]

As you have heard, since these selected gonadotropins were approved, IVF technology has been broadened to include adjunct procedures such as donor oocytes and intracytoplasmic injection and there are more IVF clinics available leading to a greater pool of patients for inclusion in studies.

[Slide.]

This brings me to the point in the presentation where we will put the questions up.

DR. GIUDICE: Thank you.

DR. SLAUGHTER: We will actually put these questions back up a little later, but just to have it, I will read through the questions for you.

[Slide.]

Please discuss what enrollment criteria should be used to adequately capture the population to be studied for ovulation induction and for assisted-reproductive-technology programs.

[Slide.]

Should enrollment criteria be stratified by age for ovulation induction and for ART? Should we stratify for use of adjunct procedures such as donor oocyte or ICSI?

[Slide.]

Should our studies be blinded or not and, if blinded, please discuss the merits of blinding the assessor, the patient or both. Discuss the merits of having placebo and/or active control arms in the studies. If an active control is used, discuss how you would define the noninferiority margin.

[Slide.]

Discuss the advantages and disadvantages of single versus multiple treatment cycles. Discuss the advantages and disadvantages of

powering studies to detect a difference in live birth rate or ongoing pregnancy rate.

[Slide.]

If studies cannot be powered to demonstrate differences in live birth rate or ongoing pregnancy rate, please discuss the clinical relevance of the following surrogates; the rate of patients with a presence of a fetal heart beat, gestational sac, positive beta-hCG, ovulation rate or follicular development rate.

[Slide.]

Is an intent-to-treat analysis appropriate for ovulation induction? If you feel that it is not, should cycles by analyzed per patient given hCG? Likewise, for ART, is an intent analysis appropriate? If not, should cycles by analyzed per retrieval or per embryo transfer. Please discuss which safety endpoints should be evaluated.

DR. GIUDICE: Thank you, Dr. Slaughter.

Questions from the Committee

DR. GIUDICE: Does the committee have any particular questions regarding Dr. Slaughter's presentation because we are going to wait to answer these questions or discuss these questions during the question discussion period later in the

afternoon. Yes; Dr. Stanford?

DR. STANFORD: I don't know if you know this stuff off the top of your head, but it would be helpful to me to know the GnRH agonists and antagonists, what were their endpoints when they were approved. What were the endpoints in those studies to approve those products?

DR. SLAUGHTER: The agonists have never been presented to the agency for consideration. The antagonists also looked at oocyte retrieval.

DR. STANFORD: On Repronex, which is another version of hMG, as I understand it, what was the endpoint on that because its indication is different. It lists with an agonist as opposed to the others don't list that.

DR. SLAUGHTER: The history behind that is that Repronex was originally approved as a generic drug. They came back to the agency to ask for a change in route of administration and, at that time, they also presented new trials that utilized GnRH agonist for downregulation.

Also let me say that the trials for Gonal-f, I believe, were not done with agonists. Two of the three trials for IVF for Follistim were done with GnRH agonists.

DR. STANFORD: But they didn't put that in the labeling. So was the endpoint of Repronex the number of oocytes retrieved?

DR. SLAUGHTER: Yes.

DR. STANFORD: Okay.

DR. GIUDICE: Yes?

DR. EMERSON: In the Follistim IVF study, what was the patient population, what was the definition of infertility in the normal ovulatory?

DR. SLAUGHTER: I believe that it included some male factor, I believe mild male factor, and unexplained and tubal factor.

DR. EMERSON: And infertility was defined by one year of--

DR. SLAUGHTER: Yes.

DR. EMERSON: Okay.

DR. GIUDICE: Dr. Hager.

DR. HAGER: Along that same line, is the reason that there is not consistency across the board is because you were not--the agency was not involved in the protocols?

DR. SLAUGHTER: Yes.

DR. HAGER: You were not requested to be involved in the protocols?

DR. SLAUGHTER: We did not receive those

protocols ahead of time to participate in them. Subsequent to that, we looked at the trials for Follistim and made some recommendations about endpoints. Again, our recommendation, and Dr. Bennett is seated in the audience, was to look at take-home baby. We were told that that was not feasible and so we recommended oocyte retrieval and based our clinically meaningful difference on the trials that we had up to that point.

DR. HAGER: So the use of oocyte retrieval was an FDA recommendation.

DR. SLAUGHTER: Yes.

DR. GIUDICE: I have a question following up on Dr. Hager's. Then, I guess for clarification of the advice that you would like from this committee, the immediate short-term goal is for us to answer these questions and so a guidance document would be developed for your interactions with sponsors before they design their clinical trials, when they have information that they are bringing to you? Can you shed some light on this, please.

DR. SLAUGHTER: What happens, usually--usually, at this point, what we have is the sponsors will come to us at a pre-IND stage and

talk to us about the indications or applications that they would ultimately like to bring in. And we have ability to interact with them as they are opening their IND and study design and to make recommendations to them at that point.

They generally do or do not take our suggestions and conduct their studies. They come in at the time that the studies have been completed at phase III and have another discussion with us before the NDA is submitted. There is also--it depends on the stage of development they are with their drug. We can have meetings with them when they are very early at phase I, when they are at phase II or at the phase-III development.

DR. GIUDICE: Thank you. Dr. Lipshultz?

DR. LIPSHULTZ: Just for my own clarification, then, have the companies been coming to you at different stages in terms of where they have already gone with the product or do they all come to you prior to institution of a protocol?

DR. SLAUGHTER: They have been coming at different stages. Sometimes as some of the examples I have shown you here, they never come to us until they are presenting their application. Sometimes, they come, depending on the drug product

and I am just giving a general overview now at a very, very early stage before they ever go into humans, and then work with us all along the process.

Sometimes they come in after they have already conducted some trials and to ask for further advice. When they don't come in at all, these are trials that have been conducted outside the U.S.

DR. LIPSHULTZ: Do you foresee this process changing and that is why this document becomes important?

DR. SLAUGHTER: The trend now is for most sponsors to come in with--to us for advice, now.

DR. LIPSHULTZ: Before they--

DR. SLAUGHTER: Before they submit the application. We still do have some who come in at a later stage, but we now have worked with a number of sponsors who are coming in early during their development and process. For the gonadotropins, they generally come in to discuss a phase III trial. There may be other drugs coming down the line that may come at an earlier stage, but, yes, the purpose is to give guidance to those--to the industry conducing these phase-III trials.

DR. GIUDICE: Yes, Dr. Crockett?

DR. CROCKETT: I just have a question regarding these guidance documents and the past history and what you foresee in the future for them. This is kind of a new thing to me. I am interested in knowing do you have guidance documents in place already for other drugs through this division? Is this kind of a book that is being put together of guidance documents.

Then my second part of the question is, as we have seen through these presentations, as our knowledge of science has grown, our recommendations probably would have changed over the last ten years. So my follow-up question is if we draft a guidance document today, when does this committee get to revise or look at it again and what does the division or the FDA plan to do with this?

DR. SLAUGHTER: Okay. I will try to answer all those questions in order. This is not a new thing. Throughout the agency, we try to keep our recommendations as uniform as possible and, in order to do that, we do draft guidance documents. In our division, we have numerous guidance documents, some that, as you know--some of the hormone-therapy guidance documents are now in draft

on the web.

What we intend to do with this guidance document, we do realize that things change over time which is why I am here today. But what we intend to do with this is take your recommendations, put a draft guidance together. That guidance will be approved internally and then will be put up on the web as a draft guidance for comment from the public.

Our intent is not necessarily to bring that draft back to the committee but, certainly, the committee or anyone else would make comments to that document at the draft stage and we would take into consideration those comments.

DR. GIUDICE: Dr. Hager?

DR. HAGER: I'm sorry; I hate to belabor this but it seems to me that what we are trying to determine is are we going to make a difference. I heard you make a comment earlier that the FDA recommended to pharmaceutical manufacturer X a change in protocol but that was not adhered to. At that point in time, then, basically you wait until the study is submitted without those changes in protocol and then reevaluate the data; is that correct? There is no intervention in between?

DR. SLAUGHTER: Let me try to answer that. I put up these guidance documents are just that. They are recommendations. They are not binding. What we do in order to keep our recommendations uniform is to draft these guidelines, these guidances. If the sponsor chooses not to follow our recommendation, then that becomes an issue that we will look at in terms of whether it affected the outcome and we felt there was any effect on the outcome as we review the application for our regulatory decision.

DR. HAGER: I understand that--I thought there is some intervention process but I guess there is not. There are no regulations--

DR. SLAUGHTER: No. The guidances are not regulations. We would have to do that in order to make them binding. We can only make them as uniform as possible and give uniform advice and have the sponsor voluntarily adhere to those guidance documents.

DR. GIUDICE: Dr. Shames.

DR. SHAMES: The purpose--we cannot, once we make recommendations for trials, compel anybody to do that particular trial. We can only stop a trial based on safety. So people are allowed to do

essentially within certain parameters any trial they want to do. The guidance process is to make everything more efficient for the industry, for us, so that they know beforehand basically what we want. Then we do work with them and try to tweak their trials so that we, all along the process, know what is going to happen.

Therefore, when we get the final information, we have the information we need so it is the most efficient way to get it done. Now, if they choose to do it some other way, it is still possible to get it approved but it is going to be--it is a little riskier, certainly, on their part. So we are trying to get the word out of what is the best way to do it.

Now, in this case, of course, we are not 100 percent sure what is the best way to do these trials so that is why we have convened this advisory committee. Of course, when we don't know what the best way and the company is telling us one thing and--it makes the process less efficient.

So if we can all sort of agree on some general parameters, then we can move forward faster.

DR. GIUDICE: Thank you. Any additional

questions? Yes; Dr. Emerson?

DR. EMERSON: I am going back to these previous trials that have been done, and knowing that you gave us some background information that were chapters out of various gynecology and endocrinology books that gave estimates of the rates of--pregnancy rates in infertile couples after a year, I think. Are those widely agreed-upon rates or are those, I guess they are quoting something 11 percent pregnancy, fecundity, rate after being infertile for twelve months. Is that a widely agreed-upon rate? Thank you.

DR. GIUDICE: Dr. Stanford?

DR. STANFORD: I guess it goes without saying that if there is a certain approach that comes out as recommended in a guidance document, that is basically what you are looking for in the actual approval. I mean, that is basically what you are saying is that if this is the endpoint you are asking for in the guidance document, that is also the primary endpoint you will look at when you are actually approving. That is a fair statement?

DR. SLAUGHTER: Yes.

DR. STANFORD: One other question. This may be a real--there may be no--I am just trying to

understand, I guess, the details of the actual stated indications but I noticed that, for Follistim and Gonal-f, they are both indicated for both ovulation induction and ART but they list it in different order. Is there any significance to that, or is that just random?

DR. SLAUGHTER: That was just how that fell out; yes. Let me just--I had a little bit more on this I left off the questions, one of the things that I left off from these questions. So it will address one of your things.

These are the indications; induction of ovulation or some of the labels do say induction of ovulation and pregnancy or marked follicle development and ART. We would like your comment on the appropriateness of these indications given all the discussions that hopefully you will have on the endpoints and analysis.

Then, the last thing, and I apologize that I left off this at the end, this came up this morning about the SART data that is collected on pregnancy and about pregnancy registries with industry. We, as part of the process, can only recommend that they maintain a pregnancy registry. We would like to have your input on that, should

manufacturers obtain approval for ovulation induction or ART who obtain approval maintain an pregnancy registry. If you do feel that they should, what information should be collected in that registry and at what point in time should the registry be terminated?

DR. GIUDICE: I have a follow-up comment to Dr. Stanford and also Dr. Shames' comments. Whenever a set of guidances are issued, not specifically to the FDA but by an organization or a body, very often they do not accommodate for the--or they do accommodate for "one size fits all." But, with the complexity of ART and ovulation induction, I think it is an important issue that this is another part of our discussion, that while these may be recommendations, clearly there should be some allowance for "one size does not fit all" for thee particular medications and indications.

DR. SHAMES: It says in our guidances that these are only recommendations. In a field such as this, we often are open to adding to the guidance or changing the guidance periodically. We can keep it even as a draft guidance and still keep it public.

And we certainly recognize that there are times when we--there is more than one way of doing things. So that is a well-taken point. Absolutely.

DR. GIUDICE: Thank you. Dr. Rice?

DR. RICE: I am assuming that, when I look at Point 13, we would make a recommendation, if we decided to recommend this based on future products to be approved or is there any room for the current list of products that are approved for these indications to begin to maintain a pregnancy registry or are you only looking to us for guidance for future products that are going to be approved?

DR. SLAUGHTER: I guess I would answer that as saying do you think these are necessary for these drug products and, if you believe they are, then I think we would certainly implicate them first with future products but would have further discussions on ART products that are already approved. Dr. Shames?

DR. SHAMES: If the committee believes it is important to have that, that will add weight to our arguments to drugs that are already approved that they might investigate doing that. We have no way of compelling them to do that but we would be

interested in your opinions about whether they should or shouldn't do it.

DR. GIUDICE: Thank you. I think we can discuss that in more detail later this afternoon.

I would like to move on now to the Open Public Hearing.

Open Public Hearing

DR. GIUDICE: Before starting this session, there is a statement that I have been asked to read, and that is that, "Both the FDA and public believe in a transparent process for information-gathering and decision-making. To ensure such transparency at the Open Public Hearing session of the advisory committee meeting, the FDA believes that it is important to understand the context of an individual's presentation.

"For this reason, the FDA encourages you, the Open Public Hearing speaker, at the beginning of your written or oral statement, to advise the committee of any financial relationship that you may have with any company or any group that is likely to be impacted by the topic of this meeting.

"For example, the financial information may include a company's or a group's payment of your travel, lodging or other expenses in

connection with your attendance at the meeting. Likewise, FDA encourages you at the beginning of your statement to advise the committee if you do not have any such financial relationships.

"If you choose not to address this issue of financial relationships at the beginning of your statement, it would not preclude you from speaking."

So there are three individuals who have requested time during--okay; are the three individuals who have requested time present? I see one hand, two hands. Two hands. Then let's begin with Dr. Kirsch, please. If you would introduce yourself, your affiliation and if you choose to make any comments with regard to the opening statement.

DR. KIRSCH: Thank you. Good afternoon. My name is Robert Kirsch. I am a Director of Regulatory Affairs at Sorono. I would like to make a brief statement to the committee and we very much appreciate your time this afternoon.

For more than 50 years, Sorono has been a global leader in the development of treatments for infertility and has been dedicated to helping couples realize their dreams of parenthood. Sorono

is a company committed to cutting-edge research, high-quality products and patient care.

Our complete portfolio of fertility drugs, including Gonal-f, Cetrotide, Luveris, Ovidrel and Crinone addresses patients' needs at every stage of the reproductive cycle. Three of these products, Gonal-f, Luveris and Ovadrel are gonadotropin products manufactured using recombinant DNA technology.

We appreciate the opportunity to provide comments on the following important issues raised by the FDA for discussion in consideration by the advisory committee. These issues are endpoints used in clinical trials;pregnancy as an endpoint, and clinical pregnancy.

The design of clinical trials intended to support the registration of new products and indications is an important subject both to the sponsors of such clinical trials--i.e., industry--and the FDA. Equally important is to recognize that, although in certain patient populations for which the underlying cause of infertility has been clearly identified or which has already been extensively researched, it may be possible to agree on a single standard endpoint.

However, it is imperative to avoid a "one size fits all" approach to research in the constantly evolving area of infertility and reproductive health. The first issue, endpoints used in clinical trials; for each population and indication studied, the endpoint chosen should reflect the primary pharmacological action of the drug. For indications involving ART, this may be development of multiple follicles which can be measured directly or indirectly as oocytes retrieved.

In ovulation induction protocols, for example, patients defined by the World Health Organization as WHO2, an appropriate primary endpoint may be P4. Conversely, in patients defined as WHO1, P4 is not fully informative as it does not provide full visibility to other critical components of drug effect.

Therefore, for this population, rate of follicular growth, estrogen production and endometrium receptivity are critical measures of drug efficacy and equally important to P4.

The second issue, pregnancy as an endpoint. On the broader subject of pregnancy, this is the ultimate desire of every patient and

her physician but does not necessarily reflect the primary pharmacological action of the drug. For those ovulatory patients whose cause of infertility remains unknown and unexplained, as many as 25 percent of all infertile patients, pregnancy may be an appropriate clinical endpoint.

For patients undergoing ART, there are other pharmacological agents used during the treatment regimen as well as multiple additional confounding factors any of which may impact pregnancy. These additional factors which are beyond the control of any sponsor, must be considered.

It is important to note, however, that information on pregnancy has always been collected and reported during gonadotropin clinical trials.

The third issue, clinical pregnancy; regarding clinical pregnancy as an endpoint, the above considerations remain. Clinical pregnancy is defined be the National Registry, SART, by the presence of an embryonic sac. Patients realize that a pregnancy detected in its early stages is, indeed, a pregnancy even though it may not always proceed to a pregnancy confirmed by ultrasound or even a live birth.

Beta-hCG is universally accepted as the diagnostic test for early pregnancy and the ultimate outcome of any pregnancy does not alter the fact that a pregnancy was established. Furthermore, in studies of our gonadotropin products, there are many instances where a clinical pregnancy has been confirmed by ultrasound with a fetal sac but without heart beat. Given that the majority of these same patients achieve a successful pregnancy outcome resulting in a live birth indicates that the most conservative approach, ultrasound with fetal sac and heart beat, could result in applying an excessively high standard in terms of product registration.

In considering these types of endpoints, one must realize that, in order to ensure that a clinical trial is adequately powered to detect statistical differences between treatment arms, a primary endpoint of pregnancy, clinical or otherwise, will require large numbers of patients to be studied, potentially thousands per study, creating additional hurdles when conducting research in infertility.

This, in fact, would serve as an additional deterrent to research in low-outcome

patient populations and in those patients who suffer from rare conditions or diseases. Clinical development programs will always have constraints that are not present in routine clinical practice. Importantly, one must determine how much is too much to ask of a particular gonadotropin drug product.

Ultimately, all divisions of the U.S. Food and Drug Administration must make decisions on approvability based on the overall benefit-risk profile of the proposed product in relation to treatment of patients and their respective underlying disease. These considerations are never black and white.

Therefore, in making recommendations which may form the basis for a guidance-for-industry document, we encourage that maximum consideration be given to ensure adequate breadth and flexibility in aspects of study design. This will facilitate the introduction of new drugs, treatment of new indications and improvements in the scientific and manufacturing technologies employed in making these therapies available to patients in the most efficient manner possible.

Sorono has been a leader in the field of

research development and commercialization of products used in the treatment of men and women experiencing infertility for more than 50 years and we intend to continue this commitment. We request that the advisory committee consider carefully what recommendations can be made in order to facilitate the clinical development and availability of new and innovative products which are both safe and effective in the most expeditious manner possible for these patients many of whom are currently underserved.

I would like to thank you all again for the opportunity to share Sorono's position with you this afternoon on some of these important issues. We look forward to hearing your recommendations and to receiving FDA's draft guidance for industry.

Thank you very much.

DR. GIUDICE: Thank you. The next speaker is Dr. Kurt Barnhardt.

DR. BARNHARDT: Good afternoon. It is a privilege to be part of this as a public member. My name is Kurt Barnhardt. I am an Assistant Professor of Obstetrics and Gynecology, a reproductive endocrinologist as well as an epidemiologist at the University of Pennsylvania.

I direct the Clinical Research Center in our department in the University of Pennsylvania and, as such, have contacts with many industry supports that sponsor our studies including, off the top of my head, Wyeth-Ayerst, Parke-Davis, Organon and Sorono.

I wanted to talk briefly and add my comments on the idea of study design and outcomes. As I mentioned, in my capacity as doing clinical research, I have designed many of my own studies as well as participated in many industry-sponsored studies and I wanted to speak a little bit about outcomes and study design in that aspect.

We all know, and we have heard some eloquent talks that infertility is a very complex subject and, often, I can have patients, I, in my practice, that I treat for infertility without gonadotropins and, oftentimes, I can treat them with gonadotropins concomitant with many other therapies as well.

As I want to point out, and as you have heard before, there are many, many factors influencing the ultimate success of the treatment and it is not surely just the pharmacologic agent that I choose.

I disagree with some of the verbiage that was used earlier and I think it was more a mistake that biochemical markers are surrogate markers of infertility treatment. I think that what I would like to say is that we would like to have, or, as a methodologist, you want your outcome to be as close to the direct action of the intervention as possible, in this case, drug therapy. The purpose of drug therapy is, again, in this case, for multiple follicle development or having someone ovulate that otherwise normally wouldn't.

After that takes place, there are many factors that influence whether a women is going to get pregnant. There are many other therapies that we use allowing someone to pursue that goal. Some women get insemination. Some don't. Some get ICSI. Some don't.

Age, obviously, as you know, requires a lot as the genetics of the woman, the receptivity of her uterus and such. But the further downstream we move from the use of the pharmacologic agent, the more determinants are going to affect that outcome. That makes it very difficult to design a study based on a downstream outcome.

I mentioned that another reason that it

makes it difficult for these outcomes is, again, the specifics of how we individualize treatment of the woman or the couple with infertility. It, again, varies very much on how we handle gametes, the quality of our laboratory and many other factors that are independent to the woman alone to try to stratify in these outcomes or to try and analyze these potential confounders afterwards, again, makes it very methodologically difficult.

I guess the only analogy, as I was driving down, that I could think of would be if I was designing a new drug for the induction of labor, I would hope that this drug would be judged on its ability to induce labor, not on its C-section rates and not on its perinatal mortality. Of course, those are important aspects of the drug, but they are not the primary endpoints of the drug under study.

Another issue I wanted to bring up was that the population is very specific and a very savvy population and that certain studies are going to be difficult to carry out practically in this population even though they might be a very good study on paper. A randomized, blinded trial, although certainly the Cadillac of studies, can

have some difficulties in patients where high drop-out for not getting the treatment they want or for therapies that don't work is going to very much influence the outcomes and the validity of a study.

The same could be said for the idea of trying crossover studies or the possibility of having multiple sequential cycles. Please recognize that this really is an individual population and the practicality of carrying out such studies should be taken into account.

So the goal, really, was for me to say please maintain some flexibilities in your outcomes. I mean, currently we are talking about gonadotropins where the goal of the therapy is to induce ovulation in those that don't ovulate or to induce multiple ovulation in those that do.

We could argue about what is the best test for ovulation. Currently, probably, the best compromise between reproducibility and invasiveness is a serum progesterone. But, as we change and that science advances, that also might change and we might adopt some other better marker of ovulation of follicular development.

Indeed, as our drugs change, as we go into subpopulations of infertile populations, we might

be talking about improving egg quality rather than egg number. If we are just talking about ovulation as an endpoint, we are going to lose the robustness of that additional information. Indeed, if we are talking about a new fertility drug that, for example--and I just made this up--maintains the miotic spindle, obviously, we can't have ovulation as an endpoint for that kind of drug. So we have to have flexibility on our study design to allow for the specific indication of that drug.

So, obviously, as a clinician, I hope all of my patients get pregnant and I strive to maintain that. But I also know there is a lot more than the pharmacologic agent I choose and the dose that I choose that is going to influence that.

So, of course, I suggest that you collect all pregnancy data as SART is collected and, of course, miscarriage rates, ectopic pregnancy rates and OHHS are important information. They are very valuable information to decide whether a drug is better or equivalent or no worse, and that is not what I am addressing in the study design. But that is information to compare and not, hopefully, the primary endpoint to power a study.

So, thank you very much for the chance to

speak.

DR. GIUDICE: Thank you, Dr. Barnhardt. Are there any additional comments or any additional speakers for the Open Public forum? Yes?

MR. TIPTON: I am Sean Tipton, Director of Public Affairs with the American Society for Reproductive Medicine. Our President Elect, Marian Dameler, was due to be here today and got detained in York, Pennsylvania which, unfortunately--it is better for people who are flying from farther away because they are less likely to get hung up, I think.

The ASRM does have a commercial relationship with a number of players in the pharmaceutical industry who may have an interest in the deliberations of this panel. Primarily, those will take the form of advertisement or sponsorship of ASRM programs.

Thanks for the opportunity to offer our views on ovulation-induction drugs. The ASRM is a professional association of more than 8500 members worldwide. Our members include the leading experts in the field of reproductive endocrinology and infertility many of whom are around the table and in the room today.

As an organization, we have dedicated considerable time and resources to issues surrounding the use of ovulation-induction drugs. We have developed patient-education materials. Our practice committee has issued a number of related reports and in our journal, Fertility and Sterility, there are frequently works on ovulation-induction drugs and we feel uniquely qualified to offer a little insight and input into the topic.

Ovulation-induction drugs are an essential component of the modern treatment of infertility with approximately one-third to one-half of all infertile women who are having ovulation problems. For many patients, ovulation-induction drugs help with their ovulatory problems and, for others, the drugs are used to maximize success of other treatments such as in vitro fertilization.

Ovulation-induction drugs are used with great effect every day by the members of the American Society for Reproductive Medicine. However, like all physicians, our members are constantly seeking new and better ways to treat their patients. Like any medication, ovulation-induction drugs carry some risk, each

specific product carrying its own particular risk, and we would support the approval of new products with lower-risk profiles.

Many of the concerns, such as reports of increased cancer risk following the use of these products, have not stood up to increased scrutiny, but it is increased scrutiny, more research and better data that are essential for these products as they are for any medication.

At present, one of the most serious potential risks from the use of these drugs is multifetal pregnancies. Multifetal pregnancies carry huge risk for the mother and the subsequent children. The ASRM practice committee has issues guidelines limiting the number of embryos to transfer in ART procedures to minimize the risk of multifetal pregnancy. This work was done with our affiliate, the Society for Assisted Reproductive Technology.

In that case, the data were clear that the number of embryos transferred could be reduced without adversely affecting the prospects for a successful outcome. Unfortunately, the data are less clear on steps to recommend to prevent multifetal superovulation pregnancies. Writing in

the January 2003 issue of Fertility and Sterility, the current chair of our practice committee and his immediate predecessor wrote, "Specific guidelines for management of ovulation-induction cycles cannot be offered because neither sufficient evidence nor broad consensus exists. The ASRM practice committee has no valid basis to offer specific recommendations for cycle-cancellation criteria."

In short, the data are not conclusive. Clearly, more data is needed, more research is needed, more and better medications to provide more and better options for the infertile patients in this country are needed.

We look forward to the deliberations of the committee and thank you for the opportunity.

DR. GIUDICE: Thank you. Are there any further speakers or comments? Okay; we have a choice, now, as a committee, either to take a break or to begin discussing one of the thirteen questions. Who would like to take a break? The hands are gradually going up around the table. Let's take a ten-minute break and then reconvene, please.

[Break.]

DR. GIUDICE: Before we begin discussion

of the questions, there have been several committee members who have expressed an interest in finding out what input they may have with regard to the final document. Dr. Shames, if you would make some comments, please, as to what the process is.

DR. SHAMES: Okay. The process--in a general sense, the advisory committee literally advises us on certain issues, general issues, specific issues. Ultimately, however, your advice is advice. It is a recommendation. We have the authority to make the decision or write the guidance or whatever.

What will happen in this situation where we are trying to write a guidance is, once we have finished this and we will look at the transcript--everything you say is being written down. We will have a direct transcript of all this. Our staff will write a guidance. It will be called a draft guidance and will be posted publicly.

You, everybody, anybody and everybody, will have a chance to make suggestions regarding our draft guidance including you. Now, you, as our advisors, have somewhat of a special status in that we can talk to you directly and you could recommend

to us directly. We can speak to you on a more direct basis than other people.

But the pharmaceutical industry will make recommendations. Public-interest groups will make recommendations. And then we will ultimately, if the process goes well, prepare a final guidance and that will go on the web. We may ask you, again, about the final guidance. We can do that. But, ultimately, we are the ones that have the authority to write the final guidance and put it out there.

So, if that is helpful. But, as advisors, you have good access to us. So when the draft guidance goes up, you should feel free to be aggressive in telling us what you think. Okay?

DR. GIUDICE: Thank you for that clarification. Dr. Slaughter, you have a comment?

DR. SLAUGHTER: May I also just say we would like, also, that we will have access to you. So we may seek your input further after this meeting is over. So we would expect that we would have continued interaction on this.

DR. GIUDICE: Thank you.

Presentation of Questions and Committee Discussion

DR. GIUDICE: The first question is a request to discuss what enrollment criteria should

be used to adequately capture the population to be studied for ovulation induction and for ART. This is quite a loaded question. There are clearly two groups, the ovulation-induction group and the ART group, so, perhaps, we can start off with the OI group bearing in mind the WHO1-WHO2 and also the infertility population. So there are really three subgroups within the ovulation-induction group.

So I open this discussion to the other members of the committee, if someone would like to begin to discuss the enrollment criteria for OI, initially. Dr. Keefe?

DR. KEEFE: Rather than answer, I have a question about the role of the FDA in ensuring sort of the precision of a study, the generalizability. I think I alluded to it in my presentation. A few treatments in medicine today are applied to such a narrow band of the American population which is changing over time. But is it the role of the FDA or does the FDA feel responsible for questioning that?

For example, I work in a state where my patients include sheet-metal workers as well as heiresses from Newport, and I see big variations in the way they metabolize drug, the way they--they

may be smoking or not, not telling you, body-mass index, weight. So there are enormous things that are tied, in part, to socioeconomic class, ethnic background and racial background.

Should you be concerned that the studies are coming out looking at investment bankers from Manhattan and how they are going to apply to my patient population? Is there part of your deal or not? Do you care?

DR. SHAMES: That is part of our deal. We do a lot to try to make sure that the clinical-trial results are generalizable to the general population. However, you can imagine that is very, very difficult. And so we do the best we can to make sure that the clinical-trial population reflects the general population or the population that is going to be using the drug. But it is something that we cannot do perfectly.

DR. GIUDICE: Dr. Emerson?

DR. EMERSON: This is just a question to cure my ignorance--well, not cure it, but alleviate it. Ovulation induction, is that term ever used for the idea of induction of menses in amenorrheic women and would this indication be covered by that?

DR. GIUDICE: Other colleagues are

certainly welcome to respond. For women who are not interested in fertility who are anovulatory, there are other ways to either protect the endometrium or to promote monthly withdrawal bleeds, for instance, with oral contraceptives. However, to promote a monthly bleed, one would not use gonadotropins or antiestrogens, for instance.

DR. EMERSON: So it is just that indication would not cover this.

DR. GIUDICE: No. I assume the rest of the committee agrees with this? Thank you.

DR. SLAUGHTER: Also, from the FDA standpoint, we do have indications for treatment of amenorrhea so this would not cover that. This is strictly for individuals seeking to become pregnant.

DR. GIUDICE: Thank you. I hope all these questions have not had the major purpose of trying to avoid answering the first question. But I think it is time. So who would like to begin the discussion? Dr. Emerson?

DR. EMERSON: So, on those grounds, then, it is very, very difficult for me to completely separate all of these questions and answer only one. So I do need to point out that I am going to

be answering this question with an eye towards what I am also going to recommend as endpoints and that a concept of a noninferiority trial would be very, very appropriate in my mind for an endpoint of pregnancy.

What that means is that when we are enrolling patients, we have to make certain that we have a patient population that is comparable to the ones in which we made the judgment that the existing active comparators are efficacious. So it obviously wouldn't do to be comparing a new therapy of ovulation induction in a population of women who are normal ovulatory. It is an idea that we have to apply some standards to say what is the level of infertility when we are trying to do ovulation induction in hypo-ovulatory women.

If we are trying to do ovulation induction or hyperovulation induction in, for instance, egg donors or the normal ovulatory women, then it is important that we understand exactly what that background is and that the active comparator is working in that population, that we aren't ending up just testing the new therapy against something that is really acting as placebo.

So that would be--my most major criterion

is how do you define the population to make certain that your comparison group is really gaining benefit from the treatment that it is on rather than it is just what they would be having otherwise.

DR. GIUDICE: Thank you. That is very well taken. Dr. Rice?

DR. RICE: I agree. I really do think that when you look at the differences between ovulation in patients who are presented for ovulation induction and what input you would use, that is typically different for us than a patient presenting for ART. I think it is very important for us to recognize that the criteria we set for a drug to be able to induce ovulation would probably be appropriate looking at a population who has hypo-ovulation.

Some of our WHO criteria take that into consideration. Again, that population base is typically going to be different than an ART population who generally may--a large percent of those patients are already ovulatory and we are trying to "superovulate" them.

So I think that, when we look at this, we have to be cognizant of what our endpoint is. If

it is a patient who doesn't ovulate, then, clearly, getting her to the point where she ovulates puts her on the same basis as that normally ovulatory person which is different than that person who is presenting for ART or IVF who ovulates but you want to increase your number of options when it comes down to the number of eggs you want to be able to--particularly be able to select that one out that probably is going to go on to be a pregnancy.

DR. GIUDICE: Okay. I think I am hearing basically the same thing. So, for WHO, we can either begin with WHO1, WHO2, or dive right into the infertility normal ovulatory group. Does someone want to begin, because right now these are the criteria for enrollment which is what the first discussion is on. Does someone want to address standard of care in terms of what we do in the office when someone walks in with infertility? Dr. Toner, are you going to rise to the occasion here?

DR. TONER: I guess, for the anovulatory, apparently anovulatory, patients, we would want to know that that, in fact, is the case by measuring basal gonadotropins, by measuring progesterone, and the hypothalamic style patient will have very low FSH and very low LH at any old time you choose to

measure and, in conjunction with a history of irregular absent menses, could probably be diagnostic of that circumstance.

Progestin withdrawal? I guess you could consider it. Absence of bleeding after progestin is another hallmark of that circumstance.

DR. GIUDICE: And after ruling out any organic causes, hypothalamic or pituitary. So that would be more in the exclusion criteria. Yes?

DR. EMERSON: We would also need to talk about the criteria for infertility, how do we document the time period that the couple has been infertile with a way that we do that because, once again, we are dealing with active comparators that have not truly been tried head-to-head against placebo.

So we don't know what those rates are. We are going to have to go on an understanding of what that is and our best guess so the best numbers that I gleaned from the background materials and what we have talked today is that a fertile couple, we would expect a roughly 20 percent fecundity rate and that, after a year of infertility, it is something down to 11 percent.

Yet, we saw roughly 20 percent in the

previously approved trial. So we are talking about roughly a 10 percent difference in the fecundity rate. As you are now going to look ahead trying to guess what difference we would tolerate in an active-comparator trial and still believe that we would have had efficacy against placebo, because that is sort of what a noninferiority or some one-sided equivalence trial is trying to do, is trying to guess what a placebo trial would have done to make certain that, when we go out there and say, this does work for ovulation induction and going on into pregnancy, we would get that.

So, the mix of patients, if it is down to 11 percent after one year infertility, if it is down to, I can't remember the numbers, 4 percent after two years of infertility or so on, that is going to be the criterion that we are using to base this.

So we want to make certain that, as we compare the active comparator, they haven't snuck in a group that would actually have a lower rate or a higher rate, that we are accepting enough worse behavior that is actually taking us down to the level that it is an ineffective treatment.

DR. TONER: Could I address that?

DR. GIUDICE: Dr. Toner.

DR. TONER: As a practical matter, though, if the clinician has a very confident idea that this is an anovulatory patient, we are not comfortable waiting a year, any clock to run. If we make the diagnosis, we want to treat now. So that interval, for this type of patient, isn't going to be appropriate and we won't get anyone enrolled.

DR. GIUDICE: Thank you. Dr. Keefe and then Dr. Layman?

DR. KEEFE: I was going to say the same thing, that we are trying to establish the diagnosis of an anovulatory infertility, the first is the diagnosis of infertility and I agree with Dr. Toner that you are not going to wait a year if somebody is documented to be anovulatory.

But the documentation of anovulation, I think in the spirit of keeping things generic and broad, I would say clinically useful methods of diagnosis of anovulation and they would include periodic progesterone measurement, LH surge detection, presence of amenorrhea and the absence of uterine factor.

As well, as was already mentioned, but

just to include the importance of ruling other causes of anovulation such as hypothyroidism and hyperprolactinemia as well as primary ovarian failure.

DR. GIUDICE: Dr. Layman?

DR. LAYMAN: I was going to reiterate some of that as well. I think you need to exclude hyperprolactinemia in both Group 1 and Group 2. If somebody is amenorrheic and you think they have Group 1, then I think they really need a progesterone withdrawal to show that they are hypoestrogenic so that they don't bleed. Otherwise, it doesn't fit into hypogonadotropic hypogonadism. Gonadotropins can be low or normal in the face of that. Depending on what assay you use, that is going to make a difference whereas, in the other group, I think one thing we have to think about is the prevalence of both disorders. Group 2 is much more common so that I think the guidelines for Group 2 may--for Group 1, you don't want to maybe be quite as stringent as Group 2 or you are just not going to get enough patients.

DR. GIUDICE: Yes?

DR. LEWIS: Also, although it may go counter to what you see in clinical practice, I

think that the patients in either Group 1 or Group 2 or the ART population, for that matter, should pretty much be of relatively normal body-mass index.

We know that extremes of obesity are quite common, or even just garden variety obesity, are quite common in the general population today but I think it is up to the clinician to adjust for that. When you are designing a clinical trial, I think it is useful to have patients just be relatively body-mass-index normal to glean the cleanest information possible from the study.

DR. GIUDICE: Any comments about BMI for WHO2 type patients?

DR. KEEFE: I have a question. I know what you are getting at, like try to reduce the variance and you would be able to better control the study, but it seems to me that that might lose some of the clinical value of the study. Those are precisely the ones that are the toughest, and so you get this beautiful result in the study.

As you move into the clinic, you might bump up against some problems with that.

DR. LEWIS: Well, of course. But, in any clinical trial, it is going to be quite different.

What you see published in the literature and what it boils down to in clinical practice, we all know that. That is the practice of medicine. If you start out with a most difficult population, to design your clinical trial, you may mask an effect that might otherwise be quite useful to the rest of the population. So, I think if you start with a more ideal population and then try to apply it, see how far you can push it one way or the other, you are going to get a better result.

DR. GIUDICE: Dr. Rice?

DR. RICE: I think we know a lot more information about the impact of BMI on ovulation. So that may be something that you could consider for a stratification based on BMI. That is one of the things that I think would be very critical. If you have a product that is coming out looking an anovulatory patients that you want to stratify it, not just on age, et cetera, but you also want to consider BMI because I think that we all have instances where we have seen patients who lose weight who then become ovulatory. That is the only thing that they have really changed.

So I think BMI could be used as a stratification.

DR. KEEFE: I have a comment about the use of the progesterone-withdrawal test as an inclusion criteria. While clearly it is useful to test the integrity of the uterine tract and the vagina, you can have a negative progesterone withdrawal and still have an ovulation. It is just listed as sort of one sentence in Spiroff. But I have found it very frequently, patients that are hyperandrogenic will not withdraw.

So you might have to give them steroid, presumably to decidualize their endometrium from the high androgens. It is not clear exactly why but it is very common that you will see that they will have a negative progesterone withdrawal and yet they are actually hyperestrogenic relatively and they have a completely intact tract.

So, rather than say progesterone withdrawal, maybe a steroid withdrawal which would include the next step where you would probably give them estrogen and then progesterone.

DR. GIUDICE: Dr. Layman and then Dr. Rice.

DR. LAYMAN: I agree that some hyperandrogenic women won't bleed. But I don't see adding estrogen, how that will help, actually. But

I think the way you get around that, and it is difficult in some PCOS women, to tell from hypogonadotropic hypogonadism whereas, if you look at hyperandrogenism, it may help you differentiate since hypogonadotropic patients usually have low testosterone. So that might be a way.

But I agree. It is not black and white for the progesterone withdrawal. And it should be like a good bleed. It shouldn't just be spotting.

DR. GIUDICE: Dr. Rice?

DR. RICE: I would just caution that the ASRP Practice Guideline is probably going to be coming out very soon with the recommendation to remove progesterone withdrawal. So just before--when you look at the draft document, make sure that the guidelines haven't changed and the utilization of progesterone withdrawal.

DR. GIUDICE: Have they substituted something else for it?

DR. RICE: No. Just making a comment.

DR. GIUDICE: Thank you. Dr. Emmi?

DR. EMMI: In light of what is being said about hyperandrogenicity and lack of withdrawal, frequently those patients will have fecund endometriums on ultrasound and that is a good way

of differentiating sometimes. If somebody is hypoestrogenic, they don't have a lot lining as opposed to people who have 20 millimeters of lining on ultrasound. So that may be a point to look at.

DR. GIUDICE: This is an interesting issue, whether or not ultrasound should be part of the enrollment criteria, whether it is for hypothalamic or Type 1 or for Type 2. Then, in looking for age and ovarian reserve, whether that is part of it.

So we can get into the infertility side in just a moment, but I think also we do need to consider the cost perspective of putting ultrasounds in clinical trials which can be quite costly, but certainly is something that can be recommended. So what does the group think in terms of ultrasound used adjunctively for entry criteria?

DR. EMMI: I don't necessarily think it would be necessary in all patients. It depends on what parameter you use to decide is a patient in Group 1 or Group 2. I mean, if you are going to use estrogen withdrawal, progesterone-withdrawal bleed, then, if they don't bleed, you are either going to look for hyperandrogenicity or you are going to look for some other reason why they

haven't withdrawn and, at that point, perhaps, adding it might keep the cost down.

DR. GIUDICE: Dr. Crockett?

DR. CROCKETT: Yes. I just wanted to throw in my two cents about this whole question of the anovulatory versus the infertility patients in the enrollment. I think, as a group, when we are looking at making guidelines, the less specific we make the guidelines, the more applicable they may be down the line.

If we are looking specifically at anovulatory women that have, by definition, whatever definition we come to, anovulation, we shouldn't include things that pertain to other parts of the fertility process in making inclusion criteria. For instance, if we start talking about endometrial lining, that is not necessarily an indicator of ovulatory function or dysfunction. It is more a function of fertility and the ability to implant and carry the pregnancy.

So I would rather us kind of back up and say let's look specifically at ovulatory and anovulatory women for what they are and leave the whole fertility side out of it.

DR. GIUDICE: Okay. Dr. Lewis?

at

DR. LEWIS: Actually, I think the reason, the rationale, for bringing up endometrium was as a reflection of estrogenicity; that is, the endometrium would be thicker if there has been ongoing estrogen production thus separating Type 1 from Type 2. At least, I think that is why it was brought up.

But the only--well, another possible complication of using ultrasound to differentiate between the two is that you can see a multifollicular pattern in both types and so some centers, particularly in Europe, like to use ultrasound to diagnose polycystic-ovary syndrome.

But that appearance can also be seen in Type 1 patients, especially if they are younger. So it is, I think, a little problematic to use ultrasound as a primary criteria. But, obviously, you have to use some sort of criteria to define that the patient has low gonadotropins and low estrogenicity. Maybe you could make it either/or, either a negative progesterone withdrawal or thin endometrium with a low circulating estradiol level. That might be an option.

DR. GIUDICE: Okay. Yes; Dr. Keefe?

DR. KEEFE: This is really halfway through

the first question. Maybe we should just wrap it up and say something like the committee recommends that enrollment criteria include the presence of oligomenorrhea and/or amenorrhea with some evidence of lack of ovulation on the basis of blood tests, progesterone, urinary tests, LH surge or other methods which could include ultrasound.

DR. GIUDICE: Thank you. Also, for Type 2 patients, there was recently a nonconsensus consensus conference in Rotterdam and I think, and you are absolutely right, we are only halfway through the first question, the whole Type 2 story is, again, very, very complicated. For us to put together today guidelines, I think, actually would probably consume the rest of the afternoon.

So I would like to know from the group if, perhaps, looking at the NIH guidelines in terms of definition of PCOS which, then, could be applied to enrollment criteria, if you think that would suffice for the document.

DR. KEEFE: Yes. Sounds good.

DR. GIUDICE: Okay. Then, moving right along; No. 2, should enrollment criteria be stratified by age for--I'm sorry; we didn't finish ART inclusion. Let's go back to that. I thought

we were out of No. 1.

DR. KEEFE: We are only a third of the way through the first question. I forgot.

DR. GIUDICE: No. We have done WHO1 and we have punted on WHO2.

DR. KEEFE: Right.

DR. GIUDICE: And now for ovulation induction for infertility, the enrollment criteria for that. This brings up many, many issues. If you have someone who is 42, you are probably not going to be doing this, or don't want them in a study for looking at a gonadotropin product.

Also, if you have--there is a whole series of things and people practice very differently. The criteria that are normally used when a couple comes in after twelve months, and, Dr. Emerson, this is to address your question about the definition of infertility, I think it is pretty standard that one takes the definition of twelve or more months of unprotected intercourse without a pregnancy, but then the age issue comes into it and whether--if someone comes in at the age of 40, one doesn't usually ask her to go home and come back in a year and then we'll talk about pregnancy and fertility.

So, we do need to discuss the issue of--and this will, also, then, have some relevance for the second question. But, going to the first, for ovulation induction for infertility, I would like for the committee to address what types of enrollment criteria they think are appropriate for the FDA to have in trials. Dr. Layman?

DR. LAYMAN: To start it, I would think you would have to pick some age group that would be agreeable to everybody, and I will throw out 25 to 34, but some age group that is reasonable would be one thing. The other would be maybe potentially eliminate ICSI with a severe male factor, to take ICSI out of it so that you are looking at fresh-cycle, first-cycle, patients.

But I would think other diagnoses would be reasonable to consider. I mean, I don't think male factor would be unreasonable. But if you start putting in ICSI with it, then you are getting in more variables.

DR. KEEFE: I am really uncomfortable with essentially disenfranchising a huge population, a growing population, of the infertile population. I don't think it is appropriate to, a priori, recommend that a certain age group be studied. It

may well be that, in the future, pharmaceutical companies find it to their advantage to target certain drug regimens. I mean, from a financial standpoint, that is where the money is; right? That is where the growth is. That is where our demographics are pushing us.

So you may want to, for a number of reasons, if you are trying to document the efficacy of ovulation, but, if we are looking at the clinical endpoint, I don't think we should box out this growing population right up front and recommend that they be excluded.

I would say that we should tailor the inclusion criteria according to age and that we should include an age-specific diagnosis of infertility, six months, 35 and older, twelve months, younger than 35, which is sort of the standard.

DR. GIUDICE: Dr. Emerson and then Dr. Emmi.

DR. EMERSON: I think this idea that there would be other issues that come up after the therapy and all of the complicating things that will have an effect on the event rate, but it is randomized trial. I don't know of a disease that

doesn't have ancillary treatments following the interventional treatment.

So, while you have to give thought to whether your treatment might cause people to change the ancillary treatments and so on, it is not an issue. This just occurs in every single trial and, by randomized, it ought to be equal if it is not caused by the treatment.

DR. GIUDICE: Dr. Lewis?

DR. LEWIS: I think that, certainly, a clinical definition of one year of infertility is widely accepted and that is fine. But I think there ought to be exceptions for ART if the infertility is due to a tubal factor. If you have occluded fallopian tubes, there is no reason to wait one year.

So I think you should have evidence of a normal uterus to be in the clinical trial and some documentation about the fallopian tubes. Some of the trials, I know, that have been done in Europe, have excluded patients with hydrosalpinges because that can have an impact on the trial and that might be something that is reasonable to exclude.

You could exclude patients with ICSI. But, in the United States, about 60 percent of--at

least 50 percent of IVF cycles include ICSI. Even if the patient doesn't have proven fertilization, a lot of centers will recommend ICSI on at least some of the eggs.

So, rather than exclude all ICSI, you might set some limits on it, ICSI with testicular sperm, perhaps. ICSI in particular cases--maybe Dr. Lipshultz has some input here to give, but I don't think you would want to exclude all ICSI because I think that would limit the population too much.

I agree that you have to have age limits and maybe 34 is a little Draconian, but maybe 38, under 38? And I don't think 25--I think you could go younger than 25. There are some patients that would be reasonable to treat.

DR. GIUDICE: Dr. Emmi and then Dr. Rice.

DR. EMMI: I was going to make one of the points that she made which is that something like 60 percent of the cycles in this country are ICSI at this point so it is difficult to factor them out and why ICSI is being done in those cases.

But I think, with the age matter, that may be something that might be nice to stratify. I think--I mean, if you look at one of the studies

that Dr. Toner was talking about with decreased follicular depletion over time, I think that the cutoff there is about 37 where you see a really rapid decline in the number of eggs in the ovary.

Maybe you could make that, 37 and above--I don't necessarily think that you have to stop at 38. I think you could go to 42 and just look at that group as a separate group.

DR. GIUDICE: Dr. Rice?

DR. RICE: It is difficult for me to set up enrollment criteria if I haven't established what my endpoint is that I am going to evaluate. If I am going to evaluate number of oocytes for a different gonadotropin, then I pretty much know that, regardless of age, I can give probably most women enough medication and get equal number of oocytes. But if I am going to set up fertilization as my endpoint, then I would set my age of enrollment differently if I wanted to say whether Product A was better than Product B if I wanted to be able to answer that question.

So you kind of got to say what your endpoint is. I think Dr. Toner showed us some data that says you can get a 42-year-old to act just like a 32-year-old when we are looking at the

number of oocytes that are retrieved. But it really comes down to the bottom line, a 42-year-old doesn't get to take home a baby as often as a 32-year-old.

So there are other confounders that come in there, but there are some variables that we may not have control over. So I kind of have to know what my endpoint is before I can say what my inclusion criteria are going to be on several of these issues.

DR. GIUDICE: I think this is where we see one size does not fit all because, if you don't know what your goal is, it is very hard to set up the criteria.

The other issue we may want to discuss, and perhaps not belabor the point, but is age really the issue that we want to talk about or is it ovarian reserve measured by other means, either ultrasound, complement of small antral follicles, ovarian volume, Clomid-challenge test, Day-3 FSH and estradiol. There is a whole series of things.

These, again, I think one could put in, and certainly the committee should discuss this, but as we advise the FDA, these are all the things that we deal with clinically that, as clinicians,

we are evaluating the performance of the ovaries in response to these medications.

Dr. Toner?

DR. TONER: I think, in that regard, you might want to set up sort of a stratification where you define upper limit of normal and then you define a borderline range, but then kind of a "no-go" range as well, for the purposes of studies.

So, for instance, on the issue of age, you might say that there is a group that is under 38. Then there is a group that is from 38 to 41, 42. But we are never going to study people past 42 because it is so futile.

On the ovarian-reserve side of the equation, you might say, we are going to consider normal, basal FSH under 10. 10 to 15 is the ambiguous area where we expect some diminished reserve but it may be okay. But we are not going to enroll anyone with an FSH of 16 or better. You know what I am saying? So there may be some advantage to demarcating where things start to get a little tough and where things become hopeless and break them out that way.

DR. KEEFE: Can I comment?

DR. GIUDICE: Dr. Keefe and then Dr.

Crockett.

DR. KEEFE: I don't think the FDA--the advisory committee should be in the business of lowering our sites as a field. I mean, it is clear that there hasn't been a lot of progress in this area, but it is going to be the pharmaceutical companies' problem if they don't account for these. Their bias is going to be already to exclude these patients.

So I don't see the advantage. I mean, if you are trying to design the perfect study to show an effect, yeah; you, yourself, as an individual designing the study, would clearly want to narrow the variance by selecting a very carefully chosen population.

But we, as the advisory committee, I don't think, should constrain the field by blocking that out. Let's say I move forward five years and I have come up with a new way of identifying that subset of 5 or 10 percent of the 42-year-olds who can get pregnant. Oh, but the FDA, though, has this thing where you have to be under 38 or 37. You know, that is not encouraged.

So, while you might want to control for that, I think what we want to do is cast a broad

net for this.

DR. GIUDICE: Dr. Crockett and then Dr. Lewis.

DR. CROCKETT: What he said. The other thing I had to add to that is that I think, once again, we are getting too narrow. What we couldn't do ten years ago, now we are talking about designing studies around today and, in ten more years, we are going to have so much more technology available that, for us to be talking about what we want as stratification for endpoints and inclusion criteria I don't, as a governing body, we should be doing that.

DR. GIUDICE: Thank you. Dr. Lewis?

DR. LEWIS: I don't think we are saying what population needs to be treated for infertility. That is not the mission of the committee or the mission of the FDA. It is simply to talk about what criteria would be useful in determining whether a drug is effective and safe or not. It doesn't mean that we are saying that this patient will never get pregnant. It doesn't have anything to do with that or with saying what the future of the field is going to be. It is simply to try to help the FDA come up with criteria that

they can use with industry to design a study to say whether the drug works or not.

I think if you start with a population where you might reasonably expect to see a better pregnancy rate, where it has been shown that a good pregnancy rate exists now, then you will go further. The trials will be more economical to devise. The drugs will get to market sooner so that we can use them in new populations who are more difficult to treat.

So I think if you look at it that way, it, perhaps, might have some utility.

DR. GIUDICE: Depending on the endpoint.

DR. LEWIS: Yes.

DR. GIUDICE: Dr. Crockett?

DR. CROCKETT: With all due respect, I disagree with you a little bit on this. I think, from a standpoint of a regulatory agency, if we say to a drug company, we have a guidance that you can't enroll somebody over the age of 42 because it is futile, that we are going to severely limit the free market and the development of that research.

I would rather not see that kind of cap put on either for ovarian reserve or age. I think it is too restrictive.

DR. GIUDICE: Dr. Hager and then Dr. Dickey.

DR. HAGER: It would seem to me that what we are saying is not to exclude but to stratify and that way those populations are included but the data are recorded accordingly. So I don't think--I would agree that we would not want to exclude them but we certainly would want to stratify those populations.

DR. GIUDICE: Thank you. Dr. Dickey?

DR. DICKEY: I think David probably said it very well. If we stratify, you can always add new populations on either end or, for that matter, new enrollment groups. What you want to avoid, I think, is comparing apples and oranges. So, by stratifying, you can begin to do that and still push the edges of science as that opportunity comes along.

DR. GIUDICE: Thank you. Dr. Rice?

DR. RICE: I think, though, you capture all of this stratification if you begin to change your endpoint because, right now, you have it where all you have to do is show follicular development. All of them showed that pretty evenly across the board.

If you take that 32-year-old and that 42-year-old and you say, my endpoint has to be fertilization, or it has to be live birth rate, in order to get said whatever indication they are going after, the stratification, then, begins and it really starts to really compare the gonadotropin compared to each other and compared to "placebo."

So I think you capture it. You keep your inclusion broad. And then you capture it by looking at your endpoints.

DR. GIUDICE: Dr. Hager?

DR. HAGER: Could we go to Question 8 and come back? I mean, it keeps coming up.

DR. GIUDICE: Question 8.

DR. SHAMES: Sure.

DR. GIUDICE: I think we have pretty much completed No. 1 because the ART piece, from what I have heard, is very tightly connected to the infertility, ovulation induction and infertility. So we have done No. 1. And we can go to No. 8. We have also done No. 2. Maybe we can just go down some of the questions, here.

Should we stratify for use of adjunct procedures such as donor oocyte or ICSI.

DR. LAYMAN: May I?

DR. GIUDICE: Yes.

DR. LAYMAN: Just in terms of ICSI, I mean ICSI is not one thing. ICSI is many different indications and to have one category of ICSI is going to be asking for trouble because you are not going to get the same rates if you do ICSI in a program who does everybody because they are afraid of missing a cycle and opposed to another one who does testicular extraction because they have a lot of male factor.

So ICSI has to be subdivided or else--you know, when you are collecting data, or else you are going to lose the meaning of the outcome.

DR. GIUDICE: So that would be stratified, also, then.

DR. LAYMAN: Substratified.

DR. GIUDICE: Or substratified; yes.

DR. LAYMAN: But the answer would be yes to No. 3.

DR. GIUDICE: Okay. Let me ask the FDA members if we need a formal vote on any of these. No. 1, we can't vote on. I think there is pretty much consensus, at least from the nods around the table, about criteria should be stratified by age and stratifying use of adjunct procedures, and then

substratifying for ICSI.

Now, this is a very interesting question; should studies be blinded or not and, if blinded, discuss the merits of blinding the assessor, the patient or both. Dr. Emmi?

DR. EMMI: I just wanted to make one comment about Question 2 before we went on. Since, when we discussed ovulation induction, we actually discussed the difference in anovulators whether they were criteria 1 or 2. Should that be for ART as well?

DR. GIUDICE: Well, how about going around the table or at least--how about if you rephrase what you would like for people to either agree or disagree with.

DR. EMMI: Well, what I propose is that there is a difference in response in PCO patients and IVF and differences in LH and whatever your belief is whether it affects the endometrial development or the number of follicles and should they be stratified out as a separate group when you are looking at ART.

DR. GIUDICE: I see nods.

DR. HAGER: I thought, as we were talking about ART, that we were talking about those

divisions anyway, tubal factor, PCO, et cetera. Was that not the consensus?

DR. GIUDICE: I don't think we actually made it clear. But we should make it clear. The indications for ART, certainly, are several. One, with the PCO patient, in particular, is sort of an interesting situation, those who fail, essentially, monofollicular development or have had one or several follicles without a pregnancy who then may progress to ART is a very different population than the woman who walks in at the age of 39 for unexplained infertility where you start gonadotropins and then you go to ART.

But it seems that the categories would stratify out. Or you can put them in. It depends on what the objective of the study is. But we have talked about male factor, anovulators and we didn't really discuss unexplained. But there is not really too much additional information that we would need to discuss unless someone has a burning issue.

DR. KEEFE: One other diagnosis that may bear or may merit stratification would be severe endometriosis because there is growing evidence from split-donor cycles that the donor, if she has

severe endometriosis, confers a reduced probability of success on the recipient. Meta-analyses have, I think, shown that there seems to be a fairly consistently lower pregnancy rate but only in severe endometriosis, as well as a low ovarian reserve. They start to resemble older patients.

DR. GIUDICE: Dr. Brzyski?

DR. BRZYSKI: In unexplained category, are we assuming that, as part of the screening for enrollment, that individuals that have some measure of ovarian reserve, because it is a fairly good likelihood that there could be impaired ovarian reserve in individuals with otherwise unexplained--you know, patent tubes, normal male, ovulatory. There is still a pretty high yield of impaired ovarian reserve in that population.

So we are assuming that, in that unexplained, that those would be screened and identified and stratified by that? Is that a true statement?

DR. GIUDICE: What does the committee think?

DR. TONER: As the advocate of ovarian-reserve screening, I think yes. You honestly just need to do it. I mean, you do a

semen analysis to detect how many sperm are there. You have got no other way to assess how many eggs are there without looking, either in the blood or on ultrasound.

DR. KEEFE: Can I answer? I agree completely with what Jim said. We would think of ovarian reserve as parallel to age. It is sort of like reproductive age. And, as Jim showed, there is an independent predictive value for each separately and, together, they interact. So, I would recommend we treat ovarian reserve like age, similarly. So even with the outcome of ICSI, in pure male factor, you can see a component of ovarian reserve.

The question is how do you measure it because our radioimmunoassays vary from center to center. So you have to have either a central lab or some standardized method. There are also problems of the cycle dependency of it.

There will be new ovarian-reserve markers coming out. Bularian-inhibiting substance has been recommended by Themin as a very useful cycle-independent marker secreted in the pre-antral phase of follicle development. So I think there will be, down the line, additional markers of

ovarian reserve that will have to be considered.

DR. SLAUGHTER: Just a comment. We do typically request these studies when they do these assays to be done in a central lab so that would be included.

DR. GIUDICE: Thank you. Dr. Layman?

DR. LAYMAN: I just have a statistical question to the statisticians here. If we get too many groups, is that going to reduce our power? I mean, that is why I was only saying a certain group. But, you guys have to be the people on that issue.

DR. EMERSON: So I am completely unable to judge any of the criteria that you have just named of what they mean medically, but, again, the overarching principles here are in doing clinical trials it that the enrollment criteria should be as close as they possibly can be to the population that you are eventually going to use this in.

That sometimes means that people that you don't really think the treatment will be all that beneficial in but it is going to be extrapolated to that population, you need to see the safety data. So you sometimes go ahead and include those patients but then say they are not really going to

be in our efficacy population. We are just going to make certain that it is still safe.

The other criterion is when you have an active comparator, what I spoke to at the very first. You need to make certain that it really is an active comparator and it is not placebo as you are setting those margins.

Then, in the stratification question, the reason why we stratify is, first and foremost, to gain precision. But there comes some point that, if you are doing a multicenter trial, you can't do too many levels of stratification. If the trial is relatively large, you don't need to stratify and you can still get the precision by adjusting for it. It sometimes takes an argument to get it through the FDA, but that is what should be being done in those instances.

The other aspect is scientific credibility of the results. If, by rights, by what we mean by statistics, you don't have to stratify on anything. But, if somebody sees that a large imbalance occurred on the trial, they aren't going to trust the trial. So those variables that are the most predictive but don't go to too many cells; just choose those that are.

From what I have heard, mainly, it is this aging aspect, be it ovarian reserve or age as a surrogate for ovarian reserve, and then the various adjunct procedures, the indications for why we are going to that sound like the largest ones.

And then you always have the site-specific issues that you want to do the stratification within site. That is going to be pretty much maxed out. Really, even if you are going to a thousand-patient trial, you don't want to stratify on much more than those things.

That is my impression from what I have been hearing. As suggested, I probably wouldn't stratify on much more than that.

DR. GIUDICE: Okay. Thank you. Perhaps we can now go to Question No. 4 regarding blinding or not. Yes; Dr. Emerson?

DR. EMERSON: Blinding; I am a statistician so you need a larger sample size and you need to blind yourself. But the issue here that I think is greatest is we have already heard aspects of the doctor effect in getting the follicles, that judgment goes into that concept. Obviously, if they know what the treatment is and, well, of course, this treatment produces more

follicles than the other and, therefore, we will sample all of those more follicles that we see here.

That is an issue. Also, the other thing that I worry about is canceling a cycle, to decide not to go with the stimulation. So I can't imagine many things that the patient would--I think, otherwise, we are only talking about fairly objective criteria, so the patient we are not worrying as much about. But it is all the investigator, the clinician, that I would worry about.

DR. GIUDICE: Dr. Emmi?

DR. EMMI: I wanted to know why cancellation is such an issue.

DR. EMERSON: We are back to my "intent to cheat." If I have a really good therapy that I am sure works, but I start seeing a patient that it doesn't look like it works, or that I am not liking the way it goes, well, I will just not try that patient very much more.

DR. GIUDICE: Dr. Rice?

DR. RICE: I think we also have to remember that, particularly if we were looking at embryo quality as an endpoint, that we still don't

have the most objective means for grading embryos. So it is clear that we must have our embryologists who are blinded as well so that we have some objective information that will come out of that if we are going to use embryology or embryo development as a criteria.

DR. GIUDICE: That is a very good point. Other discussion? Yes?

DR. EMMI: I think we are back to it is getting very complicated because not every institution uses the same criteria even for judging embryos. You know, then you would have to restrict, also--if you had somebody who wanted to cancel, you would have to have cancellation criteria. I mean, I just don't know how you can--

DR. RICE: That is the benefit of blinding, that you can take care of some of these subjective issues.

DR. GIUDICE: Dr. Hager and then Dr. Keefe.

DR. HAGER: The only reason not to double-blind the trial, in my opinion, would be if you had a placebo arm because then there would be a distinct difference. If you did not have a placebo arm, would it not be more efficacious to

double-blind rather than single-blind?

DR. GIUDICE: Comments?

DR. EMERSON: I guess my response would be that just purely logistically I think it is easier to maintain a blind from the patient and not the physician than vice versa. So I don't see any reason not to keep the patient blind in that instance, but it is just the assessor that I am most worried about.

DR. KEEFE: Okay. Are we ready to move onto question No. 5? Here we come now to the placebo and active-control arms in the studies. If an active control is used, discuss how you would define the noninferiority margin. Dr. Hager?

DR. HAGER: My feeling is that, from an ethical perspective, it is better to have an active control arm in this population of individuals.

DR. GIUDICE: Other comments? Yes?

DR. TULMAN: Just from the person you are trying to enroll is given a choice, a known choice, between a placebo or the new investigational drug, are you likely to get anybody to consent to be in it? So, from a very real patient enrollment point of view, if you are not offering a treatment that at least is one standard of care versus the new

point, you are going to have a trial that will not get off the ground.

DR. GIUDICE: That is a good point. Dr. Brzyski?

DR. BRZYSKI: To my experience with clinical trials, or not even in clinical trials, I think patients tend to believe in technology and the newest technology and so that would be one reason to blind even in an active control arm, that patients may cancel themselves if they are randomized to the traditional therapy as opposed to the new therapy which everyone believes will be the better therapy. Otherwise, why would we be studying it; right? So, it just again supports the idea of blinding the patient.

DR. GIUDICE: Thank you. One more comment? Yes, Dr. Keefe?

DR. KEEFE: I would also support not using placebo and, instead, using a treated control group. We now have published literature showing the baseline fertility in untreated patients with unexplained infertility and recently a paper came out of Belgium where they looked at the spontaneous pregnancy rate after one failed cycle of ICSI. So these were people shown to have significant male

factor and then, subsequently, there was a very low spontaneous pregnancy rate. Occasionally, people would get pregnant.

So we have pretty good evidence that the untreated group, at least from observational studies, don't do well. So I agree, it is a little bit problematic ethically to deprive treatment through placebo. So control should include treatment.

DR. GIUDICE: Okay. Thank you. No. 6 is, discuss the advantages and disadvantages of single versus multiple treatment cycles. Dr. Emerson?

DR. EMERSON: There was a second part to Question 5.

DR. GIUDICE: Oh, sorry. You're right.

DR. EMERSON: Which is how--using the active control, how do you use a noninferiority margin. The issue here, and I already alluded to this earlier and I am pretty much going to say the same thing but I will try to say it quickly, therefore, is that if we imagine that this 20 percent rate--I mean, these are the criteria to use, whether we use other data to come up with this.

But if we were having a 20 percent

fecundity rate in the previous--in what little data we have from the controlled trials, and if we believe that it should have been 10 percent or less in an infertile population, the idea is how much of a decrease will we accept. Making up a number of roughly a 6 percent decrease or an 8 percent decrease would be allowing for the idea that you might actually have a mix of patients who were a couple years out infertile and so that the 10 percent or 11 percent I was quoting was from the 12-month infertile rate. So that would still be giving you this margin.

So then a noninferiority trial is saying, we will feel confident at the end that, if we declare the new trial noninferior, it is not more than 6 percent noninferior. So we make a statement like that with 95 percent confidence. But something in the 6 percent to 8 percent range would be what I would choose.

The reason not to go all the way down to the boundary is, again, from the patient standpoint. Would you want to go on a clinical trial where you are saying that we are going to tolerate keeping on testing you on this treatment even when it is substantially below that level.

The criteria of how far I regard substantially below is the secondary benefits the new treatment may provide, that there may be less pain involved, less logistical problems, whatever other conditions might be the secondary endpoints of the trial.

DR. GIUDICE: Thank you. Dr. Rice?

DR. RICE: It clearly depends on what you define as those secondary endpoints. If I look at the trial now, and my primary endpoint is ovulation induction and number of oocytes, my secondary endpoint is pregnancy, then I will tolerate a lot of inferiority dependent on the patient population that I choose. So it depends, again, on those clinical endpoints.

DR. GIUDICE: Yes?

DR. EMERSON: There is no question that I was giving, as my example, the pregnancy endpoints.

DR. GIUDICE: Dr. Lewis?

DR. LEWIS: Yeah; I agree that obviously it does depend on what your endpoint is. But also it is going to depend on the patient's age and their diagnosis and whether you are talking about ART or ovulation induction, even a difference in pregnancy rates because you are going to--the variance is wider for some.

DR. GIUDICE: Okay. Dr. Hager?

DR. HAGER: Just a question. So are we recommending built-in alarms in the study so that, if we reach a noninferiority level of a certain percent, then the study will be broken, the code will be broken? Are we recommending that? Is that what I am hearing?

DR. GIUDICE: Dr. Emerson?

DR. EMERSON: That is an issue that can be really separated from--if you decide what your hypotheses are that you want to test, then, building in a sequential monitoring plan so that, as early as possible, you identify whether you have met that, can be addressed separately and almost all IRBs are now demanding that be--FDA and NIH are demanding that. It is my area of research so I would be glad to talk to you about it.

DR. GIUDICE: Perhaps that line of communication to the FDA can be kept open, then. Dr. Toner?

DR. TONER: I was going to ask of Dr. Emerson, I wonder whether--is there any way to standardize or generate a rule for drug companies to understand what would constitute noninferiority based on expected percentages of success? I mean,

even if we stick with your pregnancy rate endpoint and, let's say, nowadays, the pregnancy rate is 50 percent, as the expected, would you still want to subtract the same absolute number of 6 percent or would you allow it to slip down to 20?

DR. EMERSON: Would I want it to stay the same? Absolutely not. Is there any way that we can give some guidance? Well, the International Conference on Harmonization of Clinical Trials has a very long document on exactly this issue.

Every case is different and a lot depends upon what the safety concerns are, what the secondary gains are that you are trying to anticipate. There may well be some people who are willing to accept a lower fecundity rate which might mean going through more implantation procedures if they only have to go through one ovulation procedure.

So, under those considerations, as you design the trial, you have to consider all of the side effects, all of the anticipated adverse experiences, the gain that you might have, as you go through those trials and what--we are back to this issue of are you going to get people to come on the trial. No trial is very efficient if no one

will be on it, and so you don't do that.

So it does have to be modified. The issues have to be broad guidelines but, to the extent that all the data that I have seen is the table in the background material about what the fecundity rates are, the longer you have been out of fertile, and then the one comparator trial that we had from the previously approved thing.

And then you have to go with that. So it will change over time as you get more data.

DR. KEEFE: Can I comment on that?

DR. GIUDICE: Yes; one short comment.

DR. KEEFE: I think the discussion should first focus on those cases in which there is no expected benefit and then could we then do sort of a power analysis with a two-tailed component in which we are looking not just whether--you know, what is the size of the sample we would have to use to find a detrimental effect.

I think, conventionally, with power analysis, you are looking at sort of 20 percent as a reasonable detection rate. Is that done? Could you do it that way where you are essentially--because, otherwise, you are going to be encouraging companies to make very tiny little

studies where there is no difference found and make the assumption of noninferiority.

DR. EMERSON: Well, just to give you an idea because I did just look this up to see what it was. If you went, for instance, with the assuming that a 20 percent rate was the active effect, and you were willing to go down to a 14 percent rate and still call that not sufficiently inferior, that you do that, that would be about a 600 patient, person, study, total, so 300 per arm.

If you took it down to 12 percent, that would be about 150 or 160 per arm. So those are the sorts of numbers, and that is--again, in a noninferiority trial, we don't go with that 80 percent power. We really need 97.5 power. We need to say, if it is below our threshold that we consider acceptable, we want to be 97.5 percent confident that we don't approve this thing that is inferior. So our standards have changed. This is the same thing, if we are using a 95 percent confidence interval to judge this and that is what those numbers are based on.

DR. KEEFE: From the standpoint of the consumer, a lot of these people are paying out of their pockets for this. I don't know if that

factors into it, but, as the patient advocates, they really want to know if something is inferior because this is coming right out of their pockets if not their hearts.

So I would err on the side of caution before new drugs were introduced that this noninferiority has been established.

DR. EMERSON: I think one of the things you will have is we are mostly, right now, talking--or at least I am talking about--maybe I should have clarified this--but I am talking about the phase III studies, the very confirmatory studies. We have got, presumably, some phase II studies that are based on some of these surrogate endpoints to argue we have good reason to be going forward with this. And that is what you go forward with the patients.

Again, it is the clinicians who are involved with that, them knowing what sort of enrollment they can get with the patient, what seems fair for the patient, in their best judgment, and then IRBs and DSMBs take on a role there, as well.

DR. GIUDICE: That brings us right into the sixth question about single versus multiple

treatment cycles. Any comments on that?

DR. KEEFE: I'd like to pick up where I left off in my talk. I proposed that we consider the advantages of a crossover design. I know there are a lot problems with that but it seems everyone agrees there are enormous patient-specific factors at play here. So, while there is clearly a problem and you have got a lot of people dropping out because they only got through the first treatment, especially if you have proper controls in terms of two different treatments being done, a crossover design would begin to touch on some of the enormous variability from patient to patient. What do you think, Scott?

DR. EMERSON: That was an interesting proposal, as I talked earlier. I hadn't thought of that issue beforehand, of doing the crossover study, because I guess I had been thinking far more of some of the other issues. The biggest problem I would see with that is that patients dropping off the study not because they had pregnancy--that, we could handle--but more just the idea that we would have one period of treatment for them and not--but I guess I tend to think that we would gain more power from actually taking multiple treatment

cycles on the same arm or, certainly, from my endpoint of interest which would be more time until we had a pregnancy and taking advantage of--again, some of these treatments that are going to be coming up will be trying to take advantage of cryopreservation, and what will define your cycle, only one fresh cycle or the idea that you might have four cycles from a cryopreservation.

DR. GIUDICE: Dr. Hager, you had a comment?

DR. HAGER: Merely, I was just going to say it would seem to me that, if we do multiple trials, you would need to quantify or in some way stratify for fresh cycle versus frozen cycle.

DR. GIUDICE: Yes. Dr. Crockett?

DR. CROCKETT: The other thing I like about the multiple cycles is, in the single-cycle studies, I think there is more of a temptation to blast the patient with as much as you can to get the highest pregnancy rate. I worry about safety in those patients. I think, if you allow for a multiple-cycle trial, you can be more judicious about your treatment options and maybe even start lower with your gonadotropins and not put the patient at as much risk in an effort to have a

higher success rate in your study.

DR. GIUDICE: We are talking, and I will address this to Dr. Emerson--we are talking here for the phase III, so your dose-finding studies would have been done before then, I assume. Yes?

DR. EMERSON: But I think there are many times that we, as a regimen, were testing the idea of the whole regimen. So it is not--I mean, there are a lot of other ancillary things that are not related to the dose of this drug but the idea of what other ancillary treatments they might be doing or there are a number of implantations that they might do at a particular time.

Again, if it is blinded, it is post-randomization, that is okay. If it is caused by the therapy, we would be doing that. So it is sort of a level of if you were using--I am imagining that if fresh were as good as cryo and that you were more worried about multiple fetal gestations, that it might be reasonable to start out with two and see if that worked and then, the very next time, do more. And your facilitating that if you allow the multiple cycle whereas you aren't allowing it if you do the single cycle.

DR. GIUDICE: Dr. Macones, Dr. Layman?

DR. MACONES: I would just echo that. I think that, in my mind, in this case, I think we really want to try to test what people do in practice. Not being an infertility doc, but, as I see it, if you have a first cycle that fails, you go on to a second cycle. So to just limit it to a first cycle, to me, makes no clinical sense at all.

DR. GIUDICE: That will stir controversy. Dr. Layman?

DR. LAYMAN: I think, again, we have to get back to the statisticians, then, because you are going to have cycle dependency and that is going to have to be controlled for in the study, because if you have one person who has had four cycles and one person who has had one, or would it wash out in the randomization?

DR. EMERSON: It washes out in the randomization. Particularly, again, if you are looking at an endpoint such as time to live birth, or time to pregnancy, viable pregnancy or whatever, that is a concept. People who go through more cycles, well, that is a deleterious effect, that they won't look as good as the arm in which it worked the first time.

But, working on the fourth cycle is better

than never working at all and so it will capture the endpoint you like.

DR. GIUDICE: Dr. Toner and then Dr. Tulman.

DR. TONER: My objection to the possibility of multiple cycles is the lack of generalizability. At least if you are living in a nonmandated state, as I am in, most patients get one try. And that's it because they are paying their own way. To then come in with an offer of four tries for free will induce different clinical behavior on the part of clinicians. They will maybe be more conservative. What you learn that way won't reflect what actually happens.

So I would be very afraid of opening up a different kind of a treatment strategy than actually happens in practice.

DR. GIUDICE: Dr. Tulman?

DR. TULMAN: I just had two comments. One is, by having a crossover, would you, in fact, entice more people to enroll because, on some of the cycles, they might be getting what may or may not be perceived as the better treatment, although it may not be the better treatment.

And the other question I had to respond to

Dr. Toner's last comment/ was that the current practice of allowing one cycle may be based somewhat on the scientific evidence, but if the scientific evidence changes, then that might also change what is allowable for reimbursement.

DR. TONER: I was only making the comment that in Georgia--

DR. TULMAN: Okay; maybe I misread you.

DR. TONER: Most people can't afford a second try. Even if it is to be encouraged on the clinical grounds, they can't afford it. So, to now offer it to a whole bunch of people who, in the real world, would never have gotten past a first try, introduces a different kind of an observed effect.

DR. TULMAN: I guess I was just thinking further in advance in that, if we became better at doing this, and the cost were more allowable, and then the science would advance enough such that it might be that one trial might not do it but two might be good. But if you have the evidence that two is good, then that might eventually, somewhere down the road, become a more reimbursable, or a covered, benefit.

DR. GIUDICE: Dr. Crockett and then Dr.

Rice, Dr. Stanford and Dr. Lewis.

DR. CROCKETT: I really get a little afraid when we start talking about letting finances dictate how we are doing our medical science. I understand completely what you are saying about the practicality about funding for cycles. However, the fact that a patient can't afford three cycles should not dictate how our physicians are treating those patients for infertility.

For instance, you know, our natural fecundability rate is about 20 percent. It takes most normal, healthy, fertile couples at least three cycles to get pregnant, and that is normal and healthy and okay. For us to be looking at a medical technique that pushes that to 50 or 60 percent because that patient can't afford to come back for those three tries, I don't really, from a regulatory standpoint, want to encourage that type of behavior.

So I would rather try to take that economic or that time pressure out of our recommendations for that kind of trial.

DR. GIUDICE: Dr. Rice?

DR. RICE: Dr. Emerson, what does it do for the enrollment numbers if we use multiple

treatment cycles and then you can change the second cycle based on information that you gather from the first. If some of you all remember some of the ovulation-induction trials with PCOS patients, if they didn't respond to a low dose that first month, the second month, you could bump it up, or you could bump it down, depending on what the trial was.

So what does that do to your enrollment numbers?

DR. EMERSON: Well, first, I should say that those numbers that I gave you on sample size were based more on a binomial proportion, just did they get pregnant or not, rather than what I would truly advocate which is time to pregnancy, which would be a more powerful idea.

I am not, therefore, as worried about the multiple--the nongeneralizability because that will fall right out from the analysis. You will see that the time to pregnancy was this many episodes and you would be able to judge that. If one arm had a higher overall rate over a six-month period but the other arm had a higher rate at the first cycle--these are crossing survival curves and you can use statistics that will identify that if you

want.

But whatever it is, that will come out. So this idea of, then, your question of saying, now this idea that we might tailor treatment, that we would end up testing a treatment that allowed some tailoring. Of course, we do this all the time in diabetes. We don't prescribe a single dose of insulin. We give a constantly modifying dose.

So it is possible to test such things if that is the strategy that you wanted to test and as long as it was blinded, it wouldn't hurt anything, that we would be doing the same modifications on both arms.

So it is this idea of using these multiple times, we will be increasing our event rate, particularly if we can switch into something that is measuring the time as well as the probability of it occurring, we will be getting more useful information.

DR. GIUDICE: Dr. Stanford.

DR. STANFORD: I just want to comment first that the single versus multiple cycle advantages and disadvantages, again, depends on your endpoints, not only your endpoints but your adverse outcomes.

For example, just an example, if you had two regimens and one got you pregnant faster in the first cycle, but the other one got you the same pregnancy rates or maybe a little more after three cycles but with a lower multiple-pregnancy rate, I think that would be a very important thing to assess.

So my argument would be towards the total overall pregnancy rate for a course of treatment, whatever you determine that is, because then you can look at things like what is your overall multiple-pregnancy rate and maybe ovarian hyperstimulation, other things that really should be factored in.

DR. GIUDICE: Thank you.

Dr. Lewis? No. Okay.

DR. KEEFE: I have a question.

DR. GIUDICE: Dr. Keefe?

DR. KEEFE: It seems to me that multiple cycles would be most helpful if our endpoint that we were looking at was pregnancy rate, but, we can all agree that that is an extremely expensive way to bring a new drug to market. If we are going to be using other markers, especially noninferiority, it seems like to be able to fund three, four, five

IVF cycles--because what is going to happen is a significant proportion of your patients are not going to get pregnant in either group, even after three and four cycles, particularly as you get into the mid-30s.

So if we are going to use that, that is ideal. That is perfect. That is the gold standard. But, to have that burden, to put that burden, on a new drug entering, it is a little excessive, maybe.

DR. GIUDICE: I think we will probably get into that discussion when we get to No. 8. Last comment on No. 6?

DR. LAYMAN: I have one other quick question. I mean, does it make a difference when you are using multiple cycles? I have always been taught, you know, the patient is the experimental group if you are looking at the efficacy. Are you saying multiple cycles is reasonable to look at the time to event in addition to just efficacy? Is that what I understand? Is that what you are saying?

DR. EMERSON: With the multiple cycles, we would have several choices. One is you could use the cycle as the denominator and just adjust for

the fact that we have dependent observations, that we have measured some people more than one cycle and we would have to account for that in the variability.

You might also have to decide how you want to weight the fecundity rate. When I have five observations on one person and two observations on another, do I want to weight that 5 to 2 or do I want to weight that, that is one person and here is another person, which is more important.

But what I am suggesting is the time, measuring the time until they have a live birth, the fact that it is multiple cycles or not would be dealt with the exact same way. It is No. 9, but on an intent-to-treat analysis, the question is, once we have randomized you to this, just what is the time until you have had the live birth.

For some people, it is going to be at infinity. It just never happens. But we know how to analyze that data and we find that same event, and then there is not this problem because that is the unit that we are interested in, the patient.

DR. LAYMAN: So you are saying the key is you correct for the--which has been the problem of some studies in the past where there are multiple

cycles and there was no correction.

DR. EMERSON: Yes; absolutely.

DR. GIUDICE: Let's take Questions 7 and 8 together because they both have to with powering of the studies. We have been requested to discuss advantages and disadvantages of powering studies to detect a difference in live-birth rate or on ongoing pregnancy rate.

Dr. Emerson, do you want to make a comment here?

DR. EMERSON: I am not sure that the cost of this--of the numbers that I have been bandying about are any greater than there are in a lot of other diseases. So I think the importance of answering the question that is the clinical outcome is not subservient to trying to choose a surrogate outcome that is not answering the question.

In no clinical trials are we just looking for a pharmacological effect of the drug. There are plenty of examples in the history of clinical trials where the putative mechanism of the drug, the drug did it. The antiarrhythmia trials; we could give people drugs and decrease their arrhythmias. Unfortunately, we also killed them.

The point was to find--we are trying to

treat a disease here. The disease is infertility. We aren't trying to just see how many eggs we can have from a particular ovulation. The people who are wanting this treatment goes there. With an infinite amount of money, I would love to study every single step in the whole mechanism that might lead from one particular treatment to a good clinical outcome.

But no one is going to pay for it. Once they find out that it is a good clinical outcome, they are going to say, yeah, maybe you know the mechanism, maybe you don't. But I have got the outcome that I want. So we are at this stage now where there is plenty of reason to suspect that every treatment that we give may not have the same surrogate value, that we might be intervening on the whole endocrine pathway differently with one treatment than another, so the idea, here, is that if we stimulate ovulation or the additional ova that we are stimulating equally as fertile as those that we would get if we weren't stimulating as many. We don't know those answers, so it is the pregnancy rate that will answer it and, going in this sense, since you have the active comparator, that, by all appearances, works fairly well, it is

not really that big of a burden to be able to demonstrate that you are not doing harm on the important clinical outcome.

DR. GIUDICE: I think you have just opened Pandora's box. Dr. Lipshultz and then Dr. Rice and Dr. Stanford.

DR. LAYMAN: I'm sorry, but I have a question on just what you said. So what you are saying, then, is the only endpoint you think is significant is the pregnancy.

DR. EMERSON: No. I think the primary endpoint of these trials--for instance, I can imagine a scenario in which a new treatment is not as efficacious as another active treatment but has other advantages. For instance, there may be situations that I could imagine if you showed me one treatment that had a higher rate of pregnancy in the first cycle, and another treatment that had still an overall cumulative effect that was about the same, but in the first cycle, it wasn't, I can imagine many scenarios in which I would accept that.

What I was, instead, talking about was the primary endpoint and ensuring that we do not have a treatment that is not efficacious against placebo.

I mean, we are trying to estimate that in a noninferiority situation with an active comparator. The idea that we would not power a study so that you could even be sure that it was not doing as well as a placebo, I find not in good science at this point.

You would like to make certain that, as we are going forward and might approve a therapy, that we would be certain that it is working better than placebo. We won't be certain that it is working better than the active comparator, necessarily. But then we have other secondary endpoints that I would certainly look at. It is just the guarantee that we aren't doing worse than doing nothing.

DR. GIUDICE: Dr. Rice?

DR. RICE: I guess my question is more so to the FDA. How do you maintain the equality if you have previously approved products based on one outcome and now, if this committee were to make a recommendation to say now we want to use a different endpoint. Someone new comes to the market, how do you maintain that level of equality?

DR. SHAMES: Well, I think that we are here because we think, in fact, in the past, we have not--let's say this. Sometimes science

changes. Things change and we have to keep up with the science and do what we can--do the best for the public in terms of having the science that we work with mimic the science in the community.

It is not unusual for things to change and we are accused of being unfair. But this is not at all the only circumstance where that happens. This happens in other areas. Sometimes, we find safety issues that we didn't know about two or three years ago and we compel companies to do certain studies that we didn't compel other companies to do. We just find that we have to do that. We need to do the best we can at the moment based on the science, even though that is perhaps not what we did a number of years ago.

DR. RICE: I think that, looking at gonadotropins--I think you do reach a point with a certain class of products that what comes onto the market, there is a level of equality in effectiveness of an endpoint that you used to use but now the science has shown us that there are more variables that go into that ultimate outcome.

I brought this up to hope that it was based on all of the science that is coming out even though our science in understanding ovarian

reserve, et cetera, is still growing, but that the science perhaps does dictate that we look at different endpoints.

DR. SHAMES: Absolutely, it should.

DR. GIUDICE: I would like to ask the committee for some input in terms of what other primary outcomes--does everybody agree, now, that the primary outcome for, for instance, gonadotropin therapy should be pregnancy? We have heard from Dr. Emerson and I think this is where--and I think Dr. Lipshultz's question addressed this. Dr. Stanford?

DR. STANFORD: I would just like to clarify. Are you talking about, Dr. Emerson, chemical pregnancy, clinical pregnancy, or live birth?

DR. EMERSON: Well, I guess my top choice would be live birth. While I liked the suggestion earlier that it should really be live healthy birth, I think that there is trouble with healthy birth. So my first choice would be live birth and, if you have to take me to lesser things, it will be going backwards from live birth.

DR. STANFORD: I would agree with that.

DR. GIUDICE: Dr. Keefe?

DR. KEEFE: If you looked at Dr. Toner's graph at the relationship between miscarriage and take-home baby, it seems to sort of parallel. So I think it is a crude but good indicator of what is going to happen. You could imagine a number of interventions that might selectively perturb the development of the embryo after that, but I think, in general, on balance, it is probably the best sort of practical measure we have short of live birth. This would be clinical pregnancy or with a heartbeat.

DR. GIUDICE: Dr. Toner and then Dr. Lipshultz, and I saw another hand down here. Dr. Hager.

DR. TONER: I would still vote for some measure of follicular response, the primary pharmacologic action of the drug in question. I think, if we choose pregnancy, as I said this morning, if a drug these days proved more potent in egg production, it would be invisible to the fresh pregnancy rate. So an apparent benefit is completely invisible.

Now, sure; it could go the other way. But we clinicians are actually in the business of trying to get pregnancies and not trying to get

eggs despite what I might say otherwise and it wouldn't be too long before an inferior product with respect to the ultimate endpoint would come to light and change practice.

Already, all the studies that have focused on egg-production capacity have disclosed pregnancy rates. So it is not like the data would be unavailable if we stuck with the same endpoint as has been used heretofore.

DR. GIUDICE: Dr. Lipshultz?

DR. LIPSHULTZ: It's really not a question for you because, I mean, I am coming from a specialty where we don't have such a long delay in waiting for an outcome. But if you are a company developing a drug and your outcome is nine months after you start giving the medicine in question, the drug in question, let alone the start-up time and getting your patients enrolled, what are you looking at, realistically, until you get something--data together that is going to be significant enough to present and, in the meantime, science is changing and you are taking three years to do a study, with the numbers you talked about.

I mean, if you just use live birth as an endpoint, I am questioning just how long such a

study would have to take.

DR. EMERSON: In a time-to-event analysis, you actually measure the amount of data you get by the number of events. So the question is how long do you have to go until you have so many events, and you power it according to that. Of course, cancer in clinical trials routinely takes years and years and years. So this isn't anything that is out of the ordinary in the experience of clinical research.

DR. KEEFE: Can I respond to that one?

DR. GIUDICE: Yes; Dr. Keefe?

DR. KEEFE: There is a real practical issue as well. People disappear. They got their babies; like, sayonara, they are out of there. I mean, the person who is treating for infertility is not the one who is going to deliver the baby. So it is really hard just to find these people. They move. They go to different places. They deliver at another hospital. So you can't get the data, even in a really carefully controlled study.

One other point, though, about using the egg number or follicle number, there are precedents in which you get more eggs but there is a worse outcome. Klaus Dietrich, for example, just looking

at a subset, those with lower response, has shown that the flare, the old-fashioned flare, actually gives you more eggs. It gives you higher estradiol response, but a lower pregnancy rate than some of the other methods.

So it is a little tricky just to look at egg and follicle number. So I would say clinical pregnancy is kind of a middle ground.

DR. GIUDICE: Dr. Hager?

DR. HAGER: I understand the pharmaceutical perspective, the research perspective, of power and design in trying to get as many positives, if you will, regarding follicular development, regarding oocytes. At the same time, my patient population, they are not interested in a chemical pregnancy test. Their perspective is that they want to get pregnant.

At the same time, given my understanding that if we have a gestational sac with fetal heart motion, there is a 95 percent chance of carrying that pregnancy. Then, I think that gives that patient the same odds as a patient who conceives outside of an assisted-reproductive-technology situation.

So my perspective is that my endpoint

would be fetal heart motion. I realize that we would all like to have a live baby but I don't think that is the endpoint that I would like to see with these medications. I would like to take it to fetal heart motion.

DR. GIUDICE: Dr. Layman and then Dr. Brzyski.

DR. LAYMAN: I think, to me, it is different depending upon the implication. Like, for ovulation induction, for Class 1 patients versus Class 2 versus using for ART, for ART, it seems more clear to me that it should clearly be pregnancy. But, at the opposite extreme, I think Class 1, I think it is not practical to use pregnancy as the endpoint. You will never get enough patients to do it.

So I think you have to sort of bear in mind how common it is. I think Class 2 is a tougher one. It is sort of in between and I don't view all three of those the same.

DR. GIUDICE: Dr. Brzyski.

DR. BRZYSKI: I think I brought this up earlier but especially in the PCO population, there is a real problem with miscarriage rate and also in infertile populations, not only in advanced but

just the infertile puts you at increased risk of miscarriage even after the documentation of cardiac motion on ultrasound.

So I think you do lose something, and you can imagine an intervention that--I mean, there is some scientific evidence now for an off-label use of a ovulation-induction medicine that is used to induce ovulation in PCOS patients. One of the attractions to that is, because of the lower miscarriage rate in PCOS patients. So that becomes an important endpoint that you might lose if you focus on cardiac motion or positive pregnancy test.

DR. GIUDICE: That does bring us into Question 8, but I think Dr. Rice had a comment.

DR. RICE: I was just going to say, I agree with Dr. Layman. I think it really does depend on the type of patient you are looking at. If I have an anovulatory patient, then I may have a trial that really just looks at getting that patient to ovulate as the endpoint. In that patient population, that would be adequate because of what the patient came to the table with.

But when you are looking at ART patients, I think that it should be fetal-cardiac activity on a ultrasound because, again, as Dr. Hager was

saying, there is 95 percent chance of success once you have that.

I also would say to Dr. Keefe, I can bet you, and I don't think I am wrong about this, that most of the pharmaceutical companies know whether or not those women that had a positive pregnancy test delivered that baby. They track that information. I would be surprised that they didn't track some of that information.

The FDA sort of alluded to it earlier, that they know they report some of that information. At least, there is some information regarding it; is that correct? So I would be surprised that they don't know some of the outcomes of that data. So I think that data is obtainable.

DR. GIUDICE: That I think we can get to in Question 13. Let's move on. With regard to the pregnancy or the outcome, we have already discussed, not really follicular development rate but follicle development. As you can see from Question 8, the issue is, if you can't power to demonstrate differences in live birth, or ongoing pregnancy rate, then discuss the clinical relevance of these surrogate markers. Dr. Crockett?

DR. CROCKETT: I just have a question

concerning our discussion of these endpoints. In my mind, it seems like it would be easier to separate out what we look at as endpoints for efficacy versus endpoints that I would consider for safety considerations.

For instance, for efficacy of an ovulation drug, I think follicular development is an acceptable endpoint. But I might want to know more about that drug such as does it cause teratogenesis of some kind. I am not aware of that being a problem with any of our current medications, but it is conceivable that, in the future, there could be drugs that do that.

So it would probably be helpful for us, from a safety standpoint, to follow out those patients to a pregnancy or even beyond pregnancy endpoint since we know that genetic factors can carry even beyond birth. So I just wanted to raise that question.

DR. GIUDICE: Thank you. Other comments? Going through the list here, I think we have heard, with anovulatory patients, that follicular development--we haven't really defined follicular development, whether we mean follicles of a given size or estradiols over a given level or

endometrium over a given thickness. Would someone like to make some comments about that? Or should we include those as possibilities for recommendations as we advise the FDA? Okay.

DR. KEEFE: In each of these, it seems that, as you push back further, it makes it easier to do and you need less power, but it always raises the question of is there a downside to the treatment later in development.

So, just to balance these--some, for example, have argued that for ovulation induction, all we need to do is show that they have ovulated. But there are many ways to make people ovulate that almost guarantee there would be a miscarriage. You could drive the LH level so high that they may ovulate but it is not going to develop.

So I would say that, in general, for most of these things, to try to paint with broad strokes, would be that these are useful indicators of efficacy unless there is some scientific evidence to suggest a detrimental impact at later stages of development. So that would open the door to concerns of--for example, we had an intervention that drove up LH levels too high--that maybe, in that, there might be a little bit more concern

about showing more than just ovulation, ovulation with development of fetal heart, and then, if there is some evidence of a treatment causing higher rates of miscarriage after fetal heart, that efficacy should be determined at a later stage of development.

DR. GIUDICE: Okay. Yes; Dr. Emmi?

DR. EMMI: I remember reading a study a couple of years ago that showed that it wasn't just fetal heart. It was when, in the first trimester, fetal heart was obtained, whether the pregnancy would continue to go on. And that was based on age. The older patients, if they didn't have the fetal heart at about eight to ten weeks, had more chance of a miscarriage.

So I don't think that you can just go clinical pregnancy. I think you have to pick a parameter. If you are going to pick clinical pregnancy, you have to pick a time in the first trimester for that ultrasound to be performed to establish it.

DR. GIUDICE: Dr. Emerson?

DR. EMERSON: I have already stated that I sort of wouldn't go beyond the presence of the gestational sac, personally, and I would prefer not

stopping before fetal heartbeat. I do want to point out that if your whole goal is that you think that they can't get a sample size large enough to detect this, you are also guaranteeing absolutely that you won't have a large enough sample size to be able to detect a meaningful rate of fetal abnormality or miscarriage because if you don't even have a large enough sample size to observe that you have got a difference in pregnancies, by the time you have reduced that sample size, you have got a very small sample size of pregnancies to work for.

Seeing zero events, zero events, your confidence bound on a bad event rate is roughly 3 over n. The space shuttle went off 24 times without a catastrophe and the next one blew up and we now know that that rate is actually far higher than you might have thought with just 24 no failures.

So we are making a decision truly as you want to power this down for the clinical endpoints of interest is also we won't be able to detect meaningful adverse-event rates.

DR. GIUDICE: Dr. Rice?

DR. RICE: I can't see why we would think

we would not be able to enroll enough patients in, say, a trial because if you look at, just from the Follistim data she showed, what was it, under a thousand patients in those two arm, to detect a difference. We had approximately 160,000 people maybe last year who used--maybe 100,000 who used gonadotropins, 40 percent of that for IVF, the other percent just with ovulation induction.

So you are talking about 100,000 people in a given year and we can't come up with 2,000 to detect a difference in a product? That would seem unreasonable. I can't imagine that.

DR. GIUDICE: That may be true for pregnancy, for infertility, but if you get the Kallmann's patient, for instance, there you probably will not--I think what Dr. Layman was referring to was the WHO Type 1 and even subsets within that where getting enough patients for power is nearly impossible with a few thousand worldwide.

DR. RICE: Clearly, there are going to be some limitations when we get to specific disorders like that. And we have all seen, in some of those studies, you change your endpoint because you recognize that the disease is so rare that you could not potentially use pregnancy as the outcome.

But if we were looking at infertility, and I would even venture to say we were just looking at "anovulation," the anovulatory patient, that there may potentially still be the opportunity with the incidence we see the disease process in to enroll enough patients.

DR. GIUDICE: I would like to ask the committee, in terms of pregnancy rate for the list here, what the consensus is in terms of recommendation to the FDA, if pregnancy is going to be an outcome. Is it positive beta-hCG? Is it a sac on ultrasound? Is it a sac with a heartbeat? Is it a sac with a heartbeat at six weeks or eight weeks?

Are we going to give a definitive recommendation or leave the list and have that open for interpretation and discussion at the FDA by the FDA? Dr. Keefe?

DR. KEEFE: I would advocate, as was argued very persuasively, that a fetal heart is a big quantum leap, once you have the fetal heart detectable conventionally around six weeks. So I would vote for the fetal heart. You know, on balance, it is practical, doable, and it is a big hurdle that the embryo has overcome.

DR. GIUDICE: Dr. Hager?

DR. HAGER: I think the other advantage is that it is a reproducible event to measure.

DR. KEEFE: It is under the control of the clinic that is doing the study.

DR. GIUDICE: Can we take a vote on this one? Okay. Who is in favor in the definition of pregnancy as the outcome for sac and fetus with a heartbeat. Anyone opposed? Dr. Stanford.

DR. STANFORD: I would think that you could set it as a minimum standard. That is what I would be comfortable with. I think, in some cases, where there may be safety concerns that go beyond that, you would want to allow for having a higher standard in some cases where there is a reason to do so.

But, if you state it as a minimum standard, I think I would be happy with that.

DR. GIUDICE: Okay. Yes; Dr. Hager?

DR. HAGER: Just a comment. I would think that there would be a strong emphasis for follow up and tracking, et cetera. I think we have already conveyed that. So this would be a part of that is what is the ultimate outcome and, also, I would hope that we would encourage follow up regarding

anomalies and aneuploidy, et cetera.

DR. GIUDICE: Thank you. Let's move on to Questions 9 and 10 regarding intention to treat. So the first one, is an intent-to-treat analysis appropriate for ovulation induction and, if not, should cycles be analyzed per patient given hCG? Any comments on that?

We have heard Dr. Emerson. Perhaps he would like to restate his statement.

DR. EMERSON: You know, randomization--you are only protected by randomization if you do intent-to-treat. Otherwise, you are not.

DR. GIUDICE: Yes. Dr. Lewis?

DR. LEWIS: I think you have to use intent-to-treat for both ovulation induction and ART because also cancellation rates are very important and, if a given drug is associated with a higher cancellation rate for whatever reason, you want to know about it. So I think the other endpoints are really not so meaningful, or the other denominators are not so meaningful, I should say.

DR. GIUDICE: Yes; Dr. Toner?

DR. TONER: I would concur. I would just argue, hopefully, for flexibility in the

stimulation arm because, if you constrain every patient to get the same dose, and you are under on some and over on some, you are going to get hardly any to the finish line. And you will look like you didn't do well, but it is because you had your hands tied behind your back.

So I think that was the problem with some protocols that have been done to date and would need to be changed.

DR. GIUDICE: There seems to be pretty good consensus among the group about intent-to-treat. Going on to No. 11, to discuss safety endpoints that should be evaluated. Dr. Stanford?

DR. STANFORD: I will just say, again, I think multiple gestation definitely ought to be on the list. Certainly, there are others that we have talked about, ovarian-hyperstimulation syndrome, obviously ought to be on the list.

DR. GIUDICE: Others? Yes; Dr. Crockett?

DR. CROCKETT: I have a lot of concern about this. I, in particular, am concerned about the risk to the mother being given these drugs or hyperstimulation, not just the immediate effects of the ovarian-hyperstimulation syndrome but are there

oncogenic effects further down her lifetime. So, while it may not be a factor that we need in order to make a decision about certifying a drug for use, it may be important to us, as an agency, to know how those patients, both the maternal side and the fetal side, progress 20, 30, 40, 50 years out.

And we ought to consider having a registry of adverse events just like we do for many other drugs, far out.

DR. GIUDICE: Thank you. Any other comments? Dr. Hager.

DR. HAGER: Bringing up the clomiphene situation long-term, I think, as Dr. Crockett has said, there needs to be a definite registry not only for immediate disease but also the long-term oncologic effects. I would hope that would be designed into all of these trials so that we not only follow the infants but we also follow the moms.

DR. GIUDICE: I think you both make very important points and I think one of the major questions is--actually, there were several questions--who would run and have control over a central registry. Would it be one individual pharmaceutical company? Would it be the FDA?

Would it be the CDC? Would it be SART? Would it be ASRM?

Complementary to that is the whole issue of cost, who would fund the registry. These are issues that I don't know that our committee can solve, but I think, and this certainly bears very much on Question No. 13--but this is something that I think we need to at least discuss in a little bit more detail because I think we are all aware of these issues but, perhaps, we can be of some help to the FDA in making some recommendations.

Dr. Lipshultz?

DR. LIPSHULTZ: Just a comment because I randomized up against this recently. Apparently, as physicians, we all have the duty to report to the pharmaceutical company any adverse event we find in a patient, no matter what they are taking, related to that drug. I just don't think we, as physicians, do that.

We find things. We have patients who develop malignancies and we think it might be related to something, and we never let the company know because we are not part of a study. I mean, these people, after they go through IVF are no longer going to be part of a study, or part of a

group. They are going to move on with their lives but they will all have doctors.

So I don't know whether it is anybody's, any agency's or any subgroup's responsibility, as it is the medical community's responsibility to make sure that all adverse events are reported to the pharmaceutical companies if you think there is a connection.

DR. GIUDICE: Dr. Crockett and then Dr. Keefe.

DR. CROCKETT: I think this is a public-health problem. I don't think this is a pharmaceutical-industry thing. Frankly, no disrespect to our pharmaceutical people that are in the audience, but I don't trust the drug companies to take the whole public interest to heart. I think that is why we have agencies like the CDC that track other public-disease problems.

I would suggest that should become part of our responsibility in that regard.

DR. GIUDICE: Okay.

DR. KEEFE: I agree. If you really want it to work, you don't go to the pharmaceutical companies. You don't depend on doctors. You put it out there, maybe foundations, CDC, a number of

sort of public-interest groups, and the patients will help us because they will start calling, they will start reporting. They will say they have a friend who had this or that, and all the chat rooms. That is how to make it work. Then you really have data.

There are registries available. The Australians have a registry. They have published extensively on long-term outcomes and they are not finding a lot. But they are looking more at cancer, I think the developmental anomalies, aneuploidies, other issues regarding the offspring and potentially other issues regarding the patients.

But I think it has to be outside everyone and of its own domain funded externally. Then it will work.

DR. GIUDICE: Dr. Dickey, you had a comment? And then Dr. Hager?

DR. DICKEY: Judging from some of the other things that we fund externally to keep registries, that doesn't work so well either. They are dependent on state funding which comes and goes. But I do think the whole conflict of interest thing that we are so attuned to here

suggests, in fact, that it would probably be better if it came from some centralized entity as opposed to expecting the pharmaceutical industry to track that.

I don't have a great deal of faith that either doctors or patients, in fact, do a very good job long after the fact of trying to correlate back what happened. So I think you have to have a little more specific guidance out there of who you need to report to you and what kinds of things you want reported and, perhaps, then chat rooms and things can do it.

But very few of us, in fact, correlate something back to something that happened twenty years ago or thirty years ago. And I am not sure how you get somebody to fund it.

DR. GIUDICE: Dr. Hager?

DR. HAGER: There is precedent. When metronidazole was approved by the FDA as an antianaerobic agent, there was some rodent data at the time that indicated the potential for some problems. I was at the CDC. We began a registry at that time looking at deleterious effects in newborn infants.

As you know, we have discontinued that

registry because there are none. But, there is precedent for cooperation between the two agencies to develop a registry and to follow that, and it is a public-health measure.

DR. GIUDICE: Dr. Macones, Dr. Lewis and Dr. Brzyski.

DR. MACONES: I agree that I really see this as a domain of the pharmaceutical company to put together a registry for this. I think I could see this discussion happening with the NIH coming up with, developing a long-term follow up for a cohort of people who were exposed to ART drugs. I think there is some precedent for a very successful long-term follow-up physicians health study, nurses health study, that have been remarkable.

I just can't see how a 40-year follow-up study is going to be the responsibility of a pharmaceutical company. I think this needs to be independently funded and I think the NIH would be very receptive to something like this.

DR. GIUDICE: Dr. Lewis?

DR. LEWIS: A registry could provide enormously helpful information. I think you have to be very careful about privacy concerns of patients, particularly in this day and age with

HIPAA regulations. So it would have to be, obviously, voluntary. I don't think it is realistic to think the pharmaceutical industry is going to pay for it.

One thing about patients recollections, it has been shown quite clearly that if a patient develops a disease, they recall very clearly every single thing they ever took. So recall bias could be quite a problem so you really do have to think about doing it prospectively. But it is very expensive and there are a lot of obstacles in the way.

DR. GIUDICE: Dr. Brzyski?

DR. BRZYSKI: One point I would like to raise is the issue that we really don't have good information about complications, first of all, even in the general population. There is a quote--I tried to track down a reference for the commonly cited rate of 2 to 4 percent for severe birth defects. I can't find it. It is on the CDC website, but there is no citation for, like, where that rate comes from.

If you look--well, anyway. And then if you look at infertile individuals who may have other predispositions, genetically or

epigenetically, to have complications or problems down the line, really, we have no information on that population compared to, say, the ART population where there is concern about, for instance, birth defects in the children.

So when we look at a registry, if all we have is a registry of people exposed to fertility medications, I am not sure how to put that in perspective in terms of the general population or in terms of infertile individuals who were not exposed to fertility medications. So now you are doubling or tripling your effort to collect the data on those other populations so you can make some comparison about, well, is the rate of 2 per 10,000, is that bad or good compared to other individuals.

DR. GIUDICE: Thank you. Another question also is what type of information would be collected for this type of registry. I think, certainly, now that there are potentially three populations to collect information on. This does bring up even yet again a larger database but I would like to hear--perhaps we can discuss, as a group, some potential entries into this hypothetical database. Dr. Crockett?

DR. CROCKETT: Well, let's start at the lowest end which would be adverse events during the pregnancy that you wouldn't have otherwise expected. So I would want to know fetal anomalies identified by ultrasound and a fetal-loss rate beyond the first trimester, perhaps.

Beyond that, after birth, then tracking out the infant or the baby, I would want to know about specific diseases, particularly oncogenesis, learning disability, diseases which we are concerned about now with other drugs that we are introducing into our pregnancy women. From the maternal side, I have already mentioned, oncogenesis, also.

DR. GIUDICE: Thank you. Dr. Toner?

DR. TONER: I was just going to mention, kind of following up with what Bob had said, that every state in the country has its own rules about how to do birth-defect reporting. There is no consensus about what is in the list, what is off the list, who collects it, when it is observed. So that is partly why Dr. Brzyski can't find a reference, because it is done a hundred different ways.

Probably the most simple thing to do would

be to go into a state that already does it across the board and then parse them up by, are these IVF babies or are these not so you don't focus undue scrutiny on the IVFs and don't apply that same scrutiny to everybody else.

As I understood the CDC, they were trying to get such a thing to happen in Massachusetts where there is potential for linkage. But in most states, because of privacy concerns, you can't know who the baby is, who the mom was and who got treated. So it is going to be very tough, I think, except in a state potentially like Massachusetts.

DR. GIUDICE: Dr. Keefe?

DR. KEEFE: The Dutch have done that, of course, published. They looked at one particular outcome which is adverse neurological sequelae and found huge increases. Those were in group homes where the most severely impaired live. And so there is much less detection bias.

It is true there are problems, but there is such precedence for it being such a large impact and the vast majority of the abnormalities, by the way, were linked to multiple births and prematurity, as you would expect.

So now it looks like I think 2 percent of

all births in Sweden come from IVF babies. It is going to grow in the United States. I think we should start planning now. I think it has got to be done. I mean, look at DES, diethylstilbesterol. This was the previous generation's new technology. It was some well-meaning, brilliant Harvard professors of biochemistry and GYN who had this tremendous theory of the value of estrogen in early pregnancy which made a lot of sense then but, in retrospect, is laughable, how naive it was and led to some really adverse outcomes.

So I think it is way too big and way too important to say it can't be done. It is just a question of how, really.

DR. GIUDICE: But the specific question from the FDA was whether the pharmaceutical companies would need to be responsible for these registries. What I think I have heard from the group is that this is way bigger than any one pharmaceutical company. Dr. Rice and then Dr. Stanford.

DR. RICE: I agree with Dr. Keefe. I think just because it is hard doesn't mean that we shouldn't do it. I think we need to do it. I think it is going to be valuable information that

is going to find invaluable as we look at long-term outcomes. And I do think it is bigger than the pharmaceutical companies. So I don't know if I would put, per se, the burden on them to do it.

I don't know where the burden, per se, falls but I think it definitely needs to be done because we are treating too many women at this--we are treating so many women at this point that, at some point, we are going to get to a point where we can actually detect some differences, and those differences may be very significant in looking at not only fetal outcomes but maternal outcomes, also.

DR. GIUDICE: Dr. Stanford?

DR. STANFORD: Just a comment. It is probably stating the obvious but maybe just a comment that the FDA would welcome the assistance or cooperation of the drug companies in these efforts as opposed to you should do this, or this has to be done somehow, and stating that that is a desirably--obviously, that doesn't have any teeth, regulatory teeth, but just as a statement of policy that they would appreciate cooperation or assistance in terms of tracking people from trials or whatever that may come up.

DR. GIUDICE: Dr. Rice?

DR. RICE: And we are talking about a registry for ovulation-induction patients and ART patients because we get caught up in the ART part of it but the ovulation induction is mostly what we see. Our gonadotropins are usually patients who use insemination or time to intercourse, et cetera, and that. So we want to make sure that we are talking about areas.

DR. GIUDICE: Yes; I am glad you clarified that. Thank you. Dr. Brzyski?

DR. BRZYSKI: One comment, one recommendation, I would have is I hear from my colleagues that there is concern, and we have talked about it here, concern about privacy. I mentioned, myself, the stigmatization of offspring. There are, perhaps, educational efforts that need to be made in the patient populations and that would be the efforts that professional societies could make, patient-advocacy organizations could make. Even the FDA might be able to inform the public regarding the relevance of--you know, if a registry is established, the importance of participating in that and educating patients and families and physicians about the public-health

impact of that registry.

DR. GIUDICE: Thank you. Yes; Dr. Crockett?

DR. CROCKETT: I just wanted to add one more thing to that. A registry is really helpful but the other part of this is informing the patient about what knowledge is available concerning the medications that they are taking and what knowledge is not yet available.

For instance, on the product labeling, there probably should be a statement that makes the patient aware of the safety data and efficacy data that has been looked at and the terms of those studies and the limits of those studies. If we had done that with hormone-replacement therapy twenty years ago, then we may have had a little less trouble dealing with the questions of that whole breast-cancer issue.

I think people, when the FDA says something is approved, they have the assumption that it is safe and it is efficacious for the nth degree. If we are establishing safety for a short period of time that we are looking at, it needs to be made clear in the literature to the patients.

DR. GIUDICE: That is a very good point.

Dr. Rice?

DR. RICE: I think we ought to be very careful about that because we can only use the information that we have currently to make those limited statements on safety. So we don't want to be using level 2 or 3 evidence to say what safety is. So you want to be using--you want to be very careful about that because, when you used the hormone therapy as an example, based on when the information was presented to the FDA, when they wrote those guidelines, that was the best information they probably had then.

It changes over time. So I think that is the key thing. You ought to keep the flexibility so that you can change it as more--as the best evidence becomes available.

DR. CROCKETT: So that again. That is exactly right.

DR. RICE: I don't know how I said it.

DR. CROCKETT: The patient needs to know that, though.

DR. RICE: Right. The patient clearly needs to know this.

DR. CROCKETT: That this is based on our best information right now and that it may change

over time.

DR. RICE: But I think there are different--there is different quality of evidence. If that was the case, then, we could take some studies now and put a potential oncogenic risk to taking infertility drugs. But many of us would say that that evidence is not based on anything prospective. And so you wouldn't want to put that on the label to unduly scare the patient at this point.

DR. CROCKETT: You misunderstood what I said. Let me clarify.

DR. RICE: Okay.

DR. CROCKETT: I didn't want to say specifically that there may be a risk of X, Y and Z that has not been discovered yet. I mean merely that we would tell the patient what has been discovered and what the time frame is of that discovery and that there may be, in a general way--like you previously stated, there may be risks in the future that are undisclosed at the time that the drug was approved.

DR. RICE: But then you have to ask the question which risk. You know, that is the thing, which risk. There are several possible risks that

could be there.

DR. CROCKETT: There are a million possible risks.

DR. RICE: So your possible risks needs to be based on some good evidence.

DR. CROCKETT: No. There are a million possible risks that we take with anything that we take into our bodies, but the patients, when they see something from the FDA that says it is approved, the patient and the public think that that means it is safe forever and ever. And we ought to be able to tell them, no, we have looked at the safety and efficacy regarding these things in this short period of time, and there may be things that we don't know about yet.

We don't have to list what those things may be but we need to at least disclose the limitations to what we know.

DR. GIUDICE: I think it is an obligation of clinicians who are prescribing medications to inform patients about risks and benefits and also to give them the package insert so that they read all of the information. But I think what you are getting at is, in the package insert, some statement should be made, perhaps more clearly,

that there are limitations to the evidence.

DR. CROCKETT: Yes. The whole reason I thought to bring this up is because, if we are talking about a registry, it would be helpful for the patient, when they get that medicine, to know that, if they have an adverse effect that was not discovered at the time the drug was approved, that they need to report it to somebody.

DR. GIUDICE: Yes. Thank you. We have one more question to do and that is No. 12. Oh; there was part of 13. At what point should a registry be terminated. I think, perhaps, Dr. Hager described quite well the CDC experience that if there is no difference, then why continue to have a registry and spend the money, et cetera. Unless others have additional comments on that.

Okay. It states here that drug manufacturers conducting studies for female infertility currently obtain the following indications; induction of ovulation in pregnancy and multiple follicle development in ART. Please comment on the appropriateness of these indications given your discussions of endpoints and analyses.

I would like some clarification on exactly what the question is. Are you asking us to comment

on other potential studies or are these the only indications that should be sought?

DR. SLAUGHTER: These are the current label indications that are either induction of ovulation or induction of ovulation and pregnancy or multiple follicular development and ART. Would you comment on whether or not you would modify, given all the discussions that you have had on the endpoints, et cetera.

DR. GIUDICE: Dr. Emerson?

DR. EMERSON: I'm in favor of an indication that indicates it is pregnancy when that is the endpoint. So the only way that I could imagine that the multiple follicular development and ART indication should go forward is at the point that we did develop the technology for cryopreservation of oocytes. Well, then, that would become the endpoint in itself. But I would imagine that should happen when we also feel more comfortable that that is the ultimate goal.

So I would want to add pregnancy on the follicular development and I would like to make it that it is induction of ovulation and pregnancy not just induction of ovulation.

DR. GIUDICE: Dr. Stanford?

DR. STANFORD: I would support that. It seems to me what we are basically saying is that some drugs have been approved for ovulation induction and pregnancy outside the IVF setting and then other drugs in the IVF setting have been approved, or sometimes the same drugs, simply for follicular development.

It seems like make it the same standard. Make it pregnancy across the board is what we are saying, it seems to me. Now, I would also say, though, that if we talk about some potential exceptions to that, I am still--if there is some kind of rare condition where there are not enough numbers to do that, I think that may be another discussion. But I think whenever you have got the potential--I agree that the basic gold standard should be the same across the two types of therapy.

DR. GIUDICE: Other discussion? There may be rare instances where individuals may have ovulation induction and even oocytes retrieval but then elect not to have embryos transferred for the purpose of establishing a pregnancy such as an ongoing malignancy.

But, again, these are very, very rare. So I don't know how that would be handled. I guess

one can certain prescribe medications off-label or at least not for these indications but the issue is really for the study. Yes; Dr. Emerson?

DR. EMERSON: But in that instance, wouldn't they be doing this in order to have a later pregnancy? The indication why they were doing this for in the case of the cancer therapy was to protect the idea that they might later want a pregnancy.

DR. GIUDICE: Yes; but for a pharmaceutical company, if the endpoint is pregnancy and your patient population is a cancer population, that may be a very long time before you end up with an outcome.

DR. EMERSON: So I guess I am saying that I wouldn't imagine them running a clinical trial specific to cancer patients just prior to chemotherapy. But, once you are confronted with facing chemotherapy, it seems quite reasonable to pursue something that would preserve your childbearing potential and, if there was a drug that did that, we have got that indication.

DR. GIUDICE: Other comments? Dr. Keefe?

DR. KEEFE: If we added that to the indication, doesn't that sort of lock in pregnancy

as the only outcome that would be permitted in terms of the acceptability of a new drug? It seems to me that we just had that discussion earlier and there was a little bit of debate about that. So I am just a little bit concerned that that goes into the indication.

We have a chemotherapy drug and we are saying that life is the outcome. You might be able to cure a cancer, but these are in 90-year-old patients. It seems that multiple follicular--it is very clearly connected to the outcome. I am sort of voting for flexibility in that and I am just afraid that we may lose flexibility if we put it into the indication.

DR. SHAMES: Some companies could elect to have pregnancy the outcome and some people could elect to have follicular development. I mean--

DR. RICE: But if you were a company, which one would you select?

DR. SHAMES: Depending on how the trial is run, you need to match the trial to the indication. To use your oncology analogy, we don't expect everybody to get pregnant. It is a proportion of people that get pregnant. Generally, we try to match the indication to the clinical trial and what

is important. If pregnancy is important, that would be the endpoint in a particular trial and that would be the indication they got.

It is possible, under certain circumstances, for some unusual circumstances, that there would be others and that would be the indication that they would get.

DR. GIUDICE: Dr. Toner and then Dr. Slaughter.

DR. SLAUGHTER: I guess the only comment I would have is that, as we talked about, should the science advance to the point that we have a surrogate that we really do think is a predictor. It is possible to modify the guidance document at that time.

DR. GIUDICE: Thank you. Dr. Toner, you had a comment?

DR. TONER: I was wondering whether sort of a hybrid indication such as follicle development for pregnancy might better capture what the gonadotropin part of this process is about. And to say pregnancy would apply for any of the components in the whole process that would pertain to the progesterone we are using or the hCG we are using. Why are we using it? For pregnancy. So maybe we

want to be a little more specific than pregnancy undefined.

DR. GIUDICE: Dr. Rice?

DR. RICE: I guess I am confused. Didn't we discuss this already? I mean, what did we vote on? I guess I am confused about why we are backtracking here back down this because I thought that the whole--when everybody was polled, we were saying fetal cardiac activity, et cetera.

DR. KEEFE: Since I am the one who brought it up, let me explain why.

DR. RICE: Okay.

DR. KEEFE: I think once you put it in--it is a different level. It is a different burden. It is one thing to recommend that it be used as the outcome. It is another to put it as an indication. To me, it is another level of--another burden that you are incurring. That is why I brought it up. Not so much that we hadn't discussed it. think we did almost ad nauseam. But the question is do you want to kick it up to this next level where it actually enters the indication.

I agree with Jim. If you say "for" pregnancy instead of "and" pregnancy, that one word makes it so that you have the wiggle room that you

need to keep it an open shop and, at the same time, reasonable. That's all.

DR. RICE: I think you are leaving a lot of flexibility because if I was a pharmaceutical company and I had the option, I would go for just the follicular development and "for pregnancy" sounds like my secondary endpoint which I am already looking at.

So I guess I don't distinguish between the two. I mean, either you are going to look at pregnancy as your outcome by some measure or you are going to continue to look at follicular development. If we are going to continue to look at follicular development, are we really going to be ever distinguish between Product A versus Product B versus anything else that comes on the market if they only have to look at follicular development, because we have pretty much shown that, when you look at the different products, there is not a great deal of difference in follicular development.

You would have to decide, are we going to raise the level of rigor here or are we not.

DR. GIUDICE: Dr. Stanford?

DR. STANFORD: It just seems to me the

main indication--and if the FDA staff disagree with that, I would like to know, but it seems to me they are the same.

DR. GIUDICE: Dr. Shames?

DR. SHAMES: The primary endpoint is the indication. So if we decide that pregnancy is the primary important endpoint, then it should be the indication.

DR. GIUDICE: Dr. Layman?

DR. LAYMAN: I just have a point for clarification. We said we agreed that fetal heart beat was the indictor for pregnancy. But, in that Question 8 that says, if we are not taking live birth or ongoing pregnancy rate, is the way that question reads. Did we mean--I'm not clear whether we meant we want to say for pregnancy all of them or just in that question, I mean, because it is not clear.

I didn't get the feeling we agreed on what was the best pregnancy to consider. It just says, if you can't have enough power for live birth rate or ongoing, which one. And we said heart beat.

DR. GIUDICE: Right. That was what we had discussed. Now, there seems to be some ambiguity, though.

DR. SLAUGHTER: I don't believe so. The question was worded is live birth or ongoing pregnancy to suggest something further along than gestational sac with a heart beat. So I think we clearly understand that that is what we are talking about now, that we want gestational sac with heart beat.

DR. GIUDICE: Right.

DR. LAYMAN: Right. I agree, but does that mean for every study that is what we are requiring? We are not requiring live birth or farther on. That is what I am asking. Did we agree to that?

DR. GIUDICE: We agreed to that. Dr. Lewis?

DR. LEWIS: I thought the only exception was for Type 1 anovulation where we agreed it would be very difficult to power a study to look at--okay.

DR. GIUDICE: Right. Dr. Stanford, you had a comment?

DR. STANFORD: Sort of the next question; it is sort of like there is an 8A question that is, if the study cannot be powered for presence of fetal heartbeat, and that might be the Type 1, then

what is an acceptable outcome in that case. It sounds like that is the question you are putting on the table.

DR. LEWIS: I wasn't really putting it on the table because I thought we had already discussed it. I thought we had already agreed that it would--ovulation. I'm sorry; follicular development.

DR. STANFORD: So, I guess, then the issue is under what--if we all agree that fetal heart beat is a reasonable standard in general, then, under what conditions do you accept something else as a main outcome, something less than that as a main outcome, then, as an indication. We are saying that Type 1, WHO Category Type 1, anovulation would be one.

DR. GIUDICE: Right. Are we all clear on that? Any further discussion? It is amazing we got through all those questions. I would like to thank everyone for their participation today. Tomorrow's session begins at 8:30 in the morning and the discussion will be NDA 21322 on Luveris. For members of the committee, please bring your green FDA binding document and your gray Serono briefing documents.

[Whereupon, at 4:39 p.m., the meeting was recessed to be resumed at 8:30 a.m., Tuesday, September 30.]

- - -