

[--- Unable To Translate Graphic ---]

DEPARTMENT OF HEALTH AND HUMAN SERVICES  
PUBLIC HEALTH SERVICE  
FOOD AND DRUG ADMINISTRATION

OBSTETRICS AND GYNECOLOGY DEVICES PANEL  
FIFTY-SEVENTH MEETING

Volume I

**OPEN SESSION**

Monday, July 14, 1997

9:50 a.m.

Room 20B  
9200 Corporate Boulevard

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

[--- Unable To Translate Graphic ---]

Rockville Maryland

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

[--- Unable To Translate Graphic ---]

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

Gary S. Eglinton, M.D., Chairperson  
Elisa D. Harvey, D.V.M., Ph.D., Executive Secretary

Voting Members

Michael P. Diamond, M.D.  
Barbary Levy, M.D.

Temporary Voting Members

Michael R. Neumann, Ph.D., M.D.  
David F. Katz, Ph.D.  
Diane D. Davey, M.D.  
Timothy J. O'Leary, M.D., Ph.D.  
Diane Solomon, M.D.

Industry Representative

Cindy Domecus, R.A.C.

Consumer Representative

Diony Young

Staff

Colin Pollard  
Deborah Smith  
Yung Pak  
Dan Schultz, M.D.  
Kathy Daws-Kopp  
Al Montgomery  
Lillian L. Yin, M.D.

[--- Unable To Translate Graphic ---]

C O N T E N T S

<u>AGENDA ITEM</u>	<u>PAGE</u>
Meeting called to order	4
Introduction - Colin Pollard, Chief, Obstetrics and Gynecology Devices Branch, Center for Devices and Radiological Health	10
Product Development Protocol (PDP - New CRDH Initiative Elisa D. Harvey, D.V.M., Ph.D.	14
FDA Presentation "Intrapartum Continuous Monitors for Fetal Oxygen Saturation and Fetal pH"	24
Kathy Daws-Kopp, OB-GYN Devices Branch	26
Deborah Smith, M.D., FDA, Office of Women's Health	29
Industry Comments Healthdyne - Michael Ross, M.D.	33
Panel Comments on Guidance Document	37
Open Committee Discussion Invited Guest Speakers	
Cervical Cancer Screening and Detection Mark Schiffman, M.D., M.P.H., (NIH)	78
Technology of in vivo devices for detection of cervical cancer Rebecca Richards-Kortum, Ph.D. (University of Texas at Austin)	95
Clinical Trial Design for in vivo devices for detection of cervical cancer Robert Hirsch, Ph.D. (George Washington Univ)	118
Industry Presentations	
1) Trylon Corporation - Stewart Lonky, M.D., Medical Director	141
2) Polartechnics Limited - Malcolm Coppleson, Clinical Director	156
FDA Presentation	

[--- Unable To Translate Graphic ---]

"In vivo Devices for the Detection of Cervical Cancer  
and its Precursors: Submission Guidance for an IDE"  
Mridulika Virmani, Ph.D., OB-GYN Devices Branch

164

Adjournment

197

[--- Unable To Translate Graphic ---]

P R O C E E D I N G S

DR. HARVEY: We'd like to call the meeting to order.

CHAIRMAN EGLINTON: We do have an audience sign-in sheet. I'd like to make sure that everyone does sign in. If members of the audience have any comments, please remember you must be recognized before assuming the podium. You must use microphones, give your full name and your affiliations and who sponsored your trip here today, including any travel or per diem fees or involvement with any other companies.

We'd like to have the panel members introduce themselves now. We will start this way starting with Dr. Katz, please.

DR. KATZ: I'm David Katz from Duke University where I'm a professor in the Departments of Biomedical Engineering and Obstetrics & Gynecology.

DR. O'LEARY: Timothy O'Leary, Armed Forces Institute of Pathology.

DR. LEVY: I'm Barbara Levy. I'm a gynecologist practicing in Federal Way, Washington, and clinical assistant professor of OB-GYN at the University of Washington.

[--- Unable To Translate Graphic ---]

DR. DIAMOND: My name is Michael Diamond. I'm a professor of obstetrics and gynecology and director of the Division of Reproductive Endocrinology and Infertility at Wayne State University in Detroit, Michigan.

MS. DOMECUS: Cindy Domecus, senior vice president of Clinical Research, Regulatory Affairs, and Quality Assurance for Conceptus. I'm the industry rep to the panel.

DR. YIN: Lillian Yin. I'm the director of the Division of Reproductive, Abdominal, Ear, Nose, and Throat, and Radiological Devices, CDRH.

MS. YOUNG: I'm Diony Young from Geneseo, New York, and I'm editor of the journal Birth-Issues in Perinatal Care.

DR. NEUMANN: I'm Michael Neumann from Case Western Reserve University in Cleveland.

DR. SOLOMON: Diane Solomon, National Cancer Institute, Bethesda, Maryland.

CHAIRMAN EGLINTON: Gary Eglinton, director of Maternal and Fetal Medicine, Georgetown University.

DR. HARVEY: Elisa Harvey, OB-GYN Devices branch, executive secretary for the OB-GYN Devices Advisory Panel.

CHAIRMAN EGLINTON: Thank you.

The FDA press contact for today is Dr. Yin. We do

[--- Unable To Translate Graphic ---]

have a full agenda. We'd like to keep pressing on. If there are any comments, please keep them brief and concise, and, again, no outbursts from the audience, please, but come to the podium. Thank you.

DR. HARVEY: I'd like to start by acknowledging that we have several temporary voting members with us today, and I would like to read a statement which is their appointment to temporary voting status, which has been signed by Dr. Burlington, Center Director.

Pursuant to the authority granted under the Medical Devices Advisory Committee Charter, dated October 27, 1990, as amended April 20, 1995, I appoint the following people as voting members of the Obstetrics and Gynecology Devices Panel for the duration of the panel meeting on July 14-15, 1997: Diane Davey, M.D., David F. Katz, Ph.D., Michael R. Neumann, Ph.D., M.D., Timothy J. O'Leary, M.D., Ph.D., and Diane Solomon, M.D.

For the record, these individuals are special government employees and are either a consultant to this panel or voting member of another panel under the Medical Devices Advisory Committee. They have undergone the customary conflict-of-interest review, and they have reviewed the material to be considered at this meeting. And

[--- Unable To Translate Graphic ---]

as I said, it's signed by Dr. Burlington.

I would now like to introduce the panel to our new consumer representative, whose term began in January of this year. It is a four-year term. Diony Young is her name, and she has an extensive background and long experience working as a consumer advocate, particularly in the areas of prenatal and perinatal care and childbirth education. She has published extensively in these areas and has previously served on advisory panels at NIH. As you heard her say, she is also currently the editor of the peer-reviewed journal Birth-Issues in Perinatal Care. I'm sure she's going to bring an important consumer perspective to the panel, and I would ask that you give Ms. Young a warm welcome to the panel.

I would now like to read the conflict-of-interest statement and waivers which apply to this meeting. The following announcement addresses conflict-of-interest issues associated with this meeting and is made part of the record to preclude even the appearance of an impropriety.

To determine if any conflict existed, the agency reviewed the submitted agenda and all financial interests reported by the committee participants. The conflict-of-interest statutes prohibit special government

[--- Unable To Translate Graphic ---]

employees from participating in matters that could affect their or their employer's financial interests. However, the agency has determined that participation of certain members and consultants, the need for whose services outweighs the potential conflict of interest involved is in the best interest of the government.

A waiver has been granted to Dr. Michael Diamond for his interest in a firm at issue that could potentially be affected by the panel's deliberation. The waiver permits this individual to participate in all matters before the panel. Copies of this waiver may be obtained from the agency's Freedom of Information Office, Room 12A-15 of the Parklawn Building.

We would like to note for the record that the agency took into consideration certain matters regarding Drs. Michael Diamond, Michael Neumann, and Diane Solomon.

Dr. Diamond reported that department colleagues have had relationships with fetal monitor firms and have been or are involved in research relating to fetal monitoring and cervical cancer screening. However, he has no personal involvement nor any managerial responsibilities for these colleagues. In the absence of any financial interests, the agency has determined that Dr. Diamond may

[--- Unable To Translate Graphic ---]

participate fully in these deliberations.

Dr. Neumann reported a relationship with an electronic fetal monitor firm on matters not related to what is being discussed at this meeting. Since this matter is unrelated to the specific issues before the panel, the agency has determined that he may participate fully in the panel's deliberations.

Dr. Solomon reported an NIH study for which firms at issue provide materials and equipment at their own cost. In the absence of any personal or imputed financial interest, the agency has determined that she may participate fully in panel deliberations.

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 should 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 all persons making statements or presentations disclose any current or previous financial involvement with any firm whose products they may wish to comment upon. We would like to note for the record that

[--- Unable To Translate Graphic ---]

Rebecca Kortum, who is a guest speaker with us today, has acknowledged that her employer has an interest in a firm to develop methods for in vivo detection of cervical cancer.

That's the conflict-of-interest statement. I wanted to also note for the record that transcripts or videos are available of the meeting, if so desired.

Transcripts are available through the Miller Reporting Company, and that phone number is (202) 546-6666. Videos are available through Video Visions, and that phone number is (301) 438-8726.

I believe we have--most of the presentations today have already been given to us, but those presenters to the panel who have not already done so should provide FDA with a hard copy of their remarks, including overheads. And if there is anybody who needs to give FDA copies, Mr. Yung Pak, if he could stand, he will collect these from you at the podium. Thank you, Yung.

Dr. Eglinton?

CHAIRMAN EGLINTON: Mr. Colin Pollard will now give a brief overview of the purpose of this meeting and update the panel on recent activities of the devices branch.

MR. POLLARD: Thank you, Dr. Eglinton.

Before we get into the first agenda item to look

[--- Unable To Translate Graphic ---]

at the guidance document on new intrapartum monitors, I wanted to make a few announcements of some recent FDA activities.

The first thing I'd like to tell you is back on February 14th of this year, FDA approved its first humanitarian device exemption application, and this was for a fetal bladder stent. If you will recall, back last July we spent an arduous day struggling with the PMA in that regard. At that time, the regulation for the humanitarian device exemption had just been published but was not effective. The company applied for and received qualification for this status, and I have copies of the summary of safety and probable benefit for any of the panel who are interested. But for us it was a breakthrough experience, and I suspect that over the coming years for devices that have very limited target population in a given year in the U.S. that this is something that will be made available.

I would also like to mention, just sort of following up from our last panel met, that we had just issued some policy changes regarding a number of products. One was in the area of falloposcopy, where we had moved it from a PMA track to the 510(k) tract. We approved the first

[--- Unable To Translate Graphic ---]

two falloposcopes earlier this year, and we hope that that is going to be of use to women who are trying to become pregnant.

Another document that we just released for comment a couple of months ago and is available in your folder is a medical device labeling suggested format and content, and this is a document that was a result of a center-wide effort to gain some consistency across the board on medical devices and the labeling for them. You have a copy, and if you have any comments, there is a mechanism for letting us know what you think.

Finally, we have asked Dr. Deborah Smith from our Office of Women's Health to talk to you about a long-time ongoing project of FDA to gain consistency across devices and drugs in the area of contraceptive effectiveness labeling.

DR. SMITH: Thank you, Colin. Good morning, everyone.

As Colin said, over the last five years there have been a number of PHS initiatives focusing on women's health issues and contraceptive issues in particular. In 1993, the agency announced that the labeling for contraceptives would be strengthened by making it uniform. There were different

[--- Unable To Translate Graphic ---]

presentation formats as well as data in oral contraceptive labels as compared to devices and OTC products. Drs. Lisa Rehrich(?) and Susan Alpert representing CDER, Center for Device Evaluation and Research, Reproductive Health Drugs Division, and CDRH, respectively, began working on a uniform label for contraceptive products.

The table published annually by Trussell(?) et al. in contraceptive technology was agreed to be the best available data source. A draft table was developed, and members of the CDRH Office of Health and Industry Programs, also known as OHIP, proposed focus group testing as an appropriate way of determining the presentation format for this information that would be most useful to the consumer.

This was formulate into a qualitative research proposal which was submitted to my office, the Office of Women's Health, by Paula Silverberg of OHIP and was funded in FY96. After an initial round of focus groups and internal review, additional monies were awarded in FY97 to evaluate a revised table. Colin, if you will put up the first overhead?

What you see before you, first, are examples of current contraceptive efficacy information, including the labeling of one oral contraceptive; and, second, one of the

[--- Unable To Translate Graphic ---]

currently marketed IUDs.

Now, if we go, what you now see before you--and you have a copy of this in your packet--is a sample of the recently approved contraceptive efficacy table to be used in the labeling for virtually all contraceptive products that was developed as a result of this project.

In my opinion, the result is obviously an improvement and quite commendable. We all hope that you agree. I'm pleased that I was able to participate in this project as the Office of Women's Health project officer and pleased to have the opportunity to present this accomplishment to you. It represents not only a long effort, as Colin alluded to, but I think a superb effort on the part of two centers and the multiple divisions within.

Does anyone have any quick questions or comments?

[No response.]

DR. SMITH: Otherwise, thank you.

MR. POLLARD: Thank you, Dr. Smith. For the audience, there are copies of that table on the table outside the room.

The next agenda item we wanted to brief the panel on is a new initiative within the center for an alternative to the PMA. It's called the Product Development Protocol.

[--- Unable To Translate Graphic ---]

Dr. Harvey is going to tell you all about it.

xx DR. HARVEY: Thank you, Colin.

Good morning. We would like to inform you of a new initiative which is being proposed as part of FDA's re-engineering efforts called the Product Development Protocol, or PDP, as I will refer to it.

All the panel members have in your "day-of" folders a copy of the PDP information which is currently posted on FDA's Web site at [www.fda.gov](http://www.fda.gov) for public comment. I would like to emphasize that the development and implementation of the PDP process, of the PDP alternative is an ongoing process which is updated nearly weekly. We encourage any input from you as well.

I will provide a brief introduction to the elements and process of PDP. If you have any questions, Dr. Yin, who heads up the center's PDP re-engineering team, can also help to address those for us afterwards.

PDP is intended to provide an alternative pathway to market for companies developing class III devices which would otherwise be required to go through the premarket approval, or PMA, process. Actually, it's not a new idea. The statutory authority for PDP was originally granted as part of the medical device amendments to the Federal Food,

[--- Unable To Translate Graphic ---]

Drug, and Cosmetic Act of 1976. However, this alternative process was not implemented at that time because it was considered potentially complex and there was a need to focus attention on implementing the other core provisions of the medical device amendments of 1976, such as PMA and 510(k).

The intent of implementing PDP now is to reduce both the resources required by FDA to review class III devices, as well as the total time to get one of these new devices to market. However, I should stress that the requirements for safety and effectiveness will be no less stringent under PDP than they are for PMA. Only the way in which these requirements are satisfied differs.

Here is a simplified time line of the process of development of a medical device. On the right, I have noted where FDA involvement is in the process. As you can see, PDP requires extensive interaction between the sponsor and FDA much earlier in the process than has been done in the past. This is to the benefit of both the sponsor and FDA because it decreases the probability that there will be surprised which may slow or prevent the approval of the device for marketing.

The elements of PDP are the following: First, candidates for PDP are those devices which would otherwise

[--- Unable To Translate Graphic ---]

be subject to premarket approval, as I previously mentioned; two, advisory panel review, still a required part of the process, although panel input will obviously come at a much earlier stage in the time line I just showed you than it has in the past for PMA, and I'm sure we'll probably get some questions from our panel on that aspect of PDP; finally, the proposed PDP must include descriptions of the device and any anticipated changes, all preclinical and clinical protocols, manufacturing methods, facilities and controls, and proposed labeling for the device.

The following is a summary of the PDP process. An approval or disapproval decision of the proposed PDP must be made by FDA within 120 days of receipt of the PDP. When all protocols have been completed by the sponsor, they are to send a notice of completion form to FDA, including any last results which have not yet been reviewed. FDA then has up to 90 days to declare the protocol either completed or not completed. If it's declared completed, then the device may go to market.

The following are the different phases of PDP, each of which I will briefly discuss: presubmission, filing review, FDA review, preclinical phase, clinical phase, notice of completion, and FDA's declaration of the completed

[--- Unable To Translate Graphic ---]

PDP.

PDP can be thought of as a criteria-based research template in which a kind of contract between FDA and the sponsor is agreed upon. As for presubmission, at this stage the applicant should consult with FDA and, I might add, any other outside parties, consultants, et cetera, to develop the proposed PDP.

The development of the PDP will require very early and extensive interaction with FDA and possibly consultants to provide in adequate detail all of the required information.

The applicant may the submit a summary outline of the proposed PDP. FDA will have 30 days to determine from the summary outline whether the proposed PDP appears to be an appropriate candidate for this alternative process.

If it is determined that a PDP is an appropriate route for the device, then upon submission of the complete PDP, FDA performs a substantive review. It is at this stage that panel input will be sought. There will be a total review time of 120 days by which time FDA must approve or disapprove the PDP, or Product Development Protocol. You can see we're not actually reviewing data. We're reviewing protocols.

[--- Unable To Translate Graphic ---]

FDA is currently working on the details of how panel input will be accomplished, given that the presentations of sponsors' PDPs must occur in a closed session to allay concerns over the release of proprietary information in an open public forum.

Following approval of the PDP, the applicant conducts their preclinical protocols and develops their bench and animal data as described in their PDP. They also report to FDA in the form of regular progress reports as stipulated in the PDP.

As the PDP will again stipulate, the clinical phase of the PDP can commence following completion and submission of the appropriate preclinical data, and, again, regular progress reports as defined in the conditions of that company's PDP will be submitted to FDA for review. As the company progresses toward completion of the clinical protocols, inspections for conformance to good manufacturing practices or the new quality assurance regulations, as well as bioresearch monitoring regulations, will take place.

It is, of course, anticipated that sometimes not everything laid out in the original PDP will work out as planned, and modifications of device design or testing protocols or results may be necessary. These will be

[--- Unable To Translate Graphic ---]

reviewed as either substantive or non-substantive changes, which will require notification to FDA, and these may occur through meetings or progress reports. FDA review of these modifications will be accomplished within either 30 or 60 days, depending on whether the change is substantive or not substantive.

When all trials have been completed and all progress reports submitted, the applicant submits a notice of completion to FDA, and this must be reviewed by FDA within 90 days. At this time, if FDA concurs with the company that all protocols have been completed and the results are as specified in the PDP, FDA will declare the PDP complete and the product may go to market as if a PMA had been approved.

In conclusion, it is anticipated that PDP may work best, at least in the beginning, for class III devices which are not first of a kind and for those for which FDA guidance has been developed. However, it is intended that eventually PDP will be of great assistance to the rapid development of innovative devices because it should be less expensive than the conventional two-step investigation and premarket approval procedure.

The phoenix you see here represents the efforts of

[--- Unable To Translate Graphic ---]

FDA to review this long dormant provision of the Food and Drug Act. Again, I remind you that PDP is still a work in progress and that all comments from you are welcome. And for your information, there is a PDP workshop which FDA is planning now, which will be on October 22nd, to inform interested members of industry and our advisory panels on this initiative.

Thank you very much. Also, you may be interested to know that there are T-shirts and hats with the PDP logo on sale out in the lobby.

[Laughter.]

DR. HARVEY: We would be happy to entertain any questions, Dr. Yin or myself.

DR. LEVY: Elisa, where during this process is labeling addressed or reviewed?

DR. HARVEY: It should be at the point of the 120-day review time where all the protocols and ultimate labeling are submitted. Granted, there will--we recognize that it may not be the final ultimate version, and that's why there are provisions in the system for modifications.

DR. LEVY: It just seems like from a review standpoint, to review the labeling before any data has been collected is a difficult task to ask us to do.

[--- Unable To Translate Graphic ---]

DR. HARVEY: Dr. Yin, would you like to add to that?

DR. YIN: I think the most important part in the early days is the intended use or indication for us and with the patient population. You are right. We will not be able to put out adverse effect or precaution or warnings. You are absolutely correct. But during the clinical trial, that's the time that they can send in to us progress reports. I think you are absolutely correct. We may not demand a full-blown proposed labeling in the early stage in the protocol. That's a very good point. Thank you.

DR. LEVY: But then the panel never really looks at it again after that initial--

DR. YIN: The panel would set the criteria, pass and fail the criteria. So if you are not comfortable--and you may say that you'd like to look at it--you are able to address that.

DR. DIAMOND: Elisa, as I understand it, both the company and the panel will be asked to review basically a concept and come up with a clinical study design, including inclusions and exclusions, before any animal data or any in vitro data is available. Is that correct?

DR. HARVEY: It's possible that it could be that

[--- Unable To Translate Graphic ---]

way. In reality, of course, we realize that most companies have a certain amount of data before they move forward as they would with a PDP, so that they may well have animal data and be somewhere along that time line that we looked at before they actually submit a PDP, so that there may be data there that gives them information to know how they may want to move forward with their clinical studies when they apply.

DR. YIN: The important part is the concept. You are right. If when they are developing the proposed clinical study they would say that in this animal study, if it passed, then we would anticipate that. However, we would entertain feasibility study early in the game so they can modify the protocol accordingly.

DR. DIAMOND: Would FDA--I would think that companies would want to have some idea of how likely something is to be successful before they want to bring it to FDA, because they don't want to be embarrassed time and time again, bringing things to FDA when it's in the concept form, and then as soon as they go to the lab or to the bench, they find out it doesn't work. So I would think they would probably want to do some sorts of preliminary studies to have some idea that, yes, this has a chance of working before they bring it forward.

[--- Unable To Translate Graphic ---]

If that were to happen, does that put them outside of the PDP process if they have already done some preliminary work to have a little greater confidence in what they are hoping to accomplish?

DR. YIN: No. See, right now we are even willing to entertain the PDP process even though the idea is going on now. There is--a great advantage is that if we do the review piecemeal, we get it done earlier. So by the time they completed the whole study, as you see, they only have 90 days to review, so there is a great advantage. Of course, there are disadvantages also. Let's not underline there are only pros. There are cons, too.

If the product turns out not as good and they did not meet the contract or the binding protocol, they may not get approved. But now they are in PMA, they will come in and talk to you guys and show you the data and try to negotiate. So this one here is a little bit harder, but there are certain advantages.

I think you are absolutely correct. This type of process you require a company to think it through, all the way through from the beginning to the end. And the advantage of getting the clinical protocol in mind that early is that to get in with the clinical people to

[--- Unable To Translate Graphic ---]

determine what should be the endpoint, or sometimes you may want to pick certain surrogate endpoints, so then you could decide that and then bring it to FDA, rather than come to FDA and say, What do I do now? So they may engage some of the clinical people in the medical school or wherever to advise them ahead of time, because the most important part is what should be the endpoint.

If you know the indication for use, you've got to know what should we look for for the clinical endpoint, and that will be so helpful if they worked it out and bring in the protocol ahead of time.

DR. HARVEY: Thank you.

CHAIRMAN EGLINTON: And we'd like to point out Dr. Davey has joined us since the rest of the panel members have already introduced themselves. Dr. Davey, could you introduce yourself to the group here?

DR. DAVEY: Diane Davey. I'm from Lexington, Kentucky. I'm on the Pathology-Hematology Devices Panel.

CHAIRMAN EGLINTON: Thank you.

Colin?

MR. POLLARD: Thank you, Dr. Eglinton.

Good morning again. We're going to now move to the main agenda item for this morning, which is the guidance

[--- Unable To Translate Graphic ---]

document for the new types of intrapartum monitors.

I know some of you, anyhow, recall a year ago last July we convened the panel to help us look at a new technology application, namely, intrapartum monitoring using fetal pulse oximetry or continuous fetal tissue pH. As you know, these devices are on an IDE-PMA regulatory track.

We invited three guest speakers, and we had input from several manufacturers and researchers active in this area. The panel was augmented so that there were five perinatologists working with us together with the three perinatology guest speakers. Although a great deal of information was presented that day, we had a rather formidable task of analyzing that information and sorting it into a guidance document that would be useful to sponsors who were developing this technology.

Following that meeting, we formed a small working group within the center and added one important participant from FDA's Office of Women's Health, Dr. Smith, whom you have just heard from on another, unrelated matter. And we developed an initial draft guidance document, circulated it for comment from the panel perinatologists and our guest speakers last fall, and a new draft was formulated. This time we have gone more formally public. A copy of it is

[--- Unable To Translate Graphic ---]

available at FDA's hearing clerk, and you can get it also from FDA's home page on the Internet as well.

We have sent copies to everyone who participated in last year's meeting. Technically speaking, it is available for comment, with the 90-day comment period that began June 14th.

Today, two members of the working group will give you a quick once-through of the document, highlighting what we believe are key aspects. Kathy Daws-Kopp, an electrical engineer in the branch, will capulize the pre-clinical portions of the document. After that, Debbie Smith, an obstetrician-gynecologist from our Office of Women's Health, will go through the clinical study requirements that we have proposed for this new technology application.

You will have a brief opportunity for questions and comments this morning. You may also send your comments to us later after you have had a chance to thoroughly consider its implications. We will hear from--there also will be a short opportunity for affected companies to comment as well. We hope to finalize the document before the end of the year.

Kathy?

MS. DAWS-KOPP: Good morning. I'm here to present

[--- Unable To Translate Graphic ---]

the preclinical portion of the intrapartum continuous monitors for fetal oxygen saturation and fetal pH guide, as Colin said.

I would like to remind everyone again at this time that we will be entertaining comments and suggestions for this document until mid-September. Please provide your comments in writing.

First, I would like to mention the people who have worked on this document. Dr. Smith wrote the first draft of the document last fall, and Dr. Weininger provided a lot of the preclinical text and format and has been instrumental in getting this document to its current state. Mr. Kotz and I, as well as other members of the OB-GYN Devices Branch, have also contributed.

This document covers the following general areas: Inasmuch as possible, we have defined what we expect for the IDE-PMA process for these types of devices.

As stated in the introduction of the guide, the purpose of fetal surveillance includes timely recognition of the risk for or presence of fetal acidemia. Thus, appropriate intervention can be initiated. In the United States, fetal heart rate monitoring is used almost universally as the standard for fetal assessment. However,

[--- Unable To Translate Graphic ---]

while a normal fetal heart rate pattern is a good predictor of a normal fetus, an abnormal pattern has limited predictability of fetal outcome. This leads the way for development of other technologies to improve clinical management of patients. Such technologies that have come forward to us are fetal oxygen saturation and continuous tissue pH. We have tried to address both of these devices in this document, but we concentrate more on oxygen saturation.

As I have said, I will be discussing the preclinical portions of the guide, while Dr. Smith will cover the clinical portions. As such, I will address device description, theory of operation, validation, and the non-clinical portion of the performance requirements section.

In device description, which starts on page 2 of the guide, we have outlined what we believe to be the basic areas of description that will provide a complete picture of a particular device. As shown here, we expect that that would include identification of major external interfaces, by which we mean clinical human interfaces such as those which are patient contacting, power requirements, communication interfaces, by which we mean communication

[--- Unable To Translate Graphic ---]

with other pieces of hardware or other devices such as maternal monitors, and assembly drawings.

There is more under device description, including some standard PMA-type required elements: labeling, design process, manufacturing, sterilization, and system effectiveness. This last one would include any system effectiveness studies that have been done, such as reliability, life expectancy, maintainability, et cetera.

Our theory of operation section on page 4 of the guide addresses how the device works and testing that is done to verify that it operates correctly. Signal acquisition and interpretation are about how the devices gets the signal, converts it into something usable, determines the value of the signal as well as what the device expects to see in a signal.

Under verification plan and test results, we have included testing with both animal models and on the bench. The information provided here should show that the device performs as intended. This can be done with a comparison of animal and bench testing that shows a correlation between oxygen saturation as measured by the study device and oxygen saturation as measured with a co-oximeter as the gold standard.

[--- Unable To Translate Graphic ---]

Validation, on page 6 of the guide, is the final testing of the system and occurs prior to use of the device in human study subjects. This testing addresses the operation of the system as a whole, and it can also address effective design changes and interaction with external interfaces.

Non-clinical performance requirements address the following items: the description of the intended use environments includes such things as temperature, humidity, electromagnetic compatibility, and electric safety. Human factors analysis and materials/toxicity analysis are also required.

Section 812.20 of the regulation outlines IDE application requirements. The company must submit an IDE as use of this device in human subjects constitutes a significant-risk study.

Now I would like to turn the floor over to Dr. Smith, who will discuss the remainder of the document.

DR. SMITH: Thank you, again. It has been an education as well as a pleasure for me to participate with the OB-GYN devices group on the preparation of this document. We definitely anticipate your comments and already have some modifications of our own in mind.

[--- Unable To Translate Graphic ---]

The presentations and discussion at the panel meeting of a year ago that Colin alluded to have proved very useful in the development of the clinical studies portion of this guidance document. For the intrapartum human observational studies, I would like to summarize the primary assumptions that underlie the draft document.

There are two clinical assumptions. The first is that the physiologic stress of labor is such that intermittent relative fetal hypoxia is the norm and that it is associated with a progressive reduction in fetal pH, PO<sub>2</sub>, bicarbonate, and an increase in PCO<sub>2</sub> and base excess. Most fetuses have adequate reserve and are born without any acute or long-term sequelae.

The second clinical assumption characterizes the significance of fetal distress as an indication for delivery by cesarean section. Fetal distress has been reported as the indication for operative intervention at a rate as high as 45 percent. Fetal distress is typically defined as some significant and persistent abnormality of the fetal heart rate. As previously noted, these abnormalities have a low specificity for hypoxia and acidemia in the fetus, which are the physiological problems that the intervention is seeking to contravene or prevent.

[--- Unable To Translate Graphic ---]

There is a regulatory assumption to our guidance document construction, and that is that diagnostic devices must provide more than physiological information as a demonstration of clinical utility.

It should be further stated at the outset, again, as Kathy noted, that this document is meant to apply to both continuous fetal tissue pH monitoring as well as fetal oxygen saturation monitoring, although the latter is primarily referenced.

The first series of clinical studies in humans anticipated by the guidance document are clinical reliability and accuracy studies. These studies are for the purposes of profiling the range of values of SpO<sub>2</sub> or tissue pH in normal labor in term singleton fetuses and to establish the accuracy of the system during the conditions of labor.

Clinical performance studies to assess the performance of the device in discriminating, non-reassuring fetal heart rate patterns logically follow. Protocols for these studies should include an appropriately referenced standard for interpretation of fetal heart rate patterns. They should also include a comparison methodology for acid-base determination. It is expected that newborn

[--- Unable To Translate Graphic ---]

assessments will be obtained as part of the database for the analysis of these studies.

Demonstration of clinical utility is required of all devices subjected to a PMA review. In the case of the continuous fetal oxygen saturation devices and fetal tissue pH devices, we believe that an intervention study is required. The intended use of these devices is to improve the diagnostic value of current intrapartum fetal assessment using electronic fetal monitoring.

Since the identification of the particular condition of fetal distress due to significant hypoxia or acidemia represents a continuum of intrapartum evaluation, clinical action, and neonatal outcome, a study that addresses an impact on intrapartum care is deemed appropriate. We feel it cannot be assumed that the adjunctive use of the devices will have a better discriminatory function.

Assessment of preemptive obstetric interventions, as would be subjected by the second bullet in the slide, are not tenable at this point in time due to the limitations of the knowledge base for antepartum and early intrapartum abnormalities and the length of time required for follow-up for developmental impact. These were issues that were

[--- Unable To Translate Graphic ---]

discussed very extensively last year at the panel meeting.

Therefore, in this guidance document, we anticipate the submission of the results of clinical trials utilizing these new monitoring techniques in combination with fetal heart rate pattern monitoring and the impact on the rates of obstetrical interventions, specifically C-section, and impact on early neonatal outcomes.

I'd like to close after this brief description by thanking you in advance for what we know will be thoughtful and useful comments on this guidance document.

CHAIRMAN EGLINTON: Thank you.

Now we will have time for industry comments. We have on the agenda Dr. Michael Ross from Healthdyne or representing Healthdyne.

DR. ROSS: Good morning. I'm Michael Ross. I'm the Chair of Obstetrics and Gynecology at Harbor(?) UCLA. I'm here representing Healthdyne Technologies. I don't know if Colin made copies of--yes, okay. I am just going to basically read from that, and I will leave time, if people want to find the section that I'm referring to.

We have reviewed the draft document and appreciate the careful thought and expertise which contributed to this document development. We appreciate that the document

[--- Unable To Translate Graphic ---]

details both suggested requirements and additional possible study considerations for the PMA approval of fetal oxygen saturation or fetal pH monitors. However, our review of the document resulted in several items for which we request clarification.

On page 5, C.1.c).(1).a., which took me more time to figure out than the rest of my preparation here, under the heading of threshold oxygen saturation which correlates with onset of metabolic acidosis: Assuming that the proposed oxygen saturation device is demonstrated to accurately measure O<sub>2</sub> saturation in both animal and human studies throughout a range of saturation values, including both normoxia and hypoxia, we propose that previously published animal studies, which include direct fetal arterial blood sampling, with or without pulse oximeter O<sub>2</sub> saturation studies, be utilized to establish the threshold oxygen saturation which correlates with the onset of metabolic acidosis. In effect, they are asking that in order to determine the threshold for animal studies that we use established animal study protocols with direct blood sampling rather than repeating another series of animal studies using non-invasive technology, although we would demonstrate that the non-invasive technology correlates with

[--- Unable To Translate Graphic ---]

the direct readings.

Item No. 2 on page 8.1, under Neonatal human observational data: As neonatal transmission oxygen saturation devices are already FDA-approved, would the panel accept a study of the validation of scalp O2 saturation devices in neonates utilizing approved transmission oximetry devices as the comparison standard rather than invasive blood sampling in these neonates? That is, in effect, part of the validation of the scalp oximetry would require measurements in human neonates, but we wanted to minimize the exposure of the neonates to repetitive blood sampling, and so use another established standard.

Also on page 8.2.--this is Item No. 3 on mine--Clinical reliability: Is the panel requiring studies of the effect of maternal-inspired O2 supplementation and regional analgesia or only requesting consideration of these effects?

Item No. 4, once again, on page 8.2.B)., the Clinical accuracy: Would the panel accept dual sensor studies in animals rather than humans for demonstration of reproducibility?

Item No. 5, page 10, Clinical efficacy: In the control group of a clinical randomized study--I'll wait

[--- Unable To Translate Graphic ---]

until you get to that. Once again, in the control group, that being the fetal heart monitor without oxygen saturation, of a clinical randomized study, does the panel have an opinion as to required or optional use of fetal evaluation techniques such as scalp blood sampling, acoustic stimulation, scalp stimulation, and biophysical profile? As the proposed use of the oxygen saturation device is in part to avoid scalp blood sampling and the alternative techniques--those being the acoustic stim, scalp stimulation, biophysical profile--remain controversial, we propose that the control group be evaluated by the fetal heart monitor only.

If not, if the panel suggests that an alternative technique be used in the control group, what are the requirements for that? Is it any technique? Is it at the judgment of the physician? Is it option or elective?

Page 10--once again, my Item 6--again, under Clinical efficacy: In the utility of a threshold value--and arbitrarily defined here as a 30 percent saturation--will the panel accept the use of physician judgment and interpretation, similar to our interpretation of fetal heart rate monitors, in a sense--so some art in this interpretation--as to the time and degree of fetal oxygen

[--- Unable To Translate Graphic ---]

saturation values below the threshold as compared to the time and degree above the threshold for decisions requiring intervention? Or will the panel require an absolute definition that could be used by a physician? And I list here just an arbitrary definition, 5 seconds below a 30 percent threshold following greater than 50 percent of contractions. So is the panel saying that there must be an absolute definition, or is there some physician judgment also in the interpretation of these threshold values that will be permitted, just as there is judgment in the interpretation of a fetal heart tracing?

My Item No. 7, on page 10.3.B),(1): Why did the panel suggest a gestational age greater than 36 weeks? Is there any suggestion not to consider this in pre-term infants?

Then, finally, my Item 8, on page 12.J),(8), in regard to Receiver Operator Curves: We propose that the Receiver Operator Curves may be useful in animal studies for the prediction of the development of acidosis but may have limited utility in human clinical studies. We propose that a single threshold value be utilized in human clinical studies for intervention decisions rather than a series of values which would be required to obtain an ROC curve.

[--- Unable To Translate Graphic ---]

Thank you once again for your time.

CHAIRMAN EGLINTON: Are there any other comments?

We're shuffling papers. Hold on a second.

Is there any further general discussion, questions, comments from the panel, discussion points? Michael?

DR. DIAMOND: Yes, I have several issues, in going through the document, that I wanted to clear up in my mind as to what it was that FDA was suggesting. I guess I'll just start at the beginning and go forward.

If you look at starting on page 8, the very bottom of that page is the inclusion criteria for who should be included in the studies, and I, too, was wondering why this was going to be limited to fetuses 36 weeks or greater. I would think in the long run the OB community would like to utilize these devices in premature infants as well. And if we only accumulate data on term infants, we will have no idea as to what the value is earlier, where perhaps even more crucial decisions sometimes have to be made. So I would have thought that we would want to extend the age limits and not have that limitation.

Going on to (b) at the top of page 9, I guess I began to come up with some technical problems for conducting

[--- Unable To Translate Graphic ---]

these studies in that there are all sorts of non-reassuring fetal heart rate tracings that are listed, I guess it's eight different types. And to have a study that is going to allow to have--you are either going to have to lump them all together, or you are going to have to have a large number of women in order to have enough numbers in any one of those categories to make your evaluations.

I also have the problem that in order to get informed consent to participate in the study, whatever the ultimate design turns out to be, you are probably going to have to enroll these women at the initiation of their labor as opposed to when all of a sudden you have someone with severe variable decelerations. It's not a very good time to be getting informed consent from a patient to be participating in a study.

I also was unsure why there were certain medical conditions that were going to be excluded. Again, I think in the long run you are going to want to utilize this technique in women with diabetes mellitus or sickle cell anemia, and to say from the beginning that they are going to have be excluded from the protocol of anyone that would want to come up with these devices I think is excluding a patient population which you really want to study.

[--- Unable To Translate Graphic ---]

I'll stop there for a second and see if anybody has...

CHAIRMAN EGLINTON: Is there anybody who has been involved in the development of the document who would like to comment on any of these? A similar comment would be exclusion of those with ominous or abnormal fetal heart rate tracings since that is so subjective. That might be an ideal candidate to be studied with this other ultimate technology.

DR. DIAMOND: Actually, how does that differ from what's above in (b)? Being a reproductive endocrinologist, primarily, I guess I can ask that question. But I seem to remember from years ago that a lot of those things above may be in this ominous or abnormal fetal heart rate tracing category. So I think they're mutually contradictory.

DR. SMITH: I could make a couple of preliminary comments that may not be specific to every single issue but reflect some of our thinking and some of our consideration of what we thought was the discussion on this a year ago and the interval comments that we have received from various members of the panel as well as consultants.

I think that we would all generally understand and accept that there would be over time interest in the use of

[--- Unable To Translate Graphic ---]

these devices in your high-risk obstetrical situations, whether that's maternal high risk, a certain medical complication like diabetes, or, for example, in the premature infant.

I think what our preliminary response to that would be is that in the course of then planning to bring this type of device to market, if those, in fact, were to be some of the intended uses, the kinds of clinical situations that were thought to be of value, that we would clearly then have to see those things represented--represented systematically and represented in a significant way with a significant number of patients in the clinical studies, in the pre-approval studies. And I think that some of your comments and those of Dr. Ross certainly confirm why we think that pre-marketing intervention studies are necessary.

In terms of an issue like the numbers of categories of non-reassuring fetal heart rate patterns, the alternative to having so many would be to try to develop consensus on one or two to be the only ones that would be the circumstances in which one would actually utilize the device in an adjunctive way, and that probably would defeat the purpose of expanding the variety of clinical situations, both either for the fetus or for the mother, that one would

[--- Unable To Translate Graphic ---]

be--

DR. LEVY: But our point here was to have one category that we studied that was called non-reassuring rather than dividing it up into eight different--

DR. SMITH: Well, these were giving examples of--not necessarily to constitute separate arms of the study per se.

DR. LEVY: Right.

DR. SMITH: But to give examples of what would be in this category. Obviously, as we identified the protocol, elements have to include appropriately referenced standard, institutionally based or otherwise, for interpretation and characterization of these types of abnormalities or any others.

I think we definitely appreciate the sentiment that you expressed about the timing of informed consent, and certainly in any clinical studies that require--that have an intervention related to clinical utility and certainly intrapartum ones, we are confronted with this issue all the time. And the decision that we come up with is never satisfactory to all, but we certainly have, as a matter of policy, that the appropriate--the informed consent, the timing of it, and the nature of the informed consent should

[--- Unable To Translate Graphic ---]

meet all standards and requirements and should be appropriate. And so if, in fact, the document needs to speak more specifically to that, then we certainly would take that under advisement and be happy to receive more input on that.

CHAIRMAN EGLINTON: Dr. Yin?

DR. YIN: I do appreciate all the comments that come in, and, Dr. Diamond, your comment is very apropos. However, we are thinking of the companies, what they'd like for the indication for use. Because if we do require those subset of patients, you know that the numbers will be very large. So it is very difficult for the study. But if that is what the company would like to claim, that is reasonable. But then to demand that, that will really delay maybe the marketing of this product. That is what we are thinking in terms of in generalities; rather, it's that you must do everything. And I like what Dr. Ross suggested of pre-term. However, if that's not what the company wants to do, you know that, again, is a big set of patients, and the criteria would be rather difficult and different. So maybe that is why we are proposing it in a cleaner study at this time.

But I think you are right. In the long run, that

[--- Unable To Translate Graphic ---]

would be good for it. But sometimes the company would like to go on stage, so I like your comment.

DR. DIAMOND: I guess if I could make a suggestion regarding those issues, it would be that I think you are very right to separate term from pre-term and fetuses with certain medical conditions. But rather than saying, as is stated now in the document, that they have to be over 36 weeks or they have to have--or you are specifically excluding certain medical characteristics, maybe the guidance document should say that the company should specify the age range or the maturity of the fetuses they want to look at, and specify whether to include or exclude certain co-existing medical conditions as opposed to a priori placing into the inclusion or exclusion criteria and then, as you say, giving the company the option as to which to include, which to exclude, giving them the maximum flexibility.

DR. YIN: Good point. Thank you.

CHAIRMAN EGLINTON: Along that line, if there is no good physiological justification for discriminating, for segregating the population, if none of us, nobody in the FDA, none of the panel members, if nobody has any good justification for segregating the population, we really

[--- Unable To Translate Graphic ---]

should not try to segregate the population.

DR. SMITH: I think the question we would ask is for you to then give us input further, as has come from all the perinatologists thus far, is whether or not, for example, there is physiologic discrimination between a 32-week fetus and a 38-weeks fetus in terms of the acid-base physiology in labor. Does the 32-week fetus have the same, quotes, normal acid-base experience during labor as does the 37-week fetus? If, in fact, that is the case and if, in fact, we know that and can compare baseline information, profile information on the normal physiologic experience of a 32-week fetus as compared--then clearly there would not be a need to segregate or disaggregate in the clinical trial setting.

If, in fact, we don't know that yet or if, in fact, there is information to suggest a difference, then it would seem reasonable at the outset to have definition, not necessarily an ultimate exclusion or an ultimate bias against, but to have this initial definition. So we would be happy to review all that data again with you and work on that particular question.

CHAIRMAN EGLINTON: I am also thinking a couple of steps downstream. If a study is proposed that is severely

[--- Unable To Translate Graphic ---]

limited, it has only a very tightly defined, very normal population to demonstrate clinical safety and efficacy for a device. But the device is going to show up in labor and delivery after it's approved, it's going to show up in labor and delivery units all across America, like fetal heart rate monitors; it's going to find its way on to the scalp of every baby in labor and delivery, every fetus in labor and delivery, once it's there, just like the fetal heart rate monitor. It won't be labeled as such, but that's the reality. That's what's going to happen.

Is there any other commentary?

DR. DIAMOND: I'll let Michael go. I don't want to monopolize things.

DR. NEUMANN: I want to go to a different area, namely, some of the issues that Kathy brought up. One concern that I think is between the lines, but perhaps not as clearly stated as it could be, is the issue of the interface between the device, whatever it is, and the fetus. As we know from fetal scalp electrodes, there is some morbidity associated with it, and I would think that this should be indicated, especially in the case of the pH sensors, which I believe are also skin penetrating. And I'm just wondering if it is clearly enough stated that we need

[--- Unable To Translate Graphic ---]

that kind of data as well in the document.

MS. DAWS-KOPP: Yes, I think we would agree with that, and we'll note to add some more information on that.

DR. WEININGER: I'd just liked to add briefly that--

CHAIRMAN EGLINTON: Can you please identify yourself when you come to the microphone, just for the transcriptionist, who isn't looking at the videotape.

DR. WEININGER: Sandy Weininger, FDA. I would like to add that being that most of us are engineers who tried to contribute to develop this document and tried to learn as much as we can about the clinical applications, we're not clinicians and we really need your help, as Dr. Neumann has said, to identify what are the areas where the major risks are occurring, so that we can include them and we have the manufacturers address those issues in this document. In fact, when you read the document--or you've already read it--you read the entire document as if you were trying to identify where the major risks are concerned, and please tell us.

DR. LEVY: Okay. So along the lines of Dr. Neumann's comment, one of the exclusions should be contraindication to invasive monitoring, and that just has

[--- Unable To Translate Graphic ---]

to be specifically stated there as an exclusion.

CHAIRMAN EGLINTON: Dr. Diamond?

DR. DIAMOND: I have I guess what comes down to a fundamental question to pose as well, which is that the document, if you look at page 11, talks about a study design with two arms and with controls. And I guess the question I would have is in a situation like this, where you have endpoints that you can look at, which would be other than biochemical type endpoints but which will be fetuses which will have physical conditions and Apgar scores and other more accurate ways of assessing well-being, why you cannot utilize--you have your standard fetal heart rate tracing, whatever you normally do in labor, and then you are adding to it whichever one of these technologies you have. And at such a point, going along with your standard technology and that tells you you want to go ahead and intervene, and now you go ahead and you add in, whether it's your continuous pH monitoring or oxygen saturation, and then look at the fetuses that come out as a consequence--why that is not a possibility, not necessarily the only potential design, but why that is not also a potential possible design for these studies. And that way you would get around some of the problems that were alluded to earlier, with having all the

[--- Unable To Translate Graphic ---]

different types of non-reassuring heart rate tracings that you can have to make sure you have equal distributions of those or equal ominousness of tracings in the control and the treatment groups. And I would think that would be a parallel design that would have value to be allowed to be included.

MR. POLLARD: Mike, maybe I might address this. I thought one of Debbie's overheads kind of captured it, but if I understand your question, you are saying a parallel approach would be essentially looking at fetal outcome.

DR. DIAMOND: Yes.

MR. POLLARD: Right, and I think we recognize that. It may not be coming out adequately in the guidance document, in which case we certainly can beef that up to be an option. I think you haven't seen much of the emphasis of the discussion go in that direction because basically from the meeting last July, I thought there was a fairly generalized sense that it would take rather profound study sizes to be able to show those kinds of effects, and that a much more practical, if you will, from a clinical perspective, a much more practical approach would be to look at the effect on intervention. But the approach you are proposing is something that is there, that is valid. I

[--- Unable To Translate Graphic ---]

think Debbie referred to it briefly this morning, and we can build that up a little bit.

DR. DIAMOND: I'm sorry. I thought I understood her to say that that second option on her last overhead was not tenable.

MR. POLLARD: Well, when she said not tenable, I presume she really meant that would have been a very, very large study.

DR. DIAMOND: Okay.

DR. SMITH: There we are talking about, as, again, was discussed very extensively last year, the issue being raised that what we see acutely intrapartum or even in the acute neonatal period does not represent--and we have a lot of lack of specificity and a lot of lack of predictability to whether or not that actually affects long-term developmental outcomes, and that--but that's what everybody agrees is really the most important thing to understand, that if one makes an intervention and, for example, subjects a mother to an intervention that brings with it its own morbidity, that one ought to know over--not only acutely but over the long term that it was worth it.

That led to the whole discussion about what the antecedents really are, what we really mean by asphyxia,

[--- Unable To Translate Graphic ---]

what the antecedents really are to it, what the long-term effects are, the cerebral palsy question, et cetera. That would be where the gold would be. But I think what we heard and what we understand is that we don't yet have adequate characterization of the manifestations of that altered physiology in the fetus, either in the antepartum, late antepartum period, such as with biophysical monitoring, and then in the early antepartum period such that one would actually intervene at a point where you could--it would be a preemptive strike and, therefore, you would prevent 6-year-olds not being able to color with their crayons and things like that.

So we focused on the issue of the intervention, which is the contravening--what we use to contravene insult, as well as the early and acute neonatal assessment. And the document definitely calls for having information on neonatal acid-base assessment and various clinical parameters. But we have stopped there in terms of what we see as an essential requirement. We certainly, again, as Lillian suggested, we would entertain any expansions beyond that for expanded indications or intended use.

DR. DIAMOND: I very much would agree with all your initial comments, but I'm not sure that they in and of

[--- Unable To Translate Graphic ---]

themselves, as far as really what our long-term goals are, speak to the issue of a one-arm or a two-arm study.

Regardless of whether you have a one-arm or two-arm study, the intrapartum event may not be the inciting event which determines what a 6-year-old is going to be able to do or not do.

DR. SMITH: Exactly.

DR. DIAMOND: And so that does not address the issue of one arm or two arms. The question is: Is it possible to get the information that you hope to get from a two-arm study, the way this is put together, out of a one-arm design? And if so, then would a one-arm design be appropriate?

If, for example, you were going to initially blind the clinician to the results of whatever the new device you were utilizing until they make a decision, yes, I am going to do a cesarean section, no, I'm not, and at the time they would decide they were going to, then at that time allow viewing of this data to see if this alters clinical management, and then just looking at outcome as a result of that, that may be a potential alternative design.

DR. SMITH: Well, these are certainly the kinds of considerations that we'd be happy to look at further. And

[--- Unable To Translate Graphic ---]

that type of comment was not one that we received in prior circulation of the document, both industry and within. Any and all comments and suggestions are worthy of consideration, and if you'd like to--I mean, we certainly take note of that now, but would be certainly interested to discuss it further and have it go around again for comment and review.

CHAIRMAN EGLINTON: One problem with such a design would be that had that design been used in the Dublin study or randomized fetal heart rate monitoring study, they would not have been able to detect a doubling of the neonatal seizure rate in the non-monitored group. That is really the reason to have a two-arm study for something like this. And to look at fetal or neonatal or childhood outcomes, I mean, remember the collaborative perinatal project of 50,000 mother-infant pairs trying to discover the etiologies of childhood neurologic dysfunction. Such a study will probably never be repeated again. It just can't. Nobody has that--God doesn't have that much money. So we are probably never going to have the opportunity to validate this technology in a way that we all know that we should because it's just not feasible. But that's why we argued last summer about, well, if anybody could at least come up

[--- Unable To Translate Graphic ---]

with something that would cut the cesarean rate for "fetal distress," that would be useful to the public.

DR. DIAMOND: I am not familiar enough with the Dublin study to discuss that at all. But with the sort of design I just described, if the use of a device at the time a clinician was otherwise going to perform a cesarean section would allow them to decide half the time or a quarter of the time or three-quarters of the time not to do that procedure, you could still look at the surrogate endpoint of C-section rate as an endpoint, and be able to look at those infants in whom you went ahead and did cesarean section when this device also said the fetus was in trouble as opposed to those in which you were able to hold off and go for a vaginal delivery and see if those fetuses have a good outcome. But you could still use C-section rate as an endpoint.

CHAIRMAN EGLINTON: Right. But then the problem would be if you fail to perform the cesarean because the SAO2 looks good, and then you got a whole nursery full of seizing little babies, that's not good. The only way to know that is to have a two-arm study.

DR. DIAMOND: Potentially, you'll still have that with the double-arm study as well. You won't have the

[--- Unable To Translate Graphic ---]

comparison, but you'll still have that with a double--with the two-arm study as well in the group that gets randomized to that device.

CHAIRMAN EGLINTON: Maybe we're not communicating. You randomize them to standard technology, which is everybody has an electronic fetal heart rate monitor. An alternate is use Instrument B as a supplement. And you look at all of them in the nursery. That's why you need two arms. If all of them with Instrument B, because they delayed the cesarean, wind up seizing their brains out, then you have an answer. You've lowered the cesarean rate to zero, but all the babies died. Well, you have to have a two-arm study. In the other arm, you have standard practice, and all the babies are fine, and the cesarean rate is 45 percent.

Dr. Ross?

DR. ROSS: Just to perhaps make a suggestion towards Dr. Diamond's study protocol, although I'm not suggesting a one-arm study, one might consider a single-arm study, all patients receive standard fetal heart rate monitoring and oxygen saturation is blinded throughout the entire protocol. Physicians manage the patients by their routine procedures, and then one retrospectively reviews

[--- Unable To Translate Graphic ---]

whether the O2 saturation cut-off, which would have to be a priori determined, would have discriminated between those patients, among all the patients who required a cesarean section, which of the infants perhaps truly needed it and which did not. And, furthermore, among the infants who did not require an intervention, did the O2 saturation confirm their reassuring status?

DR. LEVY: But then you wouldn't have the outcomes of those patients because you wouldn't know--had you not, for example, done the cesarean section and the O2 sat. would have said it was okay not to, you don't know how that baby would have done had you not done the cesarean section. I think that's what Dr. Eglinton was saying.

DR. ROSS: Well, you would know whether the--among the patients who would have a cesarean section--once again, I'm not proposing doing this, just for Dr. Diamond's discussion. Among all the patients who had a cesarean section, you have infants that are born in perhaps a state of somewhat compromise and other infants that are apparently well, and one could determine whether the O2 saturation would have predicted that differentiation.

Now, one would not know if you did not do the C-section whether the labor course would have gotten worse,

[--- Unable To Translate Graphic ---]

whether changes would occur in the patient or the baby. But up until that point of time, the O2 saturation monitor would have predicted the well-being of the fetus as determined by a prompt C-section.

So I think it has validity if one wanted to pursue that course.

DR. LEVY: While you're standing up here, I wanted to go back and address one of your other points, which was: Do we include things like acoustic stimulation and other things? I think it makes sense to include anything that clinicians are currently using so that we have a valid comparison between new technology and the current state of affairs. For that reason, even though it's a little bit less clean, I think that an arm should be whatever the clinician currently uses to make a determination on fetal well-being prior to making a decision, and those things should be included. That's, again, the opinion of someone who hasn't practiced obstetrics in 15 years.

[Laughter.]

DR. ROSS: I wanted to ask the panel to specifically discuss Item 5 and 6 on my questions, and you're addressing Item 5. I think that is a very sensitive issue. We're trying to avoid scalp pH sampling. The other

[--- Unable To Translate Graphic ---]

modalities are controversial both in their sensitivity/specificity and potential risks in regards to acoustic stimulation specifically. And the more those other techniques are utilized in the control group, it would be a fact that the more difficult it would be to show that the oxygen saturation monitoring will have a benefit over fetal heart rate monitoring along. So it makes for a larger and more complex study.

Furthermore, if you say we should use these other techniques, which of them should be used? Should it be at the discretion of the physician? It gets to be a very messy study design, and what are the criteria for interpretation of these somewhat controversial techniques? Although some of them I use myself.

So to keep the cleanest study and to show the benefit in relation to heart rate monitoring alone, which is still used at many, if not most, hospitals throughout the country, I would suggest that we use monitoring alone and not confuse it. I would appreciate some discussion on that.

CHAIRMAN EGLINTON: Well, one point of discussion might be, since USC has trained more perinatologists than any other single institution in the country and that's where the scalp stimulation or scalp clamp test came from, I think

[--- Unable To Translate Graphic ---]

probably you'd have a whole lot of people that argue about that. That just seems to have fairly clear predictive value negative.

DR. DIAMOND: I think another issue may be, though, depending on whether or not those ancillary tests--using that word--are allowed will determine whether you're going to be able to show a benefit or whether you're going to be able to show equivalence. If you utilize all the other tests that are available, I doubt that you're going to be able to show improvements with new devices. But the FDA may then need to be willing to accept showing of equivalence between the arms as opposed to the new device improving outcome.

DR. ROSS: Right. I would agree. If the perinatologist was to use acoustic stimulation, scalp stimulation, biophysical profile, and maybe a scalp pH, he or she can certainly determine fetal well-being in conjunction with the fetal heart rate monitor as well as O2 saturation. However, this device is intended for everyone from the perinatologist to the community obstetrician who may be in the hospital or not in the hospital, as labor is managed at the present time. And so we're trying to add an objective criteria rather than a perinatologist's

[--- Unable To Translate Graphic ---]

interpretation. And since the community standard is not to utilize each and all of these techniques, I believe that's what we're trying to address.

So, again, I would suggest not to confuse it with other modalities of testing but, rather, the indication is that it will add to fetal monitoring alone.

MR. POLLARD: Maybe I could just clarify that point. As I understand this document--and it may be there is some clarity issue that we can clean up; it is a draft document--I don't think we're trying to suggest that the study centers need to use any of those ancillary methods. I think the only thing that it does say is if you do use fetal scalp blood sampling, that you follow the guidance that was given on page 8, you know, to make sure that you get appropriate data pairs for comparison. But I don't think the protocol or the suggested protocol in any way requires any of those, so hopefully that's something that we can clear up just by straightening up the language a little bit.

CHAIRMAN EGLINTON: Sure. Dr. Solomon?

DR. SOLOMON: I'd just like to comment that I would hope that such a guidance document would have some sort of permanence, and we all know that there's an evolution in the clinical standard of practice. I think the

[--- Unable To Translate Graphic ---]

concept here is that you would compare whatever at that point in time is clinical standard of practice to the addition of the device. So that in the hopes of having some kind of document that wouldn't need to be modified as often as clinical practice is modified, I think that should be the overriding concept.

CHAIRMAN EGLINTON: And I think Dr. Ross' point is very strong. The more complicated the protocol gets, the more argument is going to result at the panel discussion.

DR. SMITH: Again, I think Colin read the language. It's on page 8 as the document is currently constructed, and it was in my overhead. We say a study should "provide a comparison to an appropriately referenced...clinical standard or protocol for the evaluation of..." So evaluation takes in your clinical interpretation and then--we would just need to know what do you do, what is it that you're doing, and what is it--now, this could make it complicated, but it also does not--we are not suggesting that you need do any of these particular interventions. But, again, a la some of the discussion that took place last summer, if, notwithstanding its limited use, fetal scalp pH sampling done here or done elsewhere, is the comparison interpretation, then we do need to receive

[--- Unable To Translate Graphic ---]

information in a certain way.

DR. LEVY: I think we may have an informed consent issue, too, with this in that if in a certain clinical environment the physicians are using ancillary techniques in order to determine whether they're going to do a cesarean section or not and we have a two-arm study to tell them they can't use those things for a patient, to me, as a patient advocate, that may be an issue. That may be a problem.

If we are going to do the one-arm study, it's less of a problem. In other words, if we get to add in the information from the monitoring devices, that's less of a problem.

CHAIRMAN EGLINTON: Then the sponsor has to do the study in an institution where clinical practice is such that it permits the protocol.

DR. LEVY: Yes.

CHAIRMAN EGLINTON: I mean, there were several institutions in the country that could not participate that could not participate in the entocin tocolytic drug studies because when the protocol went to randomized against placebo, a lot of people just wouldn't participate.

On the other hand, we all know you can go to Parkland and you can do tocolytic studies randomized against

[--- Unable To Translate Graphic ---]

placebo until the cows come home because that's what they like. So you just have to do the study in the right institution.

Dr. Ross?

DR. ROSS: Perhaps I can--I appreciate the discussion on my Item No. 5. Perhaps I could prompt the committee to discuss my Item No. 6, the issue of a fixed definition for intervention. And this is an issue that I don't know the answer to because we have not yet done the clinical studies.

But, once again, would the panel feel that we need a firm threshold that's fairly rigorously defined and strongly suggested intervention or permitting a physician judgment to a significant degree, again, akin to the interpretation of heart rate monitoring in terms of interpretation of the saturation, now continuous reading? Many, many data points will be collected and obviously printed out continuously or as continuously as possible, ultimately resulting in values that progress above and below a threshold determination, likely also with patterns of oxygen saturation changes, similar to what occurs in patterns of heart rate monitoring. Are those patterns suitable for interpretation or does one need some fixed

[--- Unable To Translate Graphic ---]

definition akin to a laboratory cut-off?

CHAIRMAN EGLINTON: My first thought would be you already answered your question yourself. The more complicated it is, the harder it's going to be to get agreement in a large group. We are in this box in part because it has been so hard to make interpretation of fetal heart rate monitor patterns objective. We all know how subjective they are. And if we have another instrument, another piece of equipment that's going to be applied to the fetal scalp and women in labor are going to be subjected to interventions or not interventions on the basis of interpretation subjectively of the output from this instrument, that would get--I mean, we have this background against which we can argue this point now, and I don't--I would be really surprised if some subjective sort of waveform, area under the curve pattern analysis would ever get through this process.

Does anybody--Dr. Neumann, do you have any idea on that?

DR. NEUMANN: Well, I certainly agree with what you are saying, but on the other hand, if we put on our scientific hats, any scientific protocol that has the, quote, physician interpretation in it or anyone else's

[--- Unable To Translate Graphic ---]

interpretation in it becomes very soft. And somehow or other, we have to find a compromise between these two things.

DR. DIAMOND: Gary, if I can address that also, I think if you set an absolute definition, there are always going to be biological situations, just as was being said, where you can have those criteria met and the baby will be fine or where the baby can be in serious trouble and will subsequently manifest those problems without having reached--by being one point above whatever guidelines you set. And that is in part why I was talking about the one-arm model.

Now, again, I fully agree with you, your point that the gold standard should be a control study, but there are problems such as this issue, such as all the different types of changes in fetal heart rate that you might see, and trying to then put all of them together as a gamish(?) to make a general conclusion where your control group and your treatment group may end up being different would, I think, complicate that.

I think, further, you have complications in a randomized comparison being chosen here with C-section rate as the endpoint when it is the clinician that is deciding to

[--- Unable To Translate Graphic ---]

the do the C-section and they know whether the individual has been utilizing this device or not. So does it have to be randomized, double-blind, or looking later to see whether or not this device was utilized? How do you control for all those things?

So in this situation, the control group is not, I don't think, an ideal way to go either. That's why I throw the other out not because it's perfect and not because it doesn't have problems, but because there are problems with a randomized control group as well for all those different reasons.

CHAIRMAN EGLINTON: Dr. Yin?

DR. YIN: I'd like to change the subject. How about his Question 8, this ROC curve using animal data? I'd like to hear some discussion on that.

CHAIRMAN EGLINTON: I may not be remembering the discussion from last summer accurately, but I think the point was the ROC curve will facilitate choosing the right number. I mean, it might be 30 percent. It might be 28 percent. But that will facilitate the choice of that number. And then it would be objective. If the SAO2 is below 28 percent and the heart rate pattern looks bad, okay, deliver the baby abdominally, or whatever your intervention

[--- Unable To Translate Graphic ---]

is.

I have some nods of agreement that that matches somebody else's memory. Is that--I mean, it may be that--Dr. Ross is behind a post, so I can't really see him. But it may be--

DR. ROSS: I'm hiding.

CHAIRMAN EGLINTON: It may be that it has to be something more complicated than just a number, 27.6, you know, or multiple of a median of 2.5. Maybe it's more complicated than that. It's this and this and this and this. But I think it has to be--in the end, it has to turn out to be something objective.

DR. ROSS: Right, and I appreciate that. Once again, I don't know the answer to this. I think as simple as the industry can keep it and as objective. Nevertheless, because it's continuous values between contractions, during contractions, changes over time, it's going to be a challenge to provide that definition rather than an absolute single threshold.

I agree with--the Receiver Operator Curve, my intention was to clarify whether this was being requested to be performed in human studies; rather, I think not. I think it's more determining threshold perhaps from animal

[--- Unable To Translate Graphic ---]

investigations with confirmation of human data and then utilizing a definition in either a single-arm or double-arm human study.

CHAIRMAN EGLINTON: Dr. O'Leary had a comment, and then we have one from the audience as well.

DR. O'LEARY: Yes, the comment is on objective. I think that facilitates analysis, and, you know, when you start using fuzzy criteria for entry, you have trouble analyzing your data, the question of two-arm versus one-arm, the information that you're trying to get out of your one-arm can probably be extracted from a post hoc analysis of a two-arm study quite effectively, giving you any advantage that you would see in the one-arm, plus all of the advantages of the two-arm study.

So I think from a study design and analysis perspective, you get better information to make a set of objective criteria of determination of what, practically speaking, almost always has to be superiority because proving equivalence is a wonderful legal term, but it's a lousy statistical term. And you're almost always looking for something that's a little bit better to prove equivalence.

CHAIRMAN EGLINTON: In the audience, sir?

[--- Unable To Translate Graphic ---]

VOICE: It was covered.

CHAIRMAN EGLINTON: Okay. Can anybody tell any jokes or maybe do some tap dancing? We have an embarrassing interlude here because we can't really start the afternoon session until the published time because there may be people coming--yes, ma'am, please?

MS. YOUNG: Being very new, I was following the time and the agenda. I see that panel comments start at 11:30, so I was waiting.

I'm very new to this, so I sort of preface my comments by saying that not having been involved in any of the previous discussion on these particular devices, I have got some written comments which I provided a week or so ago, and I do have copies of the articles to which I refer because in this draft document the references--well, there have been quite a lot of other articles and studies in the literature. It just so happened that I actually had in my files some of those new studies.

On the whole subject of intrapartum fetal surveillance, it's been something I've been interested in for a long time, and I've been involved in a lot of the controversy about the use of electronic fetal monitors, about the cesarean section rate, and so on as well. And in

[--- Unable To Translate Graphic ---]

looking at this document, I should say that it's interesting to see--I know that one doesn't necessarily consider the standards of practice and clinical practice as conducted in other countries. But it's been interesting to see where Canada has gone as far as the use of electronic fetal monitors are concerned. And as I mention in my comments, the use of electronic fetal monitors in intrapartum fetal surveillance is now not necessarily considered to be the standard of practice. And, in fact, in there, the latest guidelines from the Society of Obstetricians and Gynecologists of Canada, they, in fact, recommend the use of auscultation over electronic fetal monitors as a standard practice for labor.

Looking down the road in terms of clinical practice and so not thinking about the studies that we've been talking about this morning, if the assumption is in this document, this draft document, that electronic fetal monitoring is still the standard practice, if, in fact, that changes and it looks as if--in fact, I would argue that auscultation and electronic fetal monitoring, at least as far as the American College of Obstetricians and Gynecologists is concerned and the American Academy of Pediatrics is concerned, they can be considered to be

[--- Unable To Translate Graphic ---]

equivalent procedures.

So down the road, if auscultation is considered more to be the standard practice and if these devices, this new device that's under consideration is, in fact, approved by the FDA, what is going to happen when auscultation is used more than electronic fetal monitoring and the clinician picks up something that is a complication or a problem? Is the woman going to be sort of rushed to the electronic fetal monitor, going to be hooked up to that, and then this technology is going to be used as an adjunct just to the electronic fetal monitoring, which is what I understand from this document?

What happens if auscultation is going to be used more frequently? It's used in free-standing birth centers now. Electronic fetal monitors are not used there. So I just want to know, how this technology sort of fits into my other sort of scenario, which we don't know is going to happen but could happen, in fact?

DR. ROSS: I don't know if you're addressing that to--

MS. YOUNG: Open.

CHAIRMAN EGLINTON: Dr. Ross, we'd be delighted if you could--

[--- Unable To Translate Graphic ---]

[Laughter.]

DR. ROSS: I'll look into my crystal ball.

I think it's an excellent question. There certainly is a controversy regarding electronic fetal monitoring versus auscultation performed intermittently. However, the American College of Obstetricians and Gynecologists tends to view intermittent auscultation as the relative equivalence only when used in a one-to-one patient-to-nurse relationship, and that's probably not the standard in the vast majority of institutions in the United States. So I think we remain perhaps in a two-to-one, or thereabouts, ratio of patients to nurses with the standard being electronic fetal monitoring.

Were the standard to change in time, one would have to address how to utilize both electronic fetal monitoring and pulse oximetry, but I don't have the answer to a very good question.

CHAIRMAN EGLINTON: Any other comments on that? I think that's accurate to say that ACOG has published for political reasons certain statements, and in the United States, at least in every hospital I have any familiarity with, and everyone anybody I have ever talked to has any familiarity with, electronic fetal heart rate monitoring is

[--- Unable To Translate Graphic ---]

used and, therefore, is a de facto standard. And it revolves basically around the fact that an RN costs around \$70,000 per shift per year, and you need a minimum of four of them to staff 24 hours in labor and delivery one labor room. And the economy of a fetal heart rate monitor in comparison to that precludes using a nursing staffing ratio in the United States that permits auscultation as a standard. And with current financial constraints being applied by managed care and management consulting firms such as APM and others visiting hospitals and slashing manpower costs out of hospital budgets, this is getting worse, not better. And our staffing ratios, just like every other place APM has been, have declined since they went through our manpower document, and that's been the history in every over hospital they have been in. That's not a secret. That's public information. So like it or not, the fetal heart rate monitors may be here to stay for other reasons that are not strictly medical.

I think that Ms. Young's initial comment could probably be handled just by dropping out that sentence, sentence 2 in the introduction, because we don't really have to say that it's a standard. I hope nothing that I said makes it sound like I disagree with anything she said. I

[--- Unable To Translate Graphic ---]

don't disagree with anything she said. But we could just delete that sentence, and I don't think it would harm the document at all.

Dr. Ross?

DR. ROSS: For discussion, would the panel approve a study comparing scalp oximetry and electronic heart rate monitoring versus auscultation? Or is that what you would be suggesting?

MS. YOUNG: I think that we need to. I think that the financial comments are accurate. I lament them, that we're, in fact, practicing clinically in looking after women in labor with in mind things other than quality-of-care issues. You know, there are medical-legal issues. There are financial concerns. One wonders how many other concerns are going to sort of knock out the quality-of-care issues down the road. So I think we have to be--at least I am always constantly aware of them.

I would like to see the sort of study that you mentioned being done, yes.

CHAIRMAN EGLINTON: Any other comment?

[No response.]

CHAIRMAN EGLINTON: Is there any reason why we should not break early for lunch and reconvene at 1:00, Dr.

[--- Unable To Translate Graphic ---]

Yin?

DR. YIN: Fine.

CHAIRMAN EGLINTON: Okay. We are convened for  
lunch.

[Whereupon, a luncheon recess was taken to  
reconvene at 1:00 p.m., this same day.]

[--- Unable To Translate Graphic ---]

AFTERNOON SESSION

[1:03 p.m.]

CHAIRMAN EGLINTON: Let's go ahead and come to order for the afternoon session. We will go through some of the same thing. It will be a little redundant, but to go through some of the same items we did to start the morning session because we have a different audience here, some new panel participants.

We need to have people sign in. Remember to sign in out front. If you have comments from the audience, please, you must step forward to the podium. I will recognize you, and you can speak then.

When you speak, please identify yourself and give to us a full conflict-of-interest statement, who sponsored you and whom you are representing here today.

Since we do have a new audience, let's go ahead and have the panel members introduce themselves again. Colin is itching on the front of his chair. Are we doing something wrong, Colin? Is this okay? Are we all right? We are not in trouble yet. All right.

Please have the panel members introduce themselves, beginning with Dr. Katz, please, and around this way.

[--- Unable To Translate Graphic ---]

DR. KATZ: I am David Katz from Duke University where I am on the faculty in the Departments of Biomedical Engineering and Obstetrics and Gynecology.

DR. DAVEY: Diane Davey from Lexington, University of Kentucky, and I am director of Cytopathology and co-director of Hematology.

DR. O'LEARY: Timothy O'Leary, Armed Forces Institute of Pathology, Washington, D.C., chairman of the Department of Cellular Pathology.

DR. LEVY: I am Barbara Levy, a practicing gynecologist in Federal Way, Washington, and clinical assistant professor of Obstetrics and Gynecology at the University of Washington School of Medicine.

MS. DOMECUS: Cindy Domecus, senior vice president of Clinical Research, Regulatory Affairs and Quality Assurance at Conceptus, and I am the Industry Rep on the panel.

DR. YIN: Lillian Yin, director, Division of Reproductive, Abdominal, Ear, Nose, and Throat, and Radiological Devices, with Center for Devices and Radiological Health.

MS. YOUNG: I am Diony Young, and I am a consumer member, a new consumer member to the panel. I am from

[--- Unable To Translate Graphic ---]

Geneseo, New York, and I am editor of the Journal of Birth Issues and Perinatal Care.

DR. NEUMANN: I am Michael Neumann from Case Western Reserve University in Cleveland.

DR. SOLOMON: Diane Solomon, National Cancer Institute, Bethesda, Maryland, and I am a pathologist.

CHAIRMAN EGLINTON: Gary Eglinton, director of Maternal Fetal Medicine, Georgetown University.

DR. HARVEY: Elisa Harvey, executive secretary to the Obstetrics and Gynecology Devices Panel.

CHAIRMAN EGLINTON: The FDA press contact will be Dr. Yin for this afternoon. We do have an agenda. We would like to try to stick to it. So, if we have comments from the audience, please be brief and concise.

If you come back to the podium on multiple events, multiple episodes, please re-identify yourself each time because the transcriptionist may not remember who you are, and speak up at that point.

DR. HARVEY: I have already read the conflict-of-interest waivers from this morning and, as well, introduced the temporary voting member status of some of the panel members today. So I will not redo that.

I did just want to make a small correction. For

[--- Unable To Translate Graphic ---]

those people interested in getting a video, the phone number that I gave this morning is incorrect. The correct number for Video Visions is (301) 438-8724, not 26.

That is all.

CHAIRMAN EGLINTON: Mr. Pollard, again, will introduce this afternoon's activities.

MR. POLLARD: Thank you, Dr. Eglinton, members of the panel. I just want to take a few minutes to just go over the agenda for the rest of the day and tomorrow and explain a little bit of the thinking FDA went through to get where we are.

We will be talking about the draft guidance document that you had before you and that the public should have as well on this study of in vivo devices used to detect cervical cancer and its precursors. This document was essentially a response to the development of new optical technology, and it is also the result in part of some preliminary interactions we have had with manufacturers.

We formed a working group within the Center to put this document together, and the idea really being to get something down in black and white for the panel, for FDA, for interested researchers, and for sponsors to look at to essentially deal with it early on where we have a real

[--- Unable To Translate Graphic ---]

chance for genuine impact at a meaningful point.

I would like to mention the difference here when we highlight in vivo. We are really talking about devices that are applied to the patient and pretty much instantaneously gives you that readout. I differentiate that from in vitro diagnostics, what we call IVDs, which are a range of clinical laboratory-type devices reviewed by our Division of Clinical Laboratory Devices, and I should add that there is a member of that division on the working group, and I expect that we are going to learn a lot from that experience that kind of cuts across our office.

I would also highlight that we have put together, I guess what I would call, a designer panel today, made up of members of the OB-GYN Devices Panel, but also with participation from members who are not part of the standing panel, as well as members from the Hematology Pathology Devices Panel, and I would really like to welcome Dr. Davey, Dr. O'Leary, and Dr. Solomon for their help today and tomorrow.

Dr. Neumann, of course, is no strange to our panel, having served already several years. We invited him because of his background in sensor technology and obstetrics and gynecology.

[--- Unable To Translate Graphic ---]

Dr. Katz is a relative newcomer to our panel, but we are very happy to have him with us as well, with his background in engineering and obstetrics and gynecology. We think that is very useful as well.

For the agenda, we have three guest speakers who are going to try to lay a little bit of foundation for the panel to work from for the rest of the meeting. Mark Schiffman from the National Cancer Institute is going to give us a clinical overview of cervical cancer screening. Rebecca Richards-Kortum is going to give us some information about some of the technological aspects of what we are looking at, and Dr. Hirsch from George Washington University is going to be talking about some of the statistical considerations that must be taken when we develop clinical protocols that try to answer the questions that we are interested in.

After that, Dr. Mridu Virmani from our branch is going to walk you through the draft document. Prior to that, we have left a little time for a number of the companies to give the panel their input on the draft guidance document in its current form.

There will probably be a little time at Dr. Eglinton's discretion for questions to the speakers.

[--- Unable To Translate Graphic ---]

However, everyone, including the guest speakers, has agreed to be back tomorrow when the panel will go through the document, page by page, using the discussion questions that our staff prepared.

The one last thing I would like to highlight is the document is a draft. It is a really the first public draft of that, and we have already noted a number of areas, even in preparing for the panel meeting, where we will probably clean certain things up and beef up other areas.

We are very much interested in all of your comments and suggestions and deletions and whatever. I would just say, consider the overall objective that we want good guidance to manufacturers who are developing these technologies for designing the proper kinds of studies that will show safety and effectiveness for their intended use.

After the panel meeting, we will compile and analyze all of the comments and complete the guidance document, hopefully by the end of the year.

Thank you, Dr. Eglinton.

CHAIRMAN EGLINTON: Thank you.

We will move on to the invited presentations, then. Dr. Schiffman?

xx

DR. SCHIFFMAN: Where do I stand?

[--- Unable To Translate Graphic ---]

CHAIRMAN EGLINTON: At the podium, please.

DR. SCHIFFMAN: I am Mark Schiffman of the National Cancer Institute, and I am in charge of the group that is studying the multi-stage carcinogenesis of GYN tumors. The way we do it is through epidemiology, but with molecular markers and a strong pathology component.

So the people in my group are from all three disciplines, epidemiology, molecular biology, and pathology. We have worked since 1984 on cervix, which is by far the best understood in terms of multi-stage carcinogenesis of the GYN tumors.

So my points today will be very focused because I feel like the time is so limited that I should make only points that are directly relevant to the screening issues, but it is an interesting topic when your understanding of a disease is an evolution, and we are talking about what to detect on a carcinogenetic pathway, what are the intermediates that we are looking for, which ones can we ignore. I think there is a lot of very fundamental points raised by this particular panel.

Can we have the lights down somewhat?

I am not going over this to start. I am just saying this is what I hope to explain, without looking at it

[--- Unable To Translate Graphic ---]

to understand it. There are now some major pathways to cervical cancer that are understood and that we should be addressing the screening technology to our advancing understanding of the process. I will go back to that later.

The first point as an epidemiologist to know is that HPV is the main cause of cervical cancer worldwide, as I will mention, and it infects the entire anogenital tract, but it really only causes a major cancer burden in the cervix. There is some interaction between the squamocolumnar junction, and it is HPV.

Now, we have thought of that--I realize that there are not all physicians in this audience, so excuse me. So, really, the--does this pointer work? So where the squamous epithelium of the vagina onto the portio meets the endocervical epithelium is the transformation zone, and I was noticing in the draft document that in some case, women with a hysterectomy will be admitted into the protocols.

I am making the comments as I go along. It might be a mistake in that the risk of vaginal cancer in a woman infective with the virus that causes cervical cancer is very low. Post hysterectomy, the risk of vaginal cancer is an extremely low risk, and if it is for benign disease, it is extraordinarily low. In other words, if the hysterectomy

[--- Unable To Translate Graphic ---]

was for a benign for fibroids, the native risk of vaginal cancer among women infected or not with the causal agent is so low that screening for vaginal cancer is not something that as an epidemiologist is going to be cost effective from a public health point of view at least.

So I am just making the point that it is the transformation zone and its interaction with HPV and that natural history that is fundamental to what we are talking about today.

Thanks to Ralph Richard and Koss and early investigators, we know that there is a continuum of changes that basic cancer does not arise de novo. Instead, there are microscopically evident precursor lesions. Now, we do not have that for every GYN tumor. Ovary, for example, what is the precursor to ovarian cancer? No one is sure. It may be just flat epithelium, but for the cervix, there is a very well-defined--over many years, it is the fifties now--set of precursor lesions that have been named a whole variety of things.

Now, the trouble with any of the nomenclatures is they have become outdated as we advance in our understanding. It became clear early on that carcinoma in situ could not be reliably distinguished from severe

[--- Unable To Translate Graphic ---]

dysplasia, for example, and with the Bethesda system, the low-grade changes were unified, as low-grade squamous intraepithelial lesions.

So there has never been an absolute perfect way to divide all of these precursors between completely normal and definitely invasive cancer. Now, that is a major problem. It is also an opportunity, of course, because it represents understanding. We know a lot of these details, and it may be that with ovary or something as we learn more, we will have an equally messy continuum until we figure it out, but with cervix, as we have learned more, it has created all of these messy borderlines between poorly to visible changes, to the point where I distrust this continuum now. Even the Bethesda system to me is becoming outdated, among people who work on this all the time.

Thank you, Diane. This is Diane Solomon who originated it. She is never going to be outdated, though.

DR. SOLOMON: I will talk to you later, Mark.

DR. SCHIFFMAN: Now, everyone knows a normal cervix when they see it because we are talking now about visual and optical galvanic combinations, and people may realize when something is very bad, but the issue with everything from aided visualization to the most subtle

[--- Unable To Translate Graphic ---]

techniques of any light spectrum still is the idea of continuum because this continuum between normal and cancer exists on the visual level, on the microscopic level, on the molecular level where I work mostly with DNA assays. The same areas of equivocation and uncertainty exists, and I have tried to outline some of this in the little thing in the book. I don't know where the book is, but the article that I submitted. Is that already out? Oh, okay.

So, today, we are talking more on the visual clinical level with in vivo diagnostics, and at any level, there is this continuum and it is always a pyramid, meaning things that are evidently cancer are always rare. Things that are high grade, bad-looking are more common, but still, very rare compared to the low grade and the equivocal low grade. This is cytology, but the same concept exists that there is a wealth of abnormal, but low-grade or uncertain significance of things, and in that sea of abnormality, there are the scary fewer things, and we have to, in terms of cervical cancer diagnostics in the United States, find the bad ones, but ignore as many of the low-grade ones as we can if they are not going to turn bad because these are so common that to pick them all up overwhelms colposcopy services, leads to unnecessary cost, so much so that this

[--- Unable To Translate Graphic ---]

trial which Diane is the project officer and I am the co-project officer, ALTRS, ASCUS-LSIL Triage Study, is funded by the NCI for around \$25 million because people recognized over many years of battling for that degree of funding that this is a major problem in the United States, over-treatment and over-referral and over-treatment of the many, many minor abnormalities that would almost all go away by themselves if left along, but we do not know which ones.

So, in the United States, we cannot afford--or anywhere where we have enough money to try to protect everybody, you cannot afford to ignore things that could be bad soon.

Now, how do you apply that kind of continuum and that kind of a problem to traditional screening? In any kind of assay--this is an old slide now--you have got to choose a cutpoint between the disease and the non-disease. Most of our screening statistic are based on dichotomies, disease, non-disease. Well, here, we know that there is a whole wealth of non-disease and progressively more seriously diseased to really diseased. So it is not continuous. It is sort of ordinal, in a way, and yet, we are trying to find an assay that just cuts that perfectly.

You can think of the Pap smear no matter how

[--- Unable To Translate Graphic ---]

sophisticated, no matter how many millions of neurons are clicking to make that decision on if something is disease or not disease still as pretty much in its statistical treatment as a cutpoint, a single cutpoint, and initially, I feel this is an important historical perspective.

When cytology was organized, it became possible to draw that cutpoint further and further back towards more sensitivity and pick up cancers or suspect cancers.

People started first classifying and then referring dysplasias, then minimal dysplasias, then equivocal, minimal dysplasias, and that was pushing this bar that way to where, all of a sudden, a lot of true normals, people who really are not diseased or never will really be seriously diseased in terms of cancer, which is the disease, are being picked up, and that is the referral problem.

Now, we are trying, through introducing multiple methods now in our studies, to find a combination of cutpoints on different dimensions that work so that you can maintain specificity while increasing accuracy, and I think everything that is coming before this panel, whether it is here or in vitro, is still talking about that kind of discriminate analysis. They are trying to find variables, clinical, microscopic, whatever, that discriminate seriously

[--- Unable To Translate Graphic ---]

diseased or about-to-be diseased, seriously diseased from the benign, and trying to do it very effectively and low cost.

Now, the understanding is aided by realizing that behind all of these changes in the entire pyramid, there is a family of viruses, the papilloma viruses. We have now worked in over 30 countries, and the story is the same everywhere. Eighty-five, or more, percent of cases of cervical cancer everywhere in the world are caused by infection, plus time, with one of these viruses, and it is usually one of the oncogenic types.

This is just a phylogenetic tree based on the genetic diversity of the different viruses, but the point being that there are some clearly cancer-associated ones. Sixteen is the main type everywhere, in health and diseases, also the most common type in most populations among cytologically normal women, 18, 31.

If you add 18, 16, 31, 45, and maybe 33, somewhere in that, you have reached the bulk of cancer cases, the majority in every country in the world, but then you have to add the rest of these to get to a very high percentage.

The condyloma-associated are 6, 11, 42, and a few others, and of course, there are many other types. There

[--- Unable To Translate Graphic ---]

are over 70 types of HPV and over 30 infected cervix, but you can do with about 15 cancer-associated types in explaining in terms of etiologic fraction, virtually all cancers everywhere.

It is a stable virus. It does not mutate. So, in fact, we found a main cause, and we should be able to introduce that knowledge into diagnostics and screening in a fairly definitive way once we recognize that this is responsible, with some cofactors I don't have time to go into, for the entire story, from start to finish, the most minor equivocal lesions, from my friends in the expert cytology panels. There are subtle changes that are often HPV-related, and of course, all the way up through cancers are HPV-related.

I wanted to--because maybe some people do not follow the story--say that from an epidemiologist point of view, all the five major epidemiologic criteria for cause, for saying that HPV causes cervical cancer, have been satisfied, biologic plausibility, specificity of the association, strength of the association, consistency of the association and replication, time sequence, and that was the hardest to do, time sequence, the one in yellow.

We have done long-term perspective studies over 10

[--- Unable To Translate Graphic ---]

years now, showing that HPV precedes and predicts the first onset of cervical neoplasia.

I think because everybody gets tired of talks of any type, I am going to stay to the very most focused points here.

We have studied 26,000 women, prospectively, so far who are normal and looking for the origins of cervical neoplasia, and so I feel like we have as much experience as anyone on what does it mean to have ASCUS, or LSIL, the first things that usually happen when someone is infected, the first evidence cytologically that they are infected.

What those studies have taught us, which these are in Portland Kaiser and Guanacaste, Costa Rica, anyone who wants any methodological details, backup, wants to change anything I said, I think that is maybe better done in conversations, now that I see what the format is going to be, but I have been working on this full time for most of my career.

What it shows us is this. Now I am back to the original drawing now, hopefully able to explain it. Human papilloma viruses are mainly sexually transmitted. There are some exceptions, but they are mainly sexually transmitted, and 90 percent of invasive cancer--these are

[--- Unable To Translate Graphic ---]

etiologic fraction percentages--derive from this pathway of infected women. Uninfected women may very rarely bypass this whole pathway, but those are still debated.

Conservatively, I have just left them out, but let's talk about the ones who have been sort of explained.

If we take infection, infection occurs quite easily through sexual contact. Infection could be multiple or single. Infection leads to some degree of natural immunity following resolution, but the degree is unknown. Latency is not clearly know, yes or no, and how often. It appears to exist, but we are not sure how often. This whole field is only 15 years old. So those are the key questions of viral states like latency that are not completely understood.

Anyway, most of what we learned indicates that HPV infection is a hidden pyramid that is very, very large. It is extremely common among sexually active women. It could be up to 50 percent easily that can be infected if you use PCR-type techniques in a college-aged population if they are sexually active.

In one study in Berkeley, 100 percent of 22-year-olds reporting many partners were infected on a day that we measured a large series of them. So infection

[--- Unable To Translate Graphic ---]

itself is not really important.

Now, what is infection? Infection goes from PCR positivity and very rapidly, over a year or two, like any wart disease, can cause equivocal, cytologic changes, low-grade changes, things called SIL, LSIL, but almost all of this whole complex goes back to nothing, to no disease. It resolves over time.

This was first found with other wart diseases like cutaneous warts or foot plantar warts, but the whole cervical complex of HPV infection, what is called--I call it HPV infection--is a swirling sort of transient and then with re-catching of another type set of things that if we go in and measure, we may find a certain prevalence, but we know very well that that is not the cumulative incidence for any women unless she enters a mutually monogamous relationship.

Many women have a series of infections leading to partial or total immunity to all the different types. Out of that, very common sexually transmitted disease, 1.5 million cases reported a year and many more really unreported, you get for some unknown reason some very small percentage progressed to high grade.

We know that of the low grades in long-term follow-up, something like 10 to 20 percent progress, but

[--- Unable To Translate Graphic ---]

this whole complex is much larger than the tip of the iceberg or the low grades that you see on a given day because that tends to be more serious than the even more milder stuff detectable only on a molecular level, but a day later, this could be this. It moves around quite a bit.

High-grade lesions, you are on firm footing in terms of disease endpoint and precursor once you are at high grade. These rarely regress, though they can regress. They often progress, given enough time, and I consider high-grade HSIL to be the true precursor to invasive cancer, and this to be a viral infection that is as very strong intermediate endpoint and a risk factor for the development of neoplasia. I no longer consider this neoplasia, and many people don't.

The other thing I want to say very quickly is that ASCUS does not exist. It has no morphologic meaning. We have done many studies trying to arrive at a cell that everyone agrees is ASCUS; that is, does not happen. If you get enough experts, no one will call any cell in your atypical repertoire ASCUS. They will either call it LSIL or down to normal, reactive in some way.

We have tried, as Diane knows, to use the book, the criterion book, and train people on it, and that does not improve the situation, and we try different kinds of

[--- Unable To Translate Graphic ---]

markers.

What it appears to be is that ASCUS is either normal, including reactive changes, or SIL, 90 percent LSIL, 10 percent HSIL, meaning there is a subset of ASCUS that really is confused with HSIL and is very highly HPV-positive, indicating that we will not need to triage on that. It will just be lumped with HSIL, but I call ASCUS now equivocal SIL.

With that clarification, we avoid situations like this. This was a study we did with five pathologists. We took 200 slides. These are conventional smears that had been called cytologic atypia. This was before Bethesda. We asked every one of them if it was normal. Zero points. We attributed zero points. Is it equivocal, ASCUS? Half-a-point. Or definitely SIL? One point.

You could see that the HPV DNA prevalence in those that are certain only in the aggregate is almost 100 percent. Whereas, those that everybody called normal are down at--this is the same rate as the normal population, which was 17 percent.

So all we are seeing is that this borderline morphologically may be better expressed by DNA testing than by eyeballing it because it is so difficult. Morphologic

[--- Unable To Translate Graphic ---]

changes are so difficult microscopically that it is just very difficult to reach agreement on it.

So, if you agree that ASCUS is highly equivocal, then it cannot be a gold standard for anything. In the book, in the draft, it talks about some kind of disease standard, including ASCUS. I cannot accept that in my mind because it does not have a gold standard. It is the gold standard of not having a gold standard.

I am almost finished.

The LSIL diagnosis includes, of course, cellular changes of HPV infection, mild dysplasia or CIN1. We have shown, I think convincingly, that this is just as HPV-DNA-positive as HPV infection itself. These are all just very transient or poorly defined characteristics of HPV infection, and there is no way to reliably separate these two.

Some pathologist can distinguish reliably by themselves something they see, but if you bring another expert in, that consensus disappears very quickly.

So I feel, as do--I don't know how many people--that the whole complex of HPV infection, from mild PCR only to CIN 1, histologically confirmed, are the same thing.

[--- Unable To Translate Graphic ---]

The way I sort of have indicated that is we studied 17,654 women in Portland who had never had an abnormal Pap smear in their life. We carefully confirmed that they were normal again today, and we have reviewed all of their past Paps that we could find. This was in Kaiser Permanente. We confirmed they had always been normal and then followed them.

Now, those who were HPV-positive at enrollment had very large percentages of developing an abnormality for the first time ever in their life, every time you followed them, predicted only by the fact that they were positive at enrollment. They looked like everybody else morphologically, and if you do a cumulative incidence rate you find out you cannot do it because that is the point.

The more frequently we looked at women, the more CIN we found. It comes and it goes quickly, and the quicker it comes and goes, the more likely you are to miss it. Subtle CIN could be missed microscopically. It comes and goes quicker than most observation periods, including Pap smears. I am talking about the very bottom of the pyramid of severity.

So you can talk about your sensitivity in detecting the worst low-grade lesions, but you cannot talk

[--- Unable To Translate Graphic ---]

about specificity because there is so much LSIL happening all the time that the only real gold standard of HPV infection is PCR, and no visual technique could ever find all of that.

So I want to review my points, the sort of talking points to be argued about. We should as much as possible forget about LSIL as a target for screening. LSIL was picked as a target of screening by cytologists who are trying to increase their sensitivity by getting closer and closer. They were using the morphologic proxy of the underlying causal infection, but now that we understand that it is the infection itself that they were examining, there are better ways to look for infection if you want to do that.

We have many different ways of looking at the changes of the infection, but to focus on LSIL, to use the microscopic picture, I do not think it is valid. I think HSIL and cancer are the targets for screening, now that we are so much more accurate, and that we should try to have very accurate detection of those higher-grade lesions, and hopefully, if it is accurate enough, we can increase screening interval to pay for the cost of the additional tests.

[--- Unable To Translate Graphic ---]

This is, again--I like this Polartechnics thing. It is just showing that we have to go to many different dimensions, including different--I like to combine tests in our studies to show how good they can be in combination in terms of sensitivity of detection of HSIL in cancer, while still having very good referral characteristics, meaning high specificity, and we actually have a better slide, which I lent down. I should not have. This is earlier, and it has been better, but we can get 90-percent detection in a whole population study in Costa Rica. It is a door-to-door survey of women who have been very poorly screened. We knock on doors, enroll people. So it is not selected in any way. It is a valid group.

By referring only to 8 to 10 percent of the population to colposcopy, we were able to find that good cost of benefit. That is a single screen, just using several tests at once, and these are all modalities, different tests that we tried or test combinations, thin preps plus cervicography.

This is--I don't even know--HPV DNA at the picograms level, plus so and so. When you get--I am just making that up. I don't know which is which anymore, but the point is, you can get a lot of them which are winners,

[--- Unable To Translate Graphic ---]

not just a matter of political will, I think, and financial resources, to basically take a high-risk population and turn it into a low-risk population real fast.

As we follow these women out, we find nobody else is getting new, really bad stuff. So we have cleaned the population for cancers.

So, before you are content with any one combination, think about projecting combinations. This is all ignoring LSIL, which we no longer see as a screening target.

Thank you very much.

CHAIRMAN EGLINTON: Thank you. We will have time for questions later.

Next, we have Dr. Richards-Kortum.

DR. RICHARDS-KORTUM: Thank you.

This afternoon, I would like to share my perspective on emerging optical technologies for detection of cervical cancer and its precursors.

I have been involved in academic research in this field for the past 12 years, and over the last 7 years, my group has collaborated to develop optical methods for detecting cervical precancer.

I will begin by overviewing the biophysical

[--- Unable To Translate Graphic ---]

principals which underlie this technology. In my opinion, it is really crucial to understand these principles in order to maximize the performance of these techniques, and more importantly, to understand the clinical situations in which they will fail and the factors which need to be controlled in multi-center clinical trials.

I will conclude by discussing what I feel are important considerations in evaluating the efficacy and the risk associated with these new technologies.

Although there is a lot of excitement about the potential of new technologies, optical methods have already made really important contributions to reducing the incidence and mortality associated with cervical cancer.

Physicians have been able to directly visualize the cervix since the invention of the speculum in the early 1800's. This led to a series of new diagnostic and screening modalities, all of which are based on optics. These include colposcopy which, with the use of the green filter, really represents one of the first applications of speculoscopy for in situ diagnosis, and also cytology-relied absorbing dyes are used to indicate changes associated with neoplasia.

These optical methods have provided a unique

[--- Unable To Translate Graphic ---]

window to enable us to study the progression of cervical cancer and its precursors.

While our understanding of the biology of this process has evolved dramatically, really, the optical methods haven't changed much since the fifties.

Innovations in photonic technologies in the last decade, though, I think, have the potential for us to make major advances in screening and diagnostic techniques, and these innovations fall into four categories.

The first is improvements in technologies, developments in lasers and LEDs, fiberoptics, and CCD detectors, which enable us to record optical signatures with a very high precision, enabling us to record changes that our eyes are not sensitive to.

As a result, researchers have begun to examine the use of tissue speculoscopy. In speculoscopy, we record the intensity of light returning from the tissue as a function of color or wavelength, and this can give us information that is characteristic about both the molecular and the cellular composition of tissue.

Using fiberoptics and other types of probes, we can control the delivery and have quantitative detection, and finally, through an improved understanding of the light

[--- Unable To Translate Graphic ---]

tissue interaction, we can relate these optical changes to changes in the biochemical composition and the morphologic composition of tissue.

Given these new ways to examine the interaction of light with tissue, we have a unique opportunity to design optical methods, which can give us information directly about molecular composition, morphology, and tissue architecture.

If successful, the potential advantages are numerous. Because we can make these optical measurements in real time, we have the possibility to decrease cost by reducing office visits and decrease the loss to follow-up.

Because this optical radiation penetrates the full thickness of the epithelium, we can look at that full thickness tissue without biopsy and potentially increase both sensitivity and specificity, and because we can develop software algorithms to analyze this data, either automating or semi-automating the analysis, we can reduce the need for operator training.

However, in order to achieve and understand these potential advantages, we have to consider the biophysical principles behind the technologies. These essentially fall into two categories, what we can control in terms of our

[--- Unable To Translate Graphic ---]

instrumentation and software and what the tissue controls through its interaction with light.

Through appropriate design of hardware, we can try and maximize the contrast between normal and neoplastic cervix, and essentially, we have control over three parameters, the source of light that we use to illuminate the cervix, the conduit that we use to deliver light to the cervix and collect the light remitted from the cervix, and the detector that we use to sense this light.

The range of optical parameters that we can look at are numerous. They include color or wavelength, intensity, and spacial patterns which we can detect in the form of images or through precisely designed fiberoptic probes.

We can also design software algorithms to analyze these data, taking either an empirical approach or a model-based approach, to yield results which can be correlated to the features of disease.

We also have to consider the interactions which occur between the tissue and the light at both the molecular and the morphologic levels. Basically, there are three types of interactions which can occur, scattering or a direction change in the light, absorption or a reduction in

[--- Unable To Translate Graphic ---]

intensity of the light, and emission or a conversion by the tissue to another color of light.

Now, everyone is familiar with the first two of these interactions. The sky appears blue because the atmosphere preferentially scatters blue light. Grass appears green because chlorophyll absorbs all of the other colors and reflects back green light, but how do these principals and processes affect the light that we measure from tissue?

This cartoon illustrates the trajectory of photons, or particles of light, within the tissue. When light is incident on the cervix, it can be scattered about by moleculars and cells within the issue so that its direction has changed. When it is incident on something that is highly absorbing like hemoglobin within a blood vessel, the light is preferentially absorbed at those wavelengths and we don't see it coming back.

In order for us to see the light being remitted from the surface of the cervix, it has to get turned around by scattering that occurs in the tissue.

Another interaction that we can see is when a photon is absorbed by the tissue, the tissue can remit that energy in the form of an inelastic scattering process, where

[--- Unable To Translate Graphic ---]

the color of light has changed, and that emitted light can scatter about through the cervix and can be remitted from the surface of the cervix so that we can see it.

Recently, many techniques have been described in the literature which utilize these effects for in situ detection of cervical precancer, and this chart summarizes the techniques which have been described in the literature beginning with colposcopy which essentially relies on the diffuse reflectance of visible light to identify lesions for biopsies, and all of the other techniques that have been described can essentially be viewed as variance of colposcopy because they rely on the same interaction of light with the cervix.

In cervicography, for example, a camera is used for a later review by an expert. In digital colposcopy, a CCD camera is used to capture the image, and software algorithms are used to identify lesions.

In speculoscopy, the light source is replaced with a blue light, chemiluminescent light source with peaks at 430, 540, and 580 nanometers.

All of these techniques that are shown in yellow rely on fundamentally the same principle, and that is diffuse reflectance of broad-band light from the cervix.

[--- Unable To Translate Graphic ---]

An elastic back-scattering speculoscopy, the Polar probe is one example of such technique, the conduit of light has changed, and light is delivered and reflected light is collected with fiberoptic probes to have very precise spacial geometries.

In elastic back-scattering speculoscopy, fluorescent speculoscopy and Raman speculoscopy are two examples. Filters are used to block the detector from seeing the color of light that the cervix is illuminated with, and the detectors now see the light that is produced by the cervix which can be orders of magnitude weaker than the light that is reflected at the illuminating wavelength.

In order to understand the relative merits of these new technologies, it is important to understand how the signals are produced in the tissue and how contrast between normal and neoplastic areas are achieved, and it is instructive to first consider the familiar techniques based on diffuse reflectance, including colposcopy, cervicography, and speculoscopy, where again a signal is produced through a combination of two effects, scattering and absorption.

Scattering, a direction change of the light, is characterized by the scattering coefficient which gives the probability that a scattering event will occur in a given

[--- Unable To Translate Graphic ---]

path length, and the phase function tells how likely it is to scatter light from one direction into another direction.

Scattering in cells is produced by spacial fluctuations in the index over a fraction, and most cells are very highly forward-scattering, but as the direction of light is not changed very much in any one scattering interaction.

In order for us to see light coming back from the cervix, again, it has to get turned around by these scattering events, and that can occur through one very large angle scattering or multiple small angle scattering events.

Absorption acts to reduce the intensity of light coming back from the cervix, and the absorption coefficient, which characterizes the probability of absorption, has a strong wavelength dependence or color dependence.

Hemoglobin is one of the most important absorbers that is present in the cervix, and it has absorption peaks in the blue, the green, and the orange regions of the spectrum.

So scattering and absorption account for the images that we see through the colposcope, but how do they interact to produce the hallmark findings of an abnormal colposcopy that has abnormal vascular patterns and in situ

[--- Unable To Translate Graphic ---]

whitening?

Well, first, consider the normal cervix which has a diffuse pink appearance when illuminated with white light. The scattering which occurs in the epithelium essentially randomizes the light. So we do not see the individual vessels which are found in the stroma beneath.

The hemoglobin preferentially absorbs blue light and reflects back other colors, and that is why the cervix has this diffuse pink appearance. When vessels form in the epithelium, now there is less scattering material overlying them. So we are able to see the individual vessels.

When we use the green filter in the colposcope, the hemoglobin preferentially absorbs that green light so the vessels appear dark, and we enhance the contrast between the vessels and the surrounding tissue.

What produces aceto-whitening? The image on the left here was obtained with a confocal microscope and shows images of epithelial cells which have not been stained. Areas of high signal are proportional to areas where the index of refraction is fluctuating spatially, and you can see that we can make out the nucleus and we can make out the cytoplasmic membrane.

When we apply acetic acid to these same cells and

[--- Unable To Translate Graphic ---]

image them through the confocal microscope, the image on the right results, and then the back-scattering signal from the nucleus is dramatically increased by the application of acetic acid.

This strong light scattering interrupts the transmission of light to the stroma, and so the hemoglobin absorption never has a chance to make the light appear pink.

Neoplastic areas appear whiter because the nuclei are larger and there are more back-scattering centers that are induced by the application of acetic acid.

Well, let's consider what happens now when we use these same interactions, scattering and absorption, but now we have become more quantitative in the instrumentation, first, by carefully controlling the geometry that we illuminate and detect with, and second, by recording the signal at many different illumination wavelengths in a quantitative way.

This approach is termed "elastic back-scattering speculoscropy" and is illustrated in the cartoon here. The detected light comes through the illumination fiber, and in order for us to sense it with the detection fiber, it has to tunnel through some of the tissue and undergo scattering and possibly absorption events. We do not see the light that

[--- Unable To Translate Graphic ---]

has been absorbed. We see only what scatters from this fiber over to our detection fiber.

The nice thing about this geometry is that it is very, very sensitive to the scattering phase function of the cells, and in particular, as the fiber separation and the numerical aperture of the fibers are reduced, we become very sensitive to that scattering phase function, and we detect those high-angle back-scattering events preferentially. So the effect of acetic acid becomes very important, sine it tends to increase the back scattering.

We can model this expected signal using Monte Carlo techniques, which statistically track the progression of photons through the tissue if we know the absorption and scattering properties of the tissue.

Understanding the scattering properties, in particular, the scattering phase function has been difficult to do theoretically, but recently, electromagnetic models have been introduced which relate the scattering properties to the three-dimensional ultrastructure of the cell.

Several group have proposed this methodology for precancer detection in both the cervix and the bladder.

Judy Moran, at Los Alamos National Labs, has shown that elastic scattering speculoscopy can provide useful

[--- Unable To Translate Graphic ---]

information for detection of bladder cancer. The graphs on the left illustrate the intensity of the back scattering as a function of wavelength or color, where the signal goes from 300 nanometers in the ultraviolet out to 800 nanometers in the infrared region of the spectrum.

The top graph shows spectra from malignant areas of tissue, the bottom graph from normal areas of tissue. In the slope of the spectrum, from about 330 to 370 nanometers of tissue is very different in malignant and normal samples. It decrease in malignant samples and increases in normal samples, and very accurate algorithms have been prescribed by this group for separating normal and malignant tissues with high sensitivities and specificities.

Once the wavelengths of interests have been identified, simpler probes and algorithms can be designed to take advantage of them. The Polarprobe is the one example which has been proposed for detection of cervical neoplasia, and one description of this probe, tissue is eliminated with light from 4 LEDs, in the green, the red, and the infrared regions of the spectrum.

One algorithm that has been presented in the literature takes the ratio of light back-scattered as 660 nanometers in the red to that in the infrared, to

[--- Unable To Translate Graphic ---]

discriminate normal tissues from atypia and higher pathologies, with an accuracy that ranges from 85 to 99 percent.

Understanding the precise morphologic basis of such signals at the cellular level is a subject of a lot of ongoing research. In particular, we are investigating electromagnetic models which can predict how the scattering phase function or the intensity of scattering as a function of angle depends on the precise three-dimensional structure of the cell.

In particular, our preliminary results show that fluctuations in the chromatin density in the nucleus increased the back scattering, and geometries where the fibers are very close together are sensitive to this back scattering, but further research is needed to fully understand these mechanisms.

Considering next what happens when we alter instrumentation to take advantage of inelastic interactions which take place in the tissue. If we now place a filter in front of our detector so that it is blind to the light that is being reflected back at the illuminating wavelength, but instead, is sensitive to the light that is produced by the tissue at other wavelengths, we gain an important source of

[--- Unable To Translate Graphic ---]

contrast. We are able to see molecules in the tissue which produce light inelastically, and fluorescence is one such type of inelastic interaction.

Examining tissue fluorescence really gives us two additional forms of contrast. Now our signals depend on the color of light that we illuminate with, as well as the color of light that we detect at, and we are sensitive to molecules in the tissue which produce fluorescence, and these are sensitive to the metabolic status of the tissue. They include the cofactors NADH and FAD, which are related to the redox potential, the aromatic amino acids, tryptophan and turacin, as well as molecules associated with the structural proteins, collagen and elastin in inflammatory cells.

The downside of this technique is that it is weak. Typically, fluorescence is three to four orders of magnitude weaker than the excitation light. So you cannot see it by your eye.

There are a number of ways to measure fluorescent spectra from tissue. The simplest way is to measure from a single pixel of tissue where you use one fiberoptic to illuminate the tissue and one fiberoptic to collect the resulting fluorescence.

[--- Unable To Translate Graphic ---]

You can provide spacial images of fluorescence, just by scaling this up in parallel and having many adjacent fiberoptic pairs, or you can essentially modify a colposcope to record fluorescent images of tissue by equipping it with the appropriate filters.

Our group is initially concentrated on using the single-pixel approach, and this is a photograph of the probe that we have used to do that. Laser light is delivered through three excitation fibers at three different excitation wavelengths and collect the resulting fluorescence.

This slide illustrates typical fluorescence spectra from cervical tissue. Here, I have plotted the fluorescence intensity as a function of emission wavelength. The excitation wavelength was in the UV at 340 nanometers, and the emission wavelength runs from the UV all the way out to the red region of the spectrum, and this shows data from two different patients.

The normal cervix has the highest fluorescence intensity, and as we go from inflammation to HPV to CIN2, we see the intensity of fluorescence drop and the peak emission wavelength shift to longer wavelengths or toward the red region of the spectrum.

[--- Unable To Translate Graphic ---]

Our studies indicate that this fluorescence is due to a combination of fluorescence produced by collagen and NADH at this excitation wavelength, but we see the reabsorption signature of hemoglobin superimposed on top of this fluorescent signal.

In studies where we turn our patients to colposcopy, we consistently observed a very significant patient-to-patient variation in the overall intensity of fluorescence and in the fluorescence line shape.

Furthermore, the fluorescence of columnar-normal tissue is very different than that of squamous-normal tissue. Despite that, there are still important differences between the various categories of tissue that we would like to discriminate. In particular, there is a decrease in intensity as we go from normal to inflammation, all the way through cancer, and an increase in the red shift.

In measurements from 361 sites and 92 patients at these three excitation wavelengths, we randomly divided our data into a training and validation set and developed a multi-variate statistical algorithm to separate tissue into the categories of normal, low-grade, and high-grade cell.

If we compare the diagnosis based on our multi-variate algorithm in the validation set to that from

[--- Unable To Translate Graphic ---]

colposcopy and histology, we find that the agreement with colpo and histology varies from about 70 percent to 85 percent.

There are other methods which you can use to collect this fluorescence information. A multi-pixel probe has been designed by our group where each pixel here now represents an individual spectrum that we are obtaining from a precise spatial location on the cervix, so we can scale up this approach in parallel.

Other groups have designed systems to directly image the fluorescence. This is an example of an image with the LIFE system to image bronchial tissue, indicating a region with CIS, but the same approach could be used for the cervix.

All of these technologies show promise in the preliminary trials that have been reported in the literature to improve diagnosis and potentially screening for cervical precancers. The question remains, though, how do they compare to the standards of care, and given this, what is the appropriate clinical role for these new technologies.

Recent review articles have considered the performance of the Pap smear and colposcopy in the referral setting. This slide summarizes the performances of these

[--- Unable To Translate Graphic ---]

techniques, plotting the sensitivity versus 100 minus the specificity in 30 studies where the Pap smear was compared to biopsy and 12 studies where colpo is compared to biopsy.

The diagonal line represents the agreement that would be expected by chance, and perfect agreement with biopsy would be represented by the upper left-hand corner of the graph. Clearly, there is a tremendous variation from one study to another, and part of this variation can be explained by the tradeoff that occurs between the false positive rate and the false negative rate.

This tradeoff actually provides a complete characterization of the performance of a technology. It is referred to as an ROC curve. Littenberg has recently described a method to estimate the ROC curve of a technology from a metaanalysis of sensitivity and specificity values reported in the literature.

Here, we show the estimated ROC curve of the Pap smear in green and colposcopy in yellow, again, in the referral setting, and these curves represent the performance metrics to which emerging optical technology should be compared.

This slide shows estimates of the ROC curve of fluorescent speculoscopy with a multivariate algorithm shown

[--- Unable To Translate Graphic ---]

by the white line and a neural net algorithm shown by the white dots here, relative to the Pap smear in green and colposcopy in yellow.

They both indicate that there is potential to enhance the performance of the Pap smear and colposcopy with the potential to reduce the need for operator training. Estimations like this are required from larger multi-center clinical trials in the hands of operators with varying skill levels to establish the appropriate clinical roles of these new technologies.

In conducting such clinical trials, the biophysical bases of these interactions dictate a number of important factors to be controlled to achieve reproducible results that can be compared between centers and investigators.

First, care must be taken to appropriately calibrate the optical devices. In particular, these detectors, their sensitivity, can have a strong wavelength dependence, and this must be calibrated using NIST-traceable standards.

In addition, the interaction of light with tissue causes the recorded signals to be very sensitive to the precise excitation and collection geometries. The

[--- Unable To Translate Graphic ---]

illumination details really matter. For example, we conducted a study using a single-pixel probe and a multi-pixel probe. The detectors have been calibrated using NIST-traceable standards, and an individual pixel in either device had exactly the same geometry. The only difference was that in the multi-pixel device, all pixels were illuminated simultaneously.

When we compared the resulting spectra, here is intensity versus wavelength at three different excitation wavelengths. Multi-pixel spectra are shown in white. Single-pixel spectra are shown in green. Clearly, there is a dramatic difference between these data, and this can be explained because light that is produced in one pixel can tunnel over to a neighboring pixel. In this tunneling process, it undergoes a longer path, and it is more likely that some of that light can be reabsorbed. So we see the signature of hemoglobin reabsorption more strongly in that geometry than we do in the other geometry.

The presence of any external agents which can affect the optical properties of tissue also must be carefully controlled; for example, acetic acid strength because it so strongly affects the back scattering, as well as time following application. The pressure that a contact

[--- Unable To Translate Graphic ---]

probe places on the tissue can also distort the cells and affect the scattering properties, and any drugs which affect the morphology, the absorption, or the fluorescence of the tissue also can have an effect on these signals.

Finally, biologic and demographic variables which affect the morphology or the tissue architecture, and particularly the epithelial thickness, like age, the stage in a menstrual cycle, whether the woman is pre-, peri-, or post-menopausal, can impact the resulting optical signatures.

Finally, the UV illumination associated with some of these technologies is a potential safety concern. The colposcope has been used extensively in the United States since the 1950s with no adverse effects reported from the UV illumination. Although the illumination from the colposcope is primarily in the visible region of the spectrum, there is some light in the UVA region of the spectrum between 320 and 400 nanometers. So we conducted a study to compare the relative risk of illumination with the colposcope and our fluorescence spectroscopy system.

Furthermore, ANSI and ACGIH provide standards for absolute levels of illumination of the skin in the UVA region, and we have evaluated colposcopy using these

[--- Unable To Translate Graphic ---]

standards. But the question remains: Are they appropriate standards for illumination of the cervix?

The biological effectiveness of light is highly wavelength-dependent and can be characterized by its action spectrum. This plot shows several different types of action spectrum. The potential for damage is plotted on the y-axis, and note this is a logarithmic scale, versus wavelength on the x-axis. And here I show three action spectra: one for cytotoxicity, one for protein DNA crosslink formation, and another for skin carcinogenesis.

The relative damage potential decreases dramatically as we go from the UV to the visible region of the spectrum, as much as 5 orders of magnitude.

We measured the relative spectral output in joules per square centimeter per nanometer of a colposcope and our spectroscopy system versus wavelength, and here I show from 320 to 500 nanometers. The average colposcope is shown in red here; the highest power colposcope is shown in green; and the lowest power colposcope is not really even visible on this graph.

The spectroscopy system is shown in yellow here, and at the wavelengths where we're exciting fluorescence, there's a lot more light coming from that system. But we

[--- Unable To Translate Graphic ---]

have to take into account the biologic potential as a function of wavelength and weigh this radiant exposure by that spectral effectiveness. And if we multiply the spectral radiant exposure by an action spectrum, we get a better measure of the potential for damage. The area under that curve really gives you the relative risk.

We compared the relative risk of illumination by an average power colposcope, a low and a high power colposcope to that of our fluorescence system using three different action spectra. We arbitrarily assigned the average power colposcope a relative risk of 1. The low power colposcope is about a factor of 3 lower; a high power colposcope about a factor of 2 higher; and fluorescence systems are comparable to or lower than the average power colposcope.

Now, the question remains: Are these action spectra appropriate for cervical epithelium? And in particular, we haven't taken into account the potential for HSV, HPV, and HIV activation.

This previous work has examined relative risk. ACGIH provides absolute standards for broad-band illumination in the UV region from 320 to 400 nanometers, and this standard says that the spectral effectiveness

[--- Unable To Translate Graphic ---]

should be calculated relative to that at 270 nanometers, using the relative spectral effectiveness curve as a function of wavelength that's shown here.

The effective UV radiant exposure at 270 nanometers, according to that standard, should not exceed 3 mJ/cm<sup>2</sup>, and if we use that curve to calculate the effective UV radiant exposure from a colposcope at 270 nanometers, it's well below that standard by two to three orders of magnitude.

In conclusion, optical technologies can provide instantaneous, automated, and accurate diagnoses which can be related to changes in morphology and chemistry. New research is deepening our understanding of the relationships between these signals and the tissue composition.

We have to be careful to control the illumination and collection geometries as well as any factors which can influence the tissue optical properties. Data with good signal-to-noise ratios can be achieved at illumination levels that have similar relative risks compared to colposcopy.

And I'd just like to acknowledge the contributions of my collaborators--Dr. Michele Mitchell, and Sharon Thomsen at the M.D. Anderson Cancer Center, Tom Wright, Dave

[--- Unable To Translate Graphic ---]

Sandison, and a host of graduate students and post-docs in my lab--as well as our sources of funding--the Whitaker Foundation, the NSF, and LifeSpex.

Thank you.

CHAIRMAN EGLINTON: Thank you.

Now, Dr. Hirsch?

DR. HIRSCH: Okay. Now the easy stuff.

I'm professor of epidemiology and biostatistics at George Washington University School of Public Health and also professor of statistics at the graduate school at GWU, also, incidentally, I am also a minister, which means that I can talk for a real long time about almost anything.

Actually, Dr. Harvey, if I stick to the time limit, could you give me a note to take back to my students? I'd appreciate it.

My particular area of research interest is in methodologies in epidemiologic research, recently concentrating on things that have to do with diagnostic devices. And this area of research was stimulated substantially by a recent sabbatical I took here at the Center for Devices and Radiological Health in the Division of Clinical Laboratory Devices. And the reason that I was interested in doing that as a sabbatical is that, as a

[--- Unable To Translate Graphic ---]

professor and teaching also not only at the university but also at NIH, with people who have to face reality, I was impressed with the kinds of things that we normally teach not bridging the gap between what we provide and what people need in the real world.

So what I wanted to do--and there were a number of people who took courses from me here at CDRH that really stimulated my interest in reality, and I came here and this is reality, folks. It really is. To be here working at the Center for Devices and Radiological Health and with the public health importance of the things that are worked on and also with the financial business aspect of things that are worked on, it's definitely reality.

As a result, what I tried to do is I try not to do what some other statisticians might do, and that's to tell you what you can think about, what you can do research on, the way that you can draw conclusions and analyze data. Because I'm firmly convinced that it has very little to do with reality. It's not appropriate, I don't think, for methodologists to tell us that we can't do what we want to do in reality.

So what I'm going to try to do is I'm going to try to tell you about some design issues and some analysis

[--- Unable To Translate Graphic ---]

issues that hopefully are reflecting the reality of the kinds of research that this guidance document is addressing. And the harder of those two is to talk about the design issues. Statistics is much easier than design, and the reason that it is is because statistics, we can talk about that in mathematical language. It might not seem an advantage to you, but it's certainly an advantage to those who are comfortable in talking about mathematical language, allows us to talk very precisely. We can't do that, or at least we can only do that to a very limited degree when we're talking about issues of study design. So that really is the harder of the two subjects. The hardest course I teach is my advanced epidemiologic design class.

To start out, I think that something that impressed me, as I was working in DCLD, is that there are two general approaches to looking at a diagnostic device. One of those approaches is to compare the new device to a reference procedure, and by reference procedure, if I'm going to be real, I have to admit that it's not a gold standard. But it's perhaps the best that we can find.

In the guidance document, I think that the Intended Uses 2 and 3, ASCUS triage--which now might not be interesting anymore since ASCUS doesn't exist--and, No. 3,

[--- Unable To Translate Graphic ---]

the localized biopsy sites seem to me that these called for the kind of study in which we're comparing the new device to a reference procedure.

The other kind of design is to compare a new device to an existing device, and this might be something that you would expect to not find in a PMA, but in the guidance document there is definitely interest in comparing these new devices to existing devices.

The first intended use is an adjunct to Pap smear. The existing device there is Pap smear alone. So that it's a head-to-head competition between those two, with the reference procedure somehow acting as the referee.

And Intended Use 4, as the primary screening device, the existing screening device being Pap smear, so that, again, it's a head-to-head competition between Pap smear and the new device.

If we're interested in comparing the new device to reference procedures, the first problem that we run into is that there is imprecision in the reference procedure. With biopsy especially important as a reference procedure in cervical cytology and diagnosis of cervical carcinoma, we need to recognize, of course, that biopsy misses a lot of true cases. The sensitivity, therefore, is less than

[--- Unable To Translate Graphic ---]

perfect. It's less than 100 percent.

Now, something that we could argue is whether or not the next statement is true, and I have waffled back and forth myself, but I have been supported by my clinical colleagues telling me that this is probably pretty close to true: that biopsy doesn't find false positives, at least if we are talking about really high-grade lesions. False positives for low-grade things, we don't know what that means. But--I don't know what that means.

So we perhaps can consider the specificity to be 100 percent, and that's a pretty good reference procedure that at least is half gold.

What I'd like to do is have very few numeric examples of what's going on when we analyze data from different kinds of study designs. And what I have done here is I have imagined that we are doing a study of a new device that actually has sensitivity and specificity both equal to 90 percent, so a pretty respectable diagnostic device. We don't know that when we analyze the data, but we need to know that to see what happens.

The reference device I'm assuming is a little bit worse than the new device, and I think that's a realistic way to think about things, that we're interested in

[--- Unable To Translate Graphic ---]

designing diagnostic devices that aren't less than existing procedures but, rather, better than. So the reference procedure here is assumed to have a sensitivity and specificity both equal to 80 percent. So new device, 90 percent; reference procedure, 80 percent.

This is the kind of data we would expect to observe, at least on the average, if we compared the performance of the device to the performance of the reference device. There are always assumptions, and I guess an important assumption that I should confess at this time is that in order to see this, what we're assuming is we're assuming, as statisticians say, statistical independence of these two procedures. What that means in more everyday language is that there is not a correlation, there is not an association to mistakes made. Mistakes are made by the new device. Mistakes are made by the reference device. But they're not necessarily the same mistakes.

That assumption makes it possible for me to calculate these things. That assumption is probably not very realistic because devices share technology, they share part of the pathology that they're sensitive to, and when that happens, we probably have correlation of errors.

What happens when you violate that assumption is

[--- Unable To Translate Graphic ---]

things just aren't quite as dramatic as they seem when they are statistically independent. But, anyway, in this study, we're assuming that we have 50 people who are positive on the reference procedure, 50 people who are negative on the reference procedure, and in each one of these cells of this two-by-two table, you can see that the number of observations include some people who have the disease and some people who don't have the disease. So each one of these cells is contaminated by something that we wish wasn't there.

Next, Max?

This is what we get from that kind of study. Remember, the new device has a sensitivity and specificity equal to 90 percent. Its apparent sensitivity and specificity from the study is 72 percent. So a substantial underestimate of the diagnostic performance of the new device.

Now, that's pretty depressing, and so people have come up with ways to fix that. They've come up with solutions to this problem of underestimation with an imperfect reference procedure. And the two most commonly encountered solutions are the resolution of discrepant results and retesting positive results.

[--- Unable To Translate Graphic ---]

Resolution of discrepant results is when we take the individuals for which the reference procedure and the new procedure disagree. One of them's positive, one of them's negative. Those are discrepant results. Then somehow we resolve that. We can resolve that perhaps using the reference procedure again, or maybe we can resolve it using something that's actually better than the reference procedure.

This is a common approach when the method that we're using to resolve discrepant results is expensive, and we don't want to do it on everybody. So we want to look more closely at those individuals for whom we're confused about their diagnostic classification.

Retesting positive results is something that is attractive when it's--not so much cost, but it's the ethical aspects of applying the resolution procedure to individuals who probably don't have the disease. And cervical carcinoma is certainly a good example of that in which you are nervous about biopsying people who have no pathology to suggest that they should have a biopsy.

Take both of these results and overestimation of the device's performance. Here's resolution of discrepant results. Now I'm looking at how the new device performs not

[--- Unable To Translate Graphic ---]

relative to the reference procedure, but relative to the reference procedure after the discrepant results have been resolved by--and I'm assuming here a perfect resolution. Perhaps clinical course would be a perfect way to resolve these.

The two discrepant cells, the ones that are on the upper right and lower left parts of the two-by-two table, those are the discrepant cells, and by applying this perfect procedure to those individuals, we find that we're able to move some of them to the cells in which there is not a disagreement.

Now, we can only move things back and forth between presumed disease positive and presumed disease negative. It's not fair to do the same with a new device because our purpose is to find out how well the new device performs. So all we're doing is we're getting a more accurate idea of who has the disease and who doesn't have the disease.

Let's see the results of that.

Remember, before, the apparent value of the sensitivity and specificity was 72 percent. Now it's certainly closer to 90 percent, but it is an overestimate. How much of an overestimate depends on a number of things.

[--- Unable To Translate Graphic ---]

In this particular case, much less of an overestimate than we had previously as an underestimate.

This is kind of a surprising result, I think, especially to clinicians who know that if you are confused about the diagnostic classification of an individual patient, if you have discrepant laboratory results, pathology results, then it's a good idea to get those resolved. And it is a good idea.

As far as individual patients are concerned, you can get a more precise, more of them correctly diagnosed by resolving discrepant results. It's efficient and effective.

Unfortunately, when we're looking at the performance, when we're comparing performance of a device, it's not the accuracy of each individual that's important. It's also important how those are distributed among the groups. And it turns out that when we're comparing diagnostic performance, the distribution of those who are correctly resolved and those who aren't is such to cause an overestimate.

Let's take a look at the other possibility, and this possibility, this is what we might do if it's very expensive to use the reference device, and now we're going to retest everybody who's positive. The only difference

[--- Unable To Translate Graphic ---]

here between this and resolution of discrepancy in this particular example is our ability in the group that have positive for both the reference and the new procedure to get rid of any contamination of individuals who don't really have the disease. And that gives us a closer value to the actual value, still an overestimate.

Now, one thing I don't want to do is I don't want to be the kind of statistician, epidemiologist, who says, so, you can't do any of these things, because that's just not a possibility. The reality is that you need to work with imprecise reference values, referent tests. And these examples give you some idea of the kinds of things that might influence your interpretation. It's not to say that these aren't things that you should consider as part of your study design. Just confess that if you do a comparison with an imprecise referent procedure, you're going to underestimate; if you resolve discrepant or retest positives, you're going to overestimate the performance of the device.

Another thing as I read the literature for diagnostic tests and diagnostic devices, I found that I was looking at two kinds of study designs. And I don't think that--for me, anyway, this was a relatively recent

[--- Unable To Translate Graphic ---]

revelation, and I'm not sure whether all my other colleagues appreciated it and just never told me about it or this might be a surprise to them as well. When we are interested in comparing two devices, when we are interested in comparing a new device to an existing diagnostic device, there are two approaches to take. One I call the case control approach. In the case control approach, what you do is you look at each of the device's performance relative to a reference device, and then you compare the results of those two analyses.

This is the most commonly encountered design in the medical literature as far as my rather informal but voluminous, perhaps, look at the diagnostic literature.

The advantage of this approach is that you don't use--you don't have to use the same people to look at the characteristic of the old device and the new device. And, therefore, you can do the kinds of analyses that Dr. Richards-Kortum was talking about a few moments ago in which she was talking about meta-analysis, examination of diagnostic performance. This is the kind of information that would allow you to do that.

The other design is what I called the paired approach, and this, I think, if I recall correctly, is a

[--- Unable To Translate Graphic ---]

more common approach to be seen, at least in the Division of Clinical Laboratory Devices for comparing a new and existing reference procedure. But that's a much more informal and less voluminous look at that information.

In this approach, what we do is we use the reference procedure not to look at the performance of each of the devices alone but, rather, using the reference procedure to separate people into two groups--a group that are presumed disease positive, reference positive, and a group who are disease negative, reference negative. Then in each of those groups what we do is we compare the performance of the new and existing tests, devices, directly.

There are a couple of advantages to this approach. One advantage is that there's not the same mixing of sensitivity and specificity that we have in the case control approach. Each one of those two-by-two tables reflected not only the sensitivity but also the specificity of the device here. The top two-by-two table reflects the relative, the comparative sensitivities of the device and the lower table reflects the relative or comparable specificities of the device.

One reason this approach is really, I think, the

[--- Unable To Translate Graphic ---]

better approach when one is really doing research to compare two devices, as in two of the intended uses, Intended Uses 2 and 3 in the guidance document, because it allows us direct comparison. Another thing that it does is it allows us to have a paired study. This is a paired study. The previous one was not. It would be wrong to analyze the case control data that I showed you before as if it were paired data. It's not. But here it is paired data. Here, each one of these letters that stand for a certain frequency in a cell of the two-by-two table tells us about how two tests on the same person--what the results of two tests on the same person look like, so that we're using the reference procedure not just as a statement of disease to compare the performance of a particular test, but to segregate the data.

Another advantage of this approach is that if you have a diagnostic--if you have a reference procedure that's not pure gold but, say, half gold, like biopsy, if you have a reference procedure in which you think the specificity may be 100 percent, may be perfect--if you don't see it, then it's not there. The neat thing about this is that the top table, the table in which we're assuming that everyone has the disease, if biopsies are reference procedure, everyone on that table does have the disease. And when you assume a

[--- Unable To Translate Graphic ---]

few things that statisticians like to assume, what comes out of that is an unbiased--not reflection of the sensitivity and specificity, but an unbiased reflection of how the sensitivities compare between the two tests. So this is a very good approach for half-gold reference procedures.

Next.

Okay. Well, that was the design stuff, and that's the hard stuff. Now, the easier stuff is statistics. And there are just a couple of things that I want to--don't write that note yet. I'm going to try, but there are just a couple of things I want to talk about as far as analyses are concerned. One of them is: How do we summarize the results of these studies in which we are looking either at the performance of a new device relative to a reference device or relative to an existing device?

Well, at first blush, what comes out of the statistician's mind, anyway, is the comparison should be made using the sensitivity and specificities of the test. But also in my experience, the predictive values are very often something people like to see in package inserts as part of a reflection of how well a device functions or to compare two devices and their function.

From a statistical point of view, we have little

[--- Unable To Translate Graphic ---]

to say about this distinction. You may be surprised that the statistician in me doesn't care whether you talk about the sensitivity and specificity or talk about the predictive values because when we think about the performance of two tests or the performance of a test relative to a reference, it doesn't make any difference. Statistical procedures, the procedures we use to take chance into account are absolutely identical for those two. So the issue is not a statistical one but a clinical one. It's an issue of communication rather than an issue of statistics.

The next thing, what you see a lot in the research literature, is the odds ratio. The odds ratio, I'm sure that everybody is familiar with the odds ratio. It's something that we interpret as being the ratio of two risks, and in research that has to do with diagnostic procedures, odds ratios can tell us about the risk of being positive on a test for people who are presumed disease positive compared to people who are presumed disease negative.

In that application, I don't think the odds ratio is very helpful. In that application, that application comes from that case control kind of design in which you calculate an odds ratio from each of those two-by-two tables. And then the odds ratio reflects a combination of

[--- Unable To Translate Graphic ---]

sensitivity and specificity. It blends the two, and the particular blend depends on how prevalent people with the condition are in your sample. So I don't think the odds ratio is very good there.

But in the paired study, the odds ratio might be a very good choice. In the paired study in which we have some people who are presumed disease positive and another group of people who are presumed disease negative, those two-by-two tables compared the sensitivities and the specificities of the test, respectively. There the odds ratio can be used to tell us about how the two sensitivities compare or the two specificities compare.

A very special thing about the odds ratio is that the odds ratio and only the odds ratio, as a method to compare those paired data, will reflect not only the relationship assuming statistical independence, but it will also tell you whether that--how well that assumption fits reality. The odds ratio gets bigger as errors get correlated. That's not true of other sorts of ratios and differences.

The fourth thing I have there is an ROC curve, and I don't remember whether the guidance document mentioned ROC as a possible way of summarizing results. But I'd like to

[--- Unable To Translate Graphic ---]

give you some idea of my feeling about the ROC curve. My feeling about the ROC curve is that we expect too much of it.

The purpose of the ROC curve, the reason that ROC curves were originally fascinating to methodologists when they were looking at the performance of diagnostic tests was that an ROC curve can help us to find an appropriate cut-off that corresponds to a particular disease prevalence, corresponds to a particular risk/benefit ratio.

Somehow we have gotten into the business of comparing ROC curves by comparing the areas under those curves. Statistically, that is okay because those areas reflect how well those two diagnostic procedures perform throughout the range of possible values. Unfortunately, I don't think that's relevant when you're using a diagnostic test clinically. What's important is to specify what the conditions are for the particular application. Are you screening? Are you ruling in disease, ruling out disease? What is the cost/benefit ratio? And then there is going to be either the two procedures--the two procedures that are compared in that way are either going to be similar or one's going to be better than the other in that particular circumstance. And that is not to say there is going to be

[--- Unable To Translate Graphic ---]

that same order for other circumstances. So I think we have to be careful about how much we try to get out of ROC curves.

I certainly can't stop talking about statistics until I at least say the p-value word. What I'd like to do is I'd like to propose that there are two approaches that we might consider for analysis of data that comes from these kinds of studies that are described in the guidance document. And one of those approaches is hypothesis testing, and it's almost gotten, I think--I'm afraid for a lot of us it's gotten to the point of being no longer cerebral, but totally spinal, that when we want to take chance into account, we calculate a p-value. And the advantages of this approach are that it's familiar and also we're not going to have much argument about a decisionmaking rule--p less than or equal to 0.05, we reject; p greater than 0.05 we don't.

Unfortunately, the same characteristics I think are the disadvantage of hypothesis testing that hypothesis testing boils us all down to one number, to a p-value. I mean, that's basically what we get out of it. And then from that p-value we make mostly a dichotomous decision instead of some sort of quantitative evaluation.

[--- Unable To Translate Graphic ---]

So I think that we need to ask ourselves is it a good idea to boil this information down to just one number, or should we be looking at things that reflect separately sensitivity and specificity, as well as other things.

Another big disadvantage is that hypothesis testing really doesn't provide for the conclusion of similarity, and that's something that's very important in reality. That, as a matter of fact, was one of the main things that struck me when I had people in my NIH course from the Center for Devices and Radiological Health, was that you're going to have times, lots of times in which you're interested in showing similarity, not difference.

Okay. This looks familiar to everybody, I'm sure. This is that good old two-by-two table that is part of the one of the first lectures in any beginning statistics course that tells you about what can happen when you do hypothesis testing using classical hypothesis testing. And what this does is it gives us two possible conclusions. We can accept the null hypothesis as being true, or we can reject that null hypothesis.

The null hypothesis for our sorts of studies are that the performance of two devices, for instance, are the same. So we could either believe in that or stop believing

[--- Unable To Translate Graphic ---]

that. Those are the two conclusions that we can draw.

There are also two versions of reality. The two versions are that the null hypothesis is essentially true and that the null hypothesis is importantly false.

Now, there are four things that can happen with those two possible conclusions and those two possible versions of reality. Two of them are fine because we've done the right thing. We believe in the null hypothesis when it's substantially true, or we reject it when it's importantly false. The other two, unfortunately, are mistakes. And statisticians, one of the things that they're--they're good at math, not so good at naming things. These are called Type 1 and Type 2 error, or maybe a little bit more descriptive is alpha error and beta error because alpha and beta are the probabilities of making these errors. Alpha is that 0.05 we use to evaluate our p-value. Beta is that thing that we use in our sample size calculation.

Well, this makes statisticians uncomfortable, looking at this two-by-two table. Part of it is okay. Part of it is okay because if our alpha is going to be equal to 0.05, I can tell you what your chance is of making a mistake if you reject the null hypothesis and the null hypothesis is true. It's 5 percent chance. Being a statistician is not

[--- Unable To Translate Graphic ---]

trying to avoid making mistakes, but being a statistician is trying to understand the probability of making mistakes and controlling that as much as possible. So I'm very comfortable with rejecting the null hypothesis because you're going to make a mistake if alpha is 0.05 5 percent of the time. I'm very uncomfortable about the other.

You know, when we do that sample size estimation, where does that beta come from? The beta is the probability of making a Type 2 error. Beta is the complement of statistical power. It doesn't come from anywhere except our imagination. We never know what beta is.

In sample size estimation, we estimate sample sizes for betas that we would like to imagine, but there's not a way that we can actually calculate the beta error, and this makes statisticians very uncomfortable.

In classical hypothesis testing, what I am morally obligated to do as your statistician is to tell you don't accept the null hypothesis as true. Because if you don't accept the null hypothesis as true, if you avoid that, then you'll never make a Type 2 error. And I don't want you to make a Type 2 error because I can't tell you what your chance is of making that Type 2 error.

So, well, that's kind of depressing, but

[--- Unable To Translate Graphic ---]

statisticians over the years have developed something that they teach you to say so that you don't feel like you're wimping out when you can't reject the null hypothesis. What you do is you fail to reject the null hypothesis.

Well, this is not very good for our applications in which we are often interested in showing similarities. So what I'd like to propose is a different way to take chance into account--which, of course, is not my invention, but just my suggestion.

The other way that we statisticians take chance into account is by calculating confidence intervals or, in official statistical lingo, through interval estimation. There are a couple of advantages of interval estimation. One is that how wide that confidence interval is tells us the information about how precisely we are able to estimate the value that we've calculated the confidence interval for. The interval estimation parallel to statistical power is reflected in the width of that confidence interval.

Another good thing about a confidence interval, instead of a p-value, is that you have a p-value, you're pretty much tied into whoever analyzed the data's null hypothesis. With interval estimation, you're not.

If you have a different value that you'd like to

[--- Unable To Translate Graphic ---]

make a decision about, you'd like to find out whether or not this is a likely or unlikely value for what it is that you have studied in your sample, then you can do that with a confidence interval by taking that value and seeing where it occurs in a continuum of values and interpret the result from the confidence interval.

Major disadvantage, though. The disadvantage is that what we'd need to do at the onset--and now that I no longer work for the FDA, I guess I can speak for the FDA now, right? Because I couldn't when I worked for the FDA. But my impression, now in the public sector, is that this is also a trend in the way that statisticians at the FDA are thinking about things, that they're getting more away from the hypothesis testing kind of approach and more to the interval estimation, confidence interval kind of approach. But it puts a heavy burden on the FDA and also on the sponsor of a particular device that you have to specify what equivalence means. Because equivalence can't mean absolutely the same. That's silly. That's hypothesis testing.

What we have to do is we have to say that equivalence means within 5 percent or within 10 percent or something like that. This is something that people in CDER,

[--- Unable To Translate Graphic ---]

the drug side of the FDA, have struggled with for generic drugs; that they have to specify how close the parameters that they're measuring have to be in order for a generic drug to be considered biologically equivalent to the drug that they're trying to develop the generic for.

This isn't something that we've had--that's been struggled with, I think, as far as diagnostic devices or devices in general are concerned. It's a very tough thing. It's not that interval estimation really is creating this problem. What interval estimation is doing is it's having to make us confess that that null hypothesis isn't the only thing that's important, that we have to think about things that are not identical but things that are close enough from a public health point of view.

Is that the last one? Then I must be done. Okay. My students are right. I couldn't do it.

CHAIRMAN EGLINTON: Okay. Thank you very much.

We'll be just about 15 minutes behind schedule, but we have a little bit of a make-up this afternoon as well, so we'll be all right.

Let's be back at 3 o'clock, 15 minutes.

[Recess.]

CHAIRMAN EGLINTON: All right. Let's begin again,

[--- Unable To Translate Graphic ---]

and again, please, when presenters come to the podium, please identify yourself fully and your source of funding for today's visit.

We have industry presentations now. First will be Dr. Stewart Lonky.

DR. LONKY: Thank you very much.

My name is Dr. Stewart Lonky. I am a board-certified specialist in internal medicine and a fellow of the American College of Physicians. I have been the chief medical officer of the Trylon Corporation for the past 8 years. In this role, I have been in charge of prospective research and clinical trials regarding Pap Plus Speculoscopy.

I would like to take this opportunity to thank the members of this panel and FDA for allowing me to present some of the comments from our company regarding the draft document for in vivo devices now before you.

The first and most important factor I wish this panel to consider is that this draft document provides an opportunity for FDA to address an issue which has been a source of confusion in both the professional and the public sector. While I believe that FDA has taken a step in the right direction with this draft document and its requirement

[--- Unable To Translate Graphic ---]

that biopsies of the cervix be the "gold standard" for the definition of cervical pathology, I am convinced that this panel and FDA must evaluate this document and its protocols in light of the entire cervical cancer and precancer marketplace and the public perception of it.

While the new in vivo devices regulation would appropriately set biopsy-proven cervical pathology as the gold standard for test performance, the in vivo devices will be using a metric that is quite different from the metric that in vitro devices cleared for Pap smear screening have had to measure up to.

I can guarantee that unless this situation is acknowledged and rectified, the result will be further public confusion concerning the messages sent by FDA.

To review this situation, a number of in vitro devices that were ostensibly developed for the laboratory marketplace have been cleared over the past few years. Essentially, they are computerized devices and slide preparation systems designed to improve the clarity and interpretation of Pap smear slides.

FDA has appropriately cleared this device for marketing as quality assurance backup devices for laboratories or as systems capable of producing cleaner

[--- Unable To Translate Graphic ---]

slides that are easier to interpret.

Unfortunately, the public sector in the medical community equate the Pap smear laboratory test with cervical screening in general, and FDA clearance of these devices has led to the public belief that cervical screening is at least being addressed by these technologies.

Promotional literature has made statements such as, and I quote, "the new computerized testing that can find the precancerous cells missed by even the best regular Pap smear screening," or "the device was cleared by the U.S. FDA in 1996 as a replacement for the conventional Pap smear...is significantly more effective than the Pap smear, improving the detection of...lesions...in screening populations." Although these claims may be compliant regarding their clearances, the public, the press, and physicians all believe that cervical screening is being fixed.

This panel is being asked to approve protocols for in vivo tests designed to be done in conjunction with the Pap smear that will use cervical biopsy as the method for defining the presence or absence of cervical disease. The Pap smear devices mentioned above were never tested to this metric. In their PMA applications, the experimental protocols looked at negative or normal Pap smears from

[--- Unable To Translate Graphic ---]

multiple centers and had them over-read by experts and then tested the ability of in vitro devices to find the missed slides.

Alternatively, experimental protocols allowed for the preparation of a Pap slide to be done either in the conventional method or by the new in vitro method. In no case was cervical pathology used as a gold standard. In essence, the Pap smear served as the gold standard for the Pap smear. Furthermore, these studies were done on normal archived Pap smears enriched with known abnormal Pap smears. There was no study of a screening population.

Now this panel will be designing, along with FDA, protocols that look at a study screening population and determine the true prevalence of cervical pathology by biopsy as best as can be done. These protocols will measure the ability of the Pap alone or the Pap plus an in vivo device to find these biopsy-proven abnormalities.

When a similar protocol has been followed in studies where all women were colposcoped and then biopsied, the overall sensitivity of the Pap smear has been determined to be between 20 percent and 45 percent in screening populations. These data have been criticized as being impossible or as representing an aberrant population of

[--- Unable To Translate Graphic ---]

women.

The more accurate experimental design of using cervical biopsy as the gold standard has led to companies having to defend data that are at odds with data derived from studying slides rather than cervixes. The public sector and the medical community have been faced with conflicting and confusing data regarding a test, the Pap smear, that they have always believed was equal to cervical screening, much the same as Kleenex has been equated to facial tissue. This situation will be repeated when the experimental protocols outlined in this proposal are put into place for the new in vivo devices without its parallel being put in place for in vitro devices.

This panel has the opportunity to recommend that the apples-versus-oranges situation come to a halt. It makes sense that the public getting cervical screening, as well as the medical providers performing cervical screening, be presented with device clearances that measure up to the same standards.

Cervical biopsy is the correct standard for today, since it is the condition of the cervix that we are ultimately interested in rather than the condition of a slide. FDA has the requirement to provide efficacy and

[--- Unable To Translate Graphic ---]

safety data to the public that will allow individuals to make informed decisions regarding these new tests, and this information base cannot be allowed to continue to come from two sides of the FDA that address the same clinical issue, cervical screening for pathology, but use different gold standards for the identification of patients with disease.

I urge this panel on behalf of the companies that are dedicating time and resources to solving these problems, on behalf of the medical practitioners trying to deliver the best clinical case they can, and on behalf of the women who deserve to be accurately informed about these technologies to recommend that in vitro devices be held to the same standards as in vivo devices so that the information being released to the public can be accurately evaluated.

A second area of my concern and the company's concern is noted on page 10 of the proposed guidelines under the Hypothesis section. Here, it is stated that the hypothesis of study should be, "The combination of the Pap smear and in vivo detection device detects more patients with LSIL," or worse, "than the Pap smear alone, and there is not a significant decrease in specificity." This requirement will be very likely impossible to meet, since the requirement, if you think about it, is that the new in

[--- Unable To Translate Graphic ---]

vivo test alone have a nearly 100-percent positive predictive value.

Significant increases in sensitivity are always accompanied by an increase in the number of women without the disease who are being identified as having an abnormal test result. Therefore, there should be an expected increase in false positive rates, and in many instances, this can lead to a lowering of specificity that is statistically significant.

The question that FDA and this panel should be interested in is will this be clinically significant, and if so, when will it become clinically significant.

This panel has to consider that we are looking at data for a screening test, designed to be used on an asymptomatic population presumed to be free of cervical pathology. It should also be remembered that we are not talking about starting a new testing program in a vacuum. There is already a test, the Pap smear, and its score has to be considered when new screening tests are evaluated.

The first issue we must address is what are we trying to do by screening a population, anyway. It is my contention that we are trying to correctly and reliably identify the women with cervical cancer and precancer and

[--- Unable To Translate Graphic ---]

the women who are completely free of cervical disease. The first group may require immediate attention, while the second group may require less frequent attention. To this end, the proposed protocol must give reviewers and in vivo device developers objective guidelines regarding overall test accuracy or performance that can be calculated or measured for each device.

We proposed that in addition to sensitivity and specificity, this panel recognize the importance of accurately identifying true negatives. A new in vivo test, plus the Pap smear, when compared with the Pap smear alone should take this dependable identification of true negatives into account.

A term, therefore, such as "overall accuracy," which would be the true positive detections, plus the true negative detections divided by the total population is a calculation that will answer the question that really indicated the overall cost of adding a new test to the Pap smear.

Clinicians can easily see that a significant increase in sensitivity that is accompanied by a decrease in specificity can be accompanied by either a significant drop in accuracy or no change in accuracy. These three

[--- Unable To Translate Graphic ---]

measurements would provide information regarding the safety and efficacy of a new in vivo device that could be used to describe all of the devices with the same baseline metric.

A third area of concern that I wish to address is also noted on page 10. Once again, I am addressing studies designed for a product that will be making claims to improve the screening sensitivity for identifying women with cervical pathology in the general population. In the section titled "Sample Clinical Study Design," it is recommended that in prospective studies, and I quote, "if either the Pap or the in vivo device is positive, the patient is scheduled for colposcopy," and obviously, eventually biopsy.

In the section titled "Data Analysis," it is remarked that there should be a comparison of "relative sensitivity and positive predictive value of the two devices."

Now, it should be obvious that the suggested protocol will always result in the Pap plus device having a relative sensitivity of 100 percent, since the requirement is that only if one of these tests is positive does a colpo ever take place. Now, this will inevitably be greater than that of the Pap smear alone, thus establishing an increase

[--- Unable To Translate Graphic ---]

in sensitivity, but I would ask the panel to consider the fact that this protocol is heavily biased to positive results and will lead to misinformation. After all, how can you compare one 100-percent relative sensitivity with another 100-percent relative sensitivity unless the initial studies of the devices are done in populations where every single woman gets a colposcopy and a biopsy? It is not difficult to see that in order to compare one in vivo or in vitro device to the next, there would have to be a measurement of either one test against the other or, more reasonably, a measurement of each case against the true number of cases of cervical pathology in their study population.

The panel should recommend that initial studies look at the true sensitivity of a new in vivo test or a new in vitro test, plus the Pap smear, and that the sensitivities be compared with the sensitivity of the Pap smear alone for the detection of LSIL, as proven by biopsy.

Only when this true number has been established can studies be conducted that then look at the relative sensitivities, as suggested in the current proposal. This will allow both the medical community and the public to compare one technology with the next. The panel should

[--- Unable To Translate Graphic ---]

choose to reduce confusion as much as possible.

The last major issue I wish to bring up at this panel is the discussion of the current proposal of the need for new in vivo tests when compared with the Pap smear or when used alone to show an advantage over the Pap smear in the detection of HSIL lesions.

In deference to Dr. Schiffman's presentation, I would recommend that although HSIL lesions are temporally closer to invasive malignancies than LSIL, the panel is urged to remember that these are screening protocols that we are discussing. Whether it is a first-level screening or whether these devices are designed to further enrich a population with an uncertain or ASCUS Pap smear, the level of detection should not be changed because the test is perceived to be capable of higher levels of detection or temporal relationship to disease, cancer.

If we are looking at devices that claim to be adjuncts to screening, then they should be asked to perform a screening function, and I would add to my prepared statement that I would urge the panel to continue to think of a PPD or a skin test for tuberculosis as a model screening test.

If we are looking at devices that claim to be

[--- Unable To Translate Graphic ---]

adjuncts, then they should be asked to perform a screening function and not the detection of high-grade lesions, just as you would not ask the PPD to be positive only in active cases of tuberculosis.

Once again, this panel should provide the guidance needed to ensure that we are judging screening tests as screening tests and that the public and the medical community is getting information that can be applied to each technology with the same understanding of the metric involved.

Before closing, I would like to raise a few points that I believe do need some attention. Many an obstetrician would advice against performing cervical biopsies on pregnant women. So I would recommend that pregnant women be excluded from the above protocols.

From our own experience in prospective studies of over 14,000 women using biopsy as a gold standard, it is my recollection that the panel consider that a 4-week period between screening visit and follow-up coloscopy is much too stringent, particularly given the current backlog of colposcopy cases in most medical centers and the normal lag time for receiving Pap smear reports.

Third, I would ask the panel and FDA to recognize

[--- Unable To Translate Graphic ---]

that in the case of a visual adjunctive technology, it will be impossible for the examiner to be blinded from the results of the visual exam that he or she performs on the patient. So you are not going to be able to blind all the studies.

Finally, I believe that the panel should consider that by not letting a patient know the results of the new test at the time that it is done and forcing her to be notified at the same time as women with only an abnormal Pap smear that she needs to return for a coloscopy, it may not be advisable. Although such a protocol may remove confusion regarding when patients were screened and when they were followed up, it will also remove the opportunity to measure what effect, if any, the inclusion of the new test has on the compliance of patients for follow-up. This matter may be of great concern in cervical screening protocols.

I would like to thank FDA for allowing this presentation today, and I would like to thank the members of this panel for their attention.

CHAIRMAN EGLINTON: Thank you.

Now we will have Dr. Michael Hirschorn of Polartechnics introduce Professor Malcolm Coppleson.

DR. HIRSHORN: Excuse me while I unpack my bag.

[--- Unable To Translate Graphic ---]

It's a long way from Australia.

My name is Dr. Mike Hirshorn. I am the Chief Executive, Marketing, for Polartechncis. My academic background is an M.D. with an MBA.

There are three of us here from Australia today; myself, Professor Malcolm Coppleson, whom I will introduce in a moment, and Karen Canfell, our Clinical Trial Coordinator.

The reason that we are here is because we have developed the Polarprobe, and we have been having discussions with the FDA about the best ways to put the Polarprobe through clinical trial to assess its safety and effectiveness.

As Australians it is really a wonderful thing to be able to come and work with the FDA, to go through this process to prove safety and effectiveness for the U.S., which is a much larger country and a much larger market than we have in Australia. Our products don't always turn out to be upside down just because they come from the other side of the world.

Polartechncis is an Australian public company founded in 1987. So we are ten years old. And our mission is the detection of cancer and precancer.

[--- Unable To Translate Graphic ---]

Polartechnics employs over 35 full-time people, including engineers, statisticians, clinicians, and collaborates with leading scientists all over the world.

Polartechnics recently entered into a commercialization agreement with Ethicon, a subsidiary of Johnson & Johnson, to introduce the Polarprobe to world markets, and a number of people are here from Ethicon today representing the scientific, clinical, and regulatory, and reimbursement side. It is a pleasure to be here working with them as well.

The field of in vivo cervical examination promises to revolutionize the screening and detection of cervical cancer by real time and accurate detection. The possibilities for improvement are enormous. in vivo techniques have the potential for saving both lives and saving money.

Many of you haven't seen the Polarprobe and the Polarprobe console. So I will just show it to you to give you an idea of the sort of size and dimensions of the kind of thing we are talking about. In my right hand is the probe. It's applied to the cervix of a woman during a gynecological examination. In my left hand is the console that examines the data and comes up with the tissue type

[--- Unable To Translate Graphic ---]

classifications.

So you can see it is a portable instrument for use in real time in a primary care situation. Perhaps I will put it here for those of you that would like to look at it.

We have had discussions with the FDA since 1995, as we brought the development of the device closer to its conclusion. And we have recently submitted clinical protocols for evaluation. These cover the use of the Polarprobe for triage to colposcopy, as an adjunct to the Pap smear and as a stand alone screen.

We have made a written submission, which has addressed many areas of the draft guidelines, but perhaps the most important we would like to discuss is the choice of a reference diagnosis, and this is the area that Professor Coppleson will address in most detail.

The choice of Reference Diagnosis ultimately determines the validity of the clinical trial. Our understanding of the guidelines, as we read them, that, as expressed at the moment, the guidelines don't specify choice of reference diagnosis, and so we welcome the opportunity for talking about it and sharing it with experts here from around the world and from many disciplines.

I would like to introduce Professor Malcolm

[--- Unable To Translate Graphic ---]

Coppleson now. Professor Coppleson is a co-inventor of the Polarprobe, but is probably better known for his contribution to the gynecologic oncology field, most notably in the field of colposcopy, and its contribution to the modern understanding of precancer of the cervix.

Professor Coppleson is also clinical director of Polartechnics.

Professor Coppleson's presentation will be in two parts. He will first discuss the design and operation of the Polarprobe to give you a little more background on how it works as applied to clinical trial design, and then he will comment on the IDE's mission guidelines, in particular to the choice of Reference Diagnosis.

A Reference Diagnosis is, of course, needed for the feasibility study and for the trials for the four intended uses defined in the guidelines on pages 7 to 13.

Professor Coppleson?

CHAIRMAN EGLINTON: As Professor Coppleson comes to the microphone, point out, please, our issue here is really the guidelines, and we are only scheduled for 15 minutes.

DR. HIRSHORN: Absolutely.

DR. COPPLESON: Thank you, Mr. Chairman. It is a

[--- Unable To Translate Graphic ---]

pleasure to be able to talk on these hallowed grounds, and I appreciate it very much.

My slides do contain a few introductory. Would you prefer I pass through these? It will be a short presentation.

CHAIRMAN EGLINTON: It's your choice, sir, but we are scheduled for 15 minutes.

DR. COPPLESON: I won't be longer than 15 minutes.

The major difference between the Pap smear and related technologies and the new in vivo detection devices is that the latter, including the Polarprobe, screen the cervix in real time.

The diagnosis by the Polarprobe represents the recognition of what we term the signatures of cancer, various precancers and normal tissue by virtue of their electrical properties and optical properties at various wavelengths.

You have already seen the probe, and the console, and the dedicated wire system. The scanning of the cervix takes about one to two minutes.

This describes briefly how the probe works. The probe, in contact with the cervix--on the left side here, this is the console here--the probe, in contact with the

[--- Unable To Translate Graphic ---]

cervix emits a series of tiny electrical charges and three wavelengths of light; red and green in the visible spectrum and infrared.

Not surprisingly, the issue responds to this stimulus, and the response signals are taken up in the probe, carried across to the first box in the console and digitized into an algorithm representative of the action on the cervix here.

From here this algorithm goes into the memory and decision-making box of the console, which contains the algorithms of 17 different cervix tissue types, subclassified into normal, precancer, and cancer. The diagnosis is made instantly, and the operator informed, and the sequence is restarted.

To develop the algorithms involved a prodigious amount of mathematics. Here we see the use of two discriminants, two different wavelengths of light, 14 tissue types--we now have 17--and you will see the scatter here, and the console has no difficulty in distinguishing between these types here.

You note the cluster of tissue types down in this corner of the graph. And if you notice COL2, which is normal mucus-secreting columnar epithelium, D1, a precancer,

[--- Unable To Translate Graphic ---]

D2, high-grade precancer. The console is very likely to have difficulty in distinguishing between the normal and the abnormal.

By the addition of a third discriminant, in this case an electrical parameter, you will notice that the D1 and the D2s, the precancers, are pulled apart from the COL2 making it less likely for false positive to occur.

Now, clearly, the more discriminants you have the more accurate the diagnosis, and currently we are using 15 different discriminants in the diagnosis, and these pull these tissue types further apart.

This is the electronics of the handle of the probe. The three LEDs here emit light into the optical fibers, which pass down through the probe to the tip, and here is a diagrammatic representation of the tip of the probe as it lies on the cervix tissue. You can see the light-emitting optical probes labeled red, green, and infrared, and here you see the light detector, which picks up the backscattered light from the tissue and transports it back to the console.

Here are the three peripherally placed electrodes, which are important in the electrical measures. All of this data--electrical and optical--is then taken back to the

[--- Unable To Translate Graphic ---]

console for analysis.

This is a brief description of the flow chart for each single observation in terms of the Polarprobe. The first is a confirmation of good probe contact. If this is not present, no diagnosis is made, and the operator is immediately informed of the poor contact.

The tissue is then assigned to the most likely of the 17 tissue types within the console. A validity check is then made. There is reclassifying for screening purposes of these 17 types into invasive cancer, high-grade disease, low-grade disease, and normal.

The operator is signalled in real time by indicator lights within his peripheral vision and also display on the console. The sequence is started every fourteenth of a second.

We have examined over 3,000 women in Sydney, London, Manila, Singapore, Recife, Brasil, and Moscow. We believe that the advantages are the real time tissue diagnosis. Despite its sophistication these methods are comparatively simple. We have proven it to be more acceptable to women in the study performed in London. They are not labor intensive because it is only the woman and the user and, for the same reason, it is cost-effective.

[--- Unable To Translate Graphic ---]

I would like to recommend that the reference diagnosis for in vivo methods for abnormalities is not just histopathology, but histopathology plus colposcopy.

Most gynecologists and pathologists believe that the opinion of the histopathologist is definitive in diagnosis of cervical neoplasia. Now, while this is undoubtedly true, and it's been known to be true for over 100 years with unambiguous clinically invasive disease, there is enough evidence around to say that this is not necessarily the case with the precancers. The major problem is that of subjectivity and interpretation.

The great Leopold Koss wrote, "There is no publication on the subject where one couldn't reshuffle the photographs and substitutes pictures labeled dysplasia for those labeled carcinoma in situ and vice versa."

There have been several studies which have shown these interobserver differences. There are also intraobserver differences. Cocker, Fox and Langley found that there were major differences in diagnosis when the same set of slides is examined serially by the same pathologist, whether the interval be hours or days.

These are the last such studies that I can find in literature, both from 1989, and the results were identical

[--- Unable To Translate Graphic ---]

enough to make the one slide. There was both inter- and intraobserver variation and poor agreement between observers in the diagnosis of HPV, CIN1 and CIN2, and moderately good agreement in CIN3. To my knowledge, nothing has changed since this time.

Colposcopy has brought about some remarkable changes in concepts concerning the natural history of cervical cancer because it vests the clinician with the powers of direct observation of those very same stages as seen by the microscopist, but seen in truly in vivo conditions.

It is perhaps not the advent of colposcopy because it has been around since the 1930s, but rather the understanding of the central importance of the most obvious feature of the colposcopic image, the transformation zone. The transformation zone is one of the major features in understanding cervical neoplasia.

What colposcopy does is display image qualities, such as color, such as blood vessel configuration, such as surface configuration and topography, all images in living tissue that are not apparent in the microscopic image. Thus, colposcopy complements histology and it's important to realize colposcopy has an authority in its own right.

[--- Unable To Translate Graphic ---]

Colposcopy hasn't just been a simple device for the clinician. It has been a real value to the histopathologist. The colposcopist's transformation zone in vivo is the histologist's metaplasia in fixed tissue. That is the transformation of glandular epithelium to columnar epithelium.

What colposcopy demonstrated is that, basically, there are really only three histological types. The first is fully differentiated squamous epithelium, which is not shown here, the original type which changes little from fetal life until senescence.

The second is the glandular or columnar epithelium seen here, and the third is the metaplastic epithelium, a new squamous epithelium, which is derived from the glandular epithelium.

An early phase in the metaplastic process is the development of an undifferentiated eight to ten cell epithelium, a perfectly normal step in the metaplastic process, but one immature metaplasia, which causes great diagnostic difficulties histologically.

Most commonly, the process is normal and proceeds along a normal path. And after passing through a myriad of possible intermediate stages, represented by different

[--- Unable To Translate Graphic ---]

histological appearances, ends up as a fully differentiated squamous epithelium, sometimes indistinguishable from the original epithelium laid down during embryo genesis.

Less commonly, the process proceeds in an abnormal direction and again after proceeding through a myriad of possible histological appearances, the various precancers may eventually end up as invasive cancer.

So that none of these appearances histologically are endpoints themselves, they all form part of a single process.

I share this slide to show that the mere study of histology does not tell all about an epithelium. Here you see an epithelium which is full of what you might term cancer cells, a typical carcinoma in situ. This section, which has been shown to the best histologists in this country and elsewhere in the world, and I never got an answer other than, "This is a carcinoma in situ."

This, in fact, is a section of a one-day-old neonate and illustrates--this is one of Ellis Pixley's series, of which there were several along similar lines--and it illustrates the problem with immature metaplasia when looked in a moment in time.

It's appropriate to preserve tissue that cytology

[--- Unable To Translate Graphic ---]

is matched with histopathology. It is surely appropriate with living tissue that the in vivo devices are matched with colposcopy.

The strength of histology is it measures structural or static changes in disease processes. The weakness is that it is incapable of measuring with the same precision dynamic changes in the transformation zone as they proceed towards their endpoints. The strength of colposcopy is that it displays those very dynamic changes in vivo within the transformation zone. The weakness of colposcopy is that, like histology, it has some degree of subjectivity.

So, finally, I would like to suggest that the aim in trials such as the ones under discussion of new in vivo devices should be to get as close to the truth as possible' that a combined histopathology and colposcopic assessment is the preferred diagnostic method; and that such collaboration between gynecologists and histologists is already in practice, permitting proper individualization of management and would be regarded in most leading clinics in this country as being best clinical practice as of today.

Thank you.

CHAIRMAN EGLINTON: Thank you. If there are no other presenters we don't know about yet, we will move to

[--- Unable To Translate Graphic ---]

Dr. Virmani, presenting for the FDA.

DR. VIRMANI: Good afternoon, Dr. Eglinton, panel members and the distinguished audience.

This afternoon I would like to go over with you our draft guidance document on the premarket testing of some new types of in vivo devices that use especially optical and electrical technology for detection of cervical cancer and its precursors.

This document was made available to the public on June 14th at FDA hearing here in Rockville, as well as through the Internet. We sent all of the panel members copies of the document a few weeks back along with several background articles. Copies of the document are available today outside on the table.

I would like to begin by acknowledging the working group that developed this document. Besides myself, Dr. Tillman is a biomedical engineer in the Office of Device Evaluation. Sharon Miller is an optical engineer from our Electro Optics Plant in the Office of Science and Technology. Dr. Robinowitz is a pathologist from the Division of Clinical Laboratory Devices.

And, finally, Diane Solomon, also a member of our panel today is a cytopathologist from the National Cancer

[--- Unable To Translate Graphic ---]

Institute. Dr. Solomons' extensive background, including being one of the original developers of Bethesda system for Pap smear classification was instrumental in the development of this document.

In addition, Mike Kuchinski, a microbiologist in my branch, provided input on device cleaning and disinfection, and Stan Lin, a biostatistician from our Office of Surveillance and Biometrics provided overall statistical input.

I have divided my presentation into three parts. First, I plan to give some background information on the conventional methods used for detection of cervical cancer. Next, I will discuss the regulatory approach that we have proposed for guiding manufacturers on how to bring these new devices to the market and, finally, and most importantly, I would like to walk the panel and the audience through our draft guidance, especially some of the key points.

I hope this will set the stage for tomorrow's panel discussions.

As you have already heard earlier this afternoon, the Pap smear has been used over 50 years as the primary screening tool for cervical cancer. It consists of three basic steps; first, cells are scraped from the cervix using

[--- Unable To Translate Graphic ---]

a spatula with or without a cervical brush. Most cervical neoplasia originates at the junction of the exocervix and endocervix. That is the transformation zone. Therefore it is critical that this area of cervix be sampled adequately.

Then the scraped cells from the cervix are transferred to a microscope slide and cell fixative is applied. A slide is then sent to a laboratory and, finally, the slide is read under the microscope by a trained cytotechnologist. All suspicious slides are confirmed by a pathologist and appropriately classified as to their type of abnormalities.

You have already heard that Pap smears are classified by the Bethesda system, which allows for a standardized cytologic identification of the cell sample from the cervix. The Bethesda system classification for different types of cells is, basically, as follows: normal, atypical squamous cells of undetermined significance, ASCUS, low-grade squamous intraepithelial lesions, LSIL, and high-grade squamous intraepithelial lesion, HSIL.

There are additional diagnostic categories in the Bethesda system, but this will suffice for our purpose. My purpose of reviewing the Bethesda system is that we will be talking about how to compare results from the new type of

[--- Unable To Translate Graphic ---]

device to conventional methods.

Also, our guidance document cross-referenced this method as one of the main points for comparison. The questions to be posed later on the panel you will see the Question No. 3(b) about Reference Diagnosis. I would like to focus the panel on this critical point.

In vivo diagnostic devices, IVDs, for detection of cervical cancer and its precursors are reviewed by our Division of Clinical Laboratory Devices, within the Office of Device Evaluation.

Besides the conventional laboratory devices used for Pap smear reading; that is, slides, cover slips, fixative, transport system, microscopes, et cetera, there are other new IVDs, including devices for making cell suspensions, for ten-layer or mono-layer slides and computer-assisted devices for searching or interpreting abnormal cells on the Pap smears.

We will not be discussing these IVDs today, however. Our deliberations today and tomorrow will obviously influence how, in general, we view IVD use to detect cervical cancer.

My branch, the Ob/Gyn branch, reviews the in vivo devices for Pap smears. This means a variety of cervical

[--- Unable To Translate Graphic ---]

spatulas and brushes applied directly to the cervix. Review of these devices are very straight forward.

FDA has cleared only one in vivo device, the speculite, that can serve as an agent to the Pap smear. Initially cleared in 1985 as an alternative light source for examination of the cervix, we cleared the same device in 1995 as an adjunct to Pap smear for a generalized claim of increased sensitivity. The claim did not address specificity.

During the past three years, FDA has become aware of the new types of in vivo devices using advanced optical technology that are also intended for cervical cancer detections.

These new devices differ from the currently used devices in that they are noninvasive and provide test results virtually real time. This is possible because of unique implementation of optical techniques. Earlier this afternoon Dr. Kortum and some of the sponsors using published literature and their own research described how fluorescence spectroscopy, Raman spectroscopy and light sketching techniques can probe the biochemistry and in some cases the morphology of cervical tissues.

I will briefly mention a couple of device

[--- Unable To Translate Graphic ---]

examples. Dr. Kortum already described them and copies of relevant general articles on them are included in the background packages we sent to you. I don't think that I am disclosing anything that is not in the public domain.

At this point in time, we have not approved any of these type of devices for cervical cancer detection. This draft document was developed in anticipation of these products. One device uses both the light source and an electrical energy source. It may just [inaudible] and a sketching of various wavelengths of light, and these variables are processed in real time through a discriminant analysis algorithm based on colposcopy, cytology, and biopsy evaluations. The probe directly contacts the cervix and results are given instantly.

Another device is a noncontact probe that uses only a light source to elicit an auto fluorescence and spectral backscattered response from the tissue. An algorithm defines the detection paradigm. Again, the results are given instantly.

These are just two examples, and you can be sure that there will be others.

Although these new devices use different types of optical and/or electrical energy sources, they all share the

[--- Unable To Translate Graphic ---]

following common characteristics: They all use a hand-held probe that houses the optical light source. The probe either touches or is held in close proximity to the cervix. They all employ some algorithm that takes the required signal from the device and processes the signal to arrive at a tissue type identification.

All of these devices employ a central processing unit with a hardware/software component that essentially runs the algorithm. Results are given within seconds of probe application on the cervix in some kind of discriminating display that differentiates normal from abnormal. Because of the simplicity of use they can easily be used in office or outpatient setting.

I will now discuss the regulatory aspects of these new types of cervical cancer devices. In your package, along with the view graphs, we have provided a 510(k) processing chart, which may be helpful in deciding some of these things.

In the past couple of years, FDA has been asked if these new types of devices could be cleared through a 510(k) premarket notification process or whether it is necessary to submit a premarket approval application known as PMA.

To clear a medical device for market through

[--- Unable To Translate Graphic ---]

510(k) premarket notification, a manufacturer must show that their device is substantially equal to a predicate device; that is, a device that was not on the market before 1976 or has been found substantially equivalent since then. Devices that cannot be found substantially equivalent to a predicate device require a PMA.

To answer this question, we need to examine the 510(k) decision-making process to see if these new devices can be found substantially equivalent to a predicate device or devices.

The first question is does the new devices have same intended use as the predicate device chosen for comparison? If it does, then the next question is does the new device have same technological characteristics as the predicate device. If you believe the new device has different technological characteristics, then we have to ask do the new technological characteristics raise new types of safety and effectiveness questions.

Let's see what that means for this new type of device. For the first question we ask are the indications for use the same, comparing the new device to predicate device? We will assume for the purpose of this discussion that the Pap smear is a predicate device.

[--- Unable To Translate Graphic ---]

As you can see, some of these indications are quite similar to Pap smear, but some, like triaging or biopsy site localization, are much different. That in itself could lead to a nonsubstantially equivalent finding.

If we find the indication for use to be reasonably similar, we have to ask how similar is the technology. As you can see, we think there are several features of these new devices that are significantly different from a technology viewpoint, including optical and/or electrical energy sources, hardware and software, the integral algorithm contained with an instantaneous display of the results.

Finally, we have to ask do these technological differences raise new types of safety and effectiveness question; that is, compared to how we would evaluate devices used for the cervical Pap smear. Do we now have new types of questions? We clearly believe that is the case.

These two types of safety and effectiveness questions include what kind of bioeffects do the optical radiation and electrical pulses have upon the cervix. Another type of question might be how was the underlying algorithm developed and how was it validated.

In turn, more clinical type of new question, given

[--- Unable To Translate Graphic ---]

the instantaneous results available from the technology, how does this affect acceptable sensitivity and specificity?

We have given this new technology a good deal of consideration. You can see that we believe these devices, at least the ones we have seen, will be found not substantially equivalent. This means that these new in vivo cancer detection devices will require an approved PMA before they can be marketed.

The draft guidance document before you was developed to help manufacturers design the right kind of principle, preclinical and clinical studies that will support premarket approval.

Let's now turn to the draft guidance document. Clinical studies of medical devices must be conducted in accordance with our tabulation for investigational device exemption called IDE. A sponsor who wishes to conduct a clinical trial would submit an IDE application to ask for permission to begin.

This guidance document is intended to identify the types of information that we expect to see in an IDE application that could develop the data needed to support premarket approval of one of these in vivo devices that we have been discussing.

[--- Unable To Translate Graphic ---]

It is important to recognize that some device types may not require all of the information specified, while there may be other types of devices that require additional studies we haven't identified.

Let's talk about the preclinical studies that would be needed before beginning the clinical trials. This information will be submitted in the IDE application.

My own had listed all of the preclinical testing concerns that we expect to be addressed in an IDE. I will highlight only the first two of them this afternoon; the device design and description and the device performance.

When an IDE comes into the FDA, the IDE sponsor should fully describe the device design and particularly this should include a thorough discussion of principles of operation of the device. The application should contain a complete description of design specification, such as the light source or sources delivered to the patient, the basis for algorithm development, a description of user interface, including any parameters that the user can set, any safety features for patient and operator and a system-level hazard analysis. The sponsor should fully describe all testing models, along with the details of model validation.

[--- Unable To Translate Graphic ---]

The other area of preclinical that I want to mention today falls under device performance. By this I mean performance of the device in the laboratory and, in particular, questions about optical radiation.

In our draft guide document, we have highlighted several performance specifications and an IDE for one of these types of device should fully address all optical radiation issues.

As part of this, the sponsor should have conducted a system-level hazard analysis. The IDE document should contain a complete description of all safety features.

If we turn to page 4 of the guidance document, you will see that we ask the sponsor to describe the type of laser or light-emitting diodes used for the light source as well as key performance specifications such as wavelength, power, exposure time, the exposure site, it's pulse size and anatomical site on the cervix where light is applied.

If there are multiple sites on the cervix where the light is applied, this must also be explained. For broad-band light sources, the sponsor should provide either absolute spectral output or relative spectral output and absolute total power.

If the new device emits short wavelengths, UV

[--- Unable To Translate Graphic ---]

radiation at levels approaching the Occupational Safety limits, carcinogenicity studies or other types of safety studies may be necessary.

We strongly recommend that the sponsor contact us as early as possible in the clinical application we are discussing today and tomorrow. Optical radiation poses several safety concerns for the patients and clinician.

I would like to briefly discuss two ways FDA might deal with this.

Option No. 1 identifies two published references for optical radiation exposure limits. The first addresses exposure limits for lasers. It was developed by American Standard Institute, ANSI, in 1993, and it gives maximal permissible exposure levels or MPEs.

You have heard some talk on these this morning from Dr. Kortum. The second from the American Conference of Governmental Industrial Hygienists developed threshold limit values, TLVs, for exposure to UV-emitting lamps.

Optical radiation from these new cervical detection devices can be compared to these levels, but it is important to note that both MPEs and TLV levels were developed for the skin and not the mucosal tissues. This hasn't been added yet to the draft, and we would be

[--- Unable To Translate Graphic ---]

interested in the panel's view on this.

If comparison of these standards does not satisfactorily address our concerns about radiation safety, a second option would be to use a relative risk approach, including a comparison of expected mutagenic effect from the new device to currently accepted mutagenic risk for the other diagnostic procedure; for example, optical radiation on cervix from a colposcopic examination or ionization radiation from a chest X-ray.

To accomplish this, the IDE sponsor would need to perform a risk analysis and there are a number of approaches that can be taken. On our next draft we plan to provide additional guidance on such relative risk analysis.

One last area of radiation safety concern deals with the special circumstances. These include possible contraindications to use such as porphyria, lupus or other photosensitizing disease. Patients undergoing phototherapy, patients on prescription and nonprescription photosensitizing drugs, such as what is used for psoriasis. The IDE sponsor will need to address these types of concerns.

As I mentioned, there are several other preclinical areas that must be addressed; for example,

[--- Unable To Translate Graphic ---]

information on software development, manufacturing, material safety on parts of the device that contact the patient, et cetera. We will be developing these sections in our next draft.

Now let's turn to the clinical studies. That is on page 7 of the draft. I would like to highlight that the present guidance document is a new approach for our division. Previous guidance documents, such as the one the panel contributed to in 1995 on thermal endometrial ablation had fairly well-defined clinical objectives with clear-cut clinical outcomes to be measured.

By contrast, the draft before you today is for detection devices that have several possible different indications for use.

I will get into it more in a moment. But, as a result, our draft guidance lays out the principle clinical study design for the different indications rather than detailed study protocol requirements.

Although we certainly would appreciate panel input on the study details, we are more interested in how the panel sees the proposed study design principles for each indication.

With that in mind, we have proposed a two-phase

[--- Unable To Translate Graphic ---]

study approach. First, a feasibility study or a series of feasibility studies for safety and preliminary effectiveness data. Then, when sufficient data has been developed from the feasibility study, the sponsor should conduct the pivotal clinical study that would support premarket approval for the specific indications for use desired for the device.

If more than one indication is desired, additional studies may be needed. For all clinical studies, regardless of the indications of use, the following design principles should be followed. These are taken from the different parts of the draft as well as general FDA guidance on clinical trials. It hasn't been exactly organized in this way in the document, and we will be pulling it all together later after we have all your points of views.

Study subject selection/exclusion criteria ensure the expected range of clinical presentations of the cervix. They should reflect the indication or indications for use ultimately claimed for the device. Criteria might include factors related to age, parity, menstrual status, pregnancy status, previous cervical surgeries, et cetera.

I would like to hear some panel comments on how to test premenopausal women; that is, when during the cycle and how many cycles, et cetera.

[--- Unable To Translate Graphic ---]

The protocol must precisely define the study endpoints or endpoint and correspond to the intended indication for use. For the definition of the study endpoint and the study hypothesis, the sponsor would justify the sample size employed in the study.

Sample size calculations should be based on appropriate statistical techniques and result in adequate power to detect a difference between the new method and the comparison.

And the protocol must also spell out the management regimen for the study subjects, again, corresponding with the intended indications for use.

Other common elements to be addressed are:

A risks analysis. This would identify all potential risks to the patient and the likelihood for them to occur and how the study protocol minimizes these risks as much as possible.

Informed consent. This would be presented to the study subject and explain these risks to her in an understandable way. Informed consent, of course, must conform with 21 C.F.R. part 50 of the FDA regulations on this.

User training. The study protocol should validate

[--- Unable To Translate Graphic ---]

the minimum education and training necessary for the clinician performing colposcopy and using the new in vivo devices.

Let's move on to the specifics of the feasibility study. This starts on page 7.

The primary purpose of feasibility study is to validate device performance, including its ability to reliably detect cervical cancer and its precursor lesions. For devices that actually touch the patient, the study should also demonstrate that such contact does not damage the tissue.

A feasibility study also provides useful information on performance needed to estimate device effectiveness and, consequently, contributes through the calculation of sample size for the pivotal effectiveness study.

We have proposed a prototype feasibility study of 100 subjects. The patient population for this study should be women with a positive Pap smear who are referred for colposcopy. To ensure a reasonable representation of the type of patients, the study should include at least 25 patients from each ASCUS, low-grade SIL, and high-grade SIL. This should permit an acceptable estimate of device

[--- Unable To Translate Graphic ---]

performance that can, in turn, be used to develop the pivotal study hypothesis and consequent sample size needed to prove the hypothesis.

The prototype feasibility study we have proposed calls for an initial test with the new optical device followed by colposcopy. We also believe a repeat test with the new device is recommended. This repeat test would answer questions about whether the acetic wash of the cervix, generally is colposcopy, adversely affects the performance of the new device.

It would also provide some basic information on the clinical repeatability of tests. We would be interested to hear panel input on these testing sequences. We would also like to hear any recommendations on how to test whether the Pap smear itself, if done only moments before, might affect the new device performance.

In addition, colposcopy for these feasibility studies will be performed to find physical effects of the device on the cervix, including trauma and bleeding. Colposcopy will also be used to validate the results from the Pap smear and the new device.

If the device is intended to localize lesions, colpography or a similar technology should be used to

[--- Unable To Translate Graphic ---]

document specific sites on the cervix. Depending upon the results of feasibility study, the sponsor may proceed to the pivotal clinical study of the safety and effectiveness or the sponsor may need to redesign the device or refocus the indications for use.

We would appreciate panel's input on how to make these assessments.

Once the appropriate feasibility studies are conducted, the final step is to design and conduct the pivotal effectiveness study that would support premarket approval.

Let's turn now to the section of draft on pivotal studies of the safety and effectiveness that would support premarket approval. This can be found on pages 8 to 13.

As you can see, it is organized around the specific indications of use selected by the sponsor. The sponsor proposed indications for use will then determine appropriate study design needed to support the PMA approval.

My next overheard is a list of possible indications for use for these kinds of detection devices.

First, adjunct to Pap smear. This would be the use of in vivo devices together with Pap smear at the time of primary screening.

[--- Unable To Translate Graphic ---]

Second, triage of ASCUS for colposcopy. This would be the use of a device to help triage patients with a Pap smear finding to ASCUS to colposcopy; the premise being that not all of these patients really need colposcopy.

Next, used at the time of colposcopy, this would be the use of device as an adjunct to colposcopic examination to help select biopsy sites on the cervix.

And, last, replacement of the Pap smear as a primary screening tool for early detection of cervical cancer. This, obviously, is a fairly radical indication for use, and we would have to study this very carefully.

There may be other indications for these devices that may require other clinical study designs. We would be interested to hear any ideas the panel may have in this regard.

Companies planning to pursue combined indications for use should include a study design for each indications for use.

Now, let's look at how this works out indication by indication.

First, we have the example of using the device as an adjunct to Pap smear; that is, information from the new optical device is added to the findings from Pap smear for

[--- Unable To Translate Graphic ---]

primary screening.

The study design should specify whether the new device is used before or after the Pap smear is conducted. The study hypothesis here would be something like the combination of new device with Pap smear detects more patients with low-grade SIL or above than the Pap smear alone.

There should be no significant decrease in the specificity. Details of this proposed study design are included in the guidance document on pages 9 through 11. All patients will receive Pap smear and in vivo detection device during the primary screening examination.

If either the results of Pap smear or in vivo device is positive, the patient will be scheduled for colposcopy.

For this type of a study, the sponsor would need to compare sensitivity and relative specificity, as well as positive and negative predictive values of the two devices for ASCUS, low-grade SIL, and high-grade SIL.

FDA would not require a determination of absolute specificity for this indication because that would require biopsy validation of women who have a negative Pap smear, negative colposcopy, and a negative result with the new

[--- Unable To Translate Graphic ---]

device.

I know the draft doesn't exactly read like this, and we will be making some corrections. We haven't spelled out how much greater the sensitivity would need to be for premarket approval. We would be interested in panel's input on this.

For the next example, if the new device is to be used to triage women with a Pap smear finding of ASCUS for colposcopy or not, then the hypothesis would be something like the new detection device will identify a subpopulation of high-risk patients requiring colposcopic follow-up from a larger population of ASCUS patients who don't have a biologic reason for follow-up.

This indication is interesting because we may reasonably sacrifice some sensitivity to gain specificity. Some study design details are spelled out on pages 11 and 12. The study design should be able to determine the sensitivity, specificity, and positive predictive value of the new device for ASCUS and low-grade SIL and high-grade SIL.

For this type of a study, results from the new device should be validated against colposcopy and directed biopsy.

[--- Unable To Translate Graphic ---]

All patients are examined first using the in vivo device and then by standard colposcopy procedure. If there is a patient-contacting probe, evidence should be provided that in vivo device will not interfere with performance of results of colposcopy.

Again, the draft here calls for positive predictive value. We will be correcting this to ensure that the study develops data on sensitivity, specificity, and positive and negative predictive values. The important thing at the end of the day is for us to be able to completely convey the diagnostic performance characteristics.

The next possible indication for use for these new devices is to assist in the selection of site on the cervix for biopsy at the time of colposcopy.

In other words, the device must have some kind of localization capability. From possible study hypothesis, the new devices will identify sites on the cervix for biopsy as well as the acetic acid wash used conventionally for colposcopy.

Obviously, that hypothesis will need to have additional details built in. Some of the details of the design for this indication are included in the draft on

[--- Unable To Translate Graphic ---]

pages 12 and 13.

Study subjects should present with an abnormal Pap smear. The in vivo device is used to localize sites. Then an acetic acid wash is performed and the colposcopist would record the area of lesion and other areas that need to be biopsied. The results would then be compared.

The sponsor should document the cytologic criteria used in clinical study for referral of patients for colposcopy. The protocol should precisely describe how the clinician will determine and document that the device reading and the biopsy were taken from the exact same location and to compare in vivo device's results to the colposcopy results.

For this kind of a study, the sponsor would also need to compare the sites selected by colposcopy and the area selected by device that would not be selected by colposcopy.

Finally, our last example of a possible indication for use is a primary screening tool for cervical cancer and its precursors. That could mean replacement of Pap smear.

For such a break-through application, a study must demonstrate safety and effectiveness in all possible subgroups of women, especially older women or women whose

[--- Unable To Translate Graphic ---]

transformation zone may be obscured or we look at it totally within the endocervical canal.

This study must demonstrate that device is as good as the Pap with a high degree of confidence for both sensitivity and specificity.

In conclusion, these study designs I just presented represents some possible approaches for the clinical utility of this technology. Each indication requires its own clinical efficacy study demonstrating the safety and effectiveness of new in vivo cervical devices.

As always, reasonable alternative study designs will be considered by FDA on a case-by-case basis. We will also consider other reasonable indications for use.

Finally, I would like to emphasize that this guidance document is still evolving as science advances and as we learn more about the technology and how it can be applied.

We also expect to make major revisions to the draft after considering comments today and tomorrow from panel and the public, as well as any comments we receive during the 90-day comment period.

We look forward to your discussion tomorrow and for input on the various points I have highlighted when you

[--- Unable To Translate Graphic ---]

address the discussion questions we prepared, and we would like you to go through the document tomorrow page by page.

Thank you very much for your attention.

CHAIRMAN EGLINTON: All right. Thank you.

Do any members of the panel have questions for any of today's presenters or for Dr. Virmani?

DR. HARVEY: I'd like to interrupt just for a second. There's a set of keys that has been found up at the guard's desk, so everyone should check to see if they've lost their keys.

CHAIRMAN EGLINTON: Dr. Diamond?

DR. DIAMOND: I was very interested by Dr. Schiffman's presentation this morning in that--again, I'm a little bit out of my own realm, but at least what I remember when I used to look at this sort of this a little bit more was that the virus was a prominent part but not as prominent a part as his presentation and the literature that he provided to us would suggest. I was just curious as to whether I really missed the boat and I'm way out of date, which could always be, or whether other panel members felt the same way.

DR. SOLOMON: I'm sorry. I'm not really quite sure that I understand your question.

[--- Unable To Translate Graphic ---]

DR. DIAMOND: The question, I guess, is basically: Is it commonly accepted today that the true issue is viruses as opposed to histopathology?

DR. SOLOMON: Well, first of all, let me confess that I work with Mark on a lot of projects, and we tend to have the same viewpoint about this.

It's not really a question of the two really being separable. Cervical neoplasia is virally induced process, but if you view the spectrum of histopathologic changes, not all of them necessarily represent a true pre-invasive lesion. So that we're recognizing that there are lesser degrees of cytologic abnormality that correlate with HPV infection, and that if you look at higher-grade lesions that tend to be the precursor lesion to what may develop into invasive cancer, that also has a viral etiology to it, but is what we view as a lesion different from the lower-grade lesions that tend to regress over time.

DR. DIAMOND: Let me try it one other way. I almost took from his presentation this morning, rather than looking at a Pap smear, we ought to be doing HPV typing on everybody and having that be the primary determinant, with Pap smears and other endpoints such as what we heard about this afternoon being--

[--- Unable To Translate Graphic ---]

DR. SOLOMON: I don't think that Mark--

DR. DIAMOND: That's not what you meant to say?

DR. SOLOMON: No.

DR. DIAMOND: Okay.

DR. SOLOMON: HPV testing would be extremely sensitive, but you would identify a huge number of women who would be HPV positive who would not necessarily even have a cytologic abnormality or even ever develop a cytologic abnormality. But, Mark, you go ahead and answer that.

DR. SCHIFFMAN: What I meant to say was that high levels of HPV 16, for example, might be as much of a risk marker for a true cancer precursor, meaning high grade, than a colposcopic appearance of a simple aceto-white lesion or an LSIL Pap smear. They're all parts of the same process. The low-grade process is the signs of viral infection, on whatever level. You know, it's signs or microscopic or DNA. That neoplasia, in the sense of something as we recognize it with genetic alterations and a real propensity for invasion and everything starts with high grade, that's what I was trying to say. But I meant to say that we could attack it on any of those types of levels, whatever combination is the most cost-effective for the setting. I didn't in any way mean to say it was just--you know, because looking for virus

[--- Unable To Translate Graphic ---]

alone at the molecular level by PCR is the gold standard for sensitivity, and I'm repeating what Diane said now, but it's so non-specific that we're always counseling women, so you have HPV, you know, wait a couple months, you probably won't have it anymore, it goes away.

DR. DIAMOND: And that goes for even types like 16?

DR. SCHIFFMAN: I have the curves of disappearance in normal, initially normal women, and 16 goes away slower, which is probably one of the reasons it's worse, but it still goes away.

DR. LEVY: One of my concerns with the guidance document is that it talks about increasing sensitivity to the detriment of specificity, and we're going to get ourselves, I'm afraid, with this kind of requirement of the companies, into the same quagmire that we're in with the in vitro testing devices; that is, we're going to have a huge number of women, therefore, identified at risk, and then where do we go from there?

So I think, given Mark's presentation this morning--or this afternoon, that we should really be thinking carefully about how we draft this guidance document so that we get some clinically meaningful outcomes to these

[--- Unable To Translate Graphic ---]

devices as opposed to just finding every possible abnormality on the cervix that doesn't pass the clinical so-what test. That would be my goal as we deliberate over the next day or so, that we can really come up with something that's clinically more meaningful than anything we have right now.

My very great fear is that we'll come up with something like some of the in vitro devices that are going to vastly increase the number of women who are labeled at risk and increase the intervention that we do without changing the outcome as far as preventing invasive cervical cancer.

CHAIRMAN EGLINTON: Any other comments or questions from the panel?

MS. YOUNG: Is this the only opportunity to ask questions of sponsors?

CHAIRMAN EGLINTON: Well, we'll have--I assume probably most of them also will be here tomorrow as we go through our discussions of the draft document. But you can certainly go ahead and ask a question now. We have a few extra minutes.

MS. YOUNG: Well, in terms of the Polarprobe, I just wondered, as far as the device is concerned, what does

[--- Unable To Translate Graphic ---]

it require as far as maintenance is concerned? And what is its life expectancy?

DR. HARVEY: Don't forget to identify yourself.

DR. COPPLESON: Professor Coppleson from Sydney, Polartechncs. The life expectancy is thought to be at least two years. This has to be determined. As far as maintenance is concerned, we have rigid sterilization procedures that we go through. These are as for endoscopy instruments, and there will be what is called a single-use sheath which will be discarded after each use, is planned for the device.

Is that really what you wanted to know?

MS. YOUNG: As far as the sheath is concerned, sheaths are also used, I think latex ones, condoms or condom-like sheaths are used, for example, in transvaginal ultrasound devices. And apparently there's quite a high leakage rate in those sheaths.

DR. COPPLESON: This will not be latex. This is a rigid plastic sheath. It's rather high-tech in various ways. It's not what it sounds like. It's not a latex sheath.

MS. YOUNG: Okay. And one other question about maintenance. When I talk about maintenance, I'm talking

[--- Unable To Translate Graphic ---]

about the sort of device--it's got a number of different parts to it.

DR. COPPLESON: Correct.

MS. YOUNG: Would they be sort of checked on a regular basis routinely, monthly, daily, or whatever, depending on the use of a particular--one particular probe in terms of whether all the parts are functioning properly?

DR. COPPLESON: There is self-calibration at the beginning of each probing session for both the electrical measurements and the optical measurements. And this calibration will be constant on a regular basis, daily basis. Each probe can identify itself. And before the actual probing begins, the operator will go through in the session a sequence which will tell him that everything is calibrated, everything is working. There is also what is termed an operator error device, and that has on the handle of the probe a series of lights. And one green light has to be on which indicates the system is functioning well and everything is calibrated. Then the other system is two green lights, normal disorder, red light means the device needs to be repositioned and a series of blue lights indicate the degree of abnormality.

CHAIRMAN EGLINTON: I think the point would be

[--- Unable To Translate Graphic ---]

that any manufacturer would have to convince the FDA that there's a standard boot-up check sequence and an ongoing continuity check or whatever, there's some system of guaranteed ongoing safety.

Thank you, sir.

DR. KATZ: I just have a quick follow-up question, Professor. Does the self-calibration involve any materials external to the device?

DR. COPPLESON: You're talking to a gynecological oncologist, and for this reason, I'd like to introduce somebody else from Polartech, Karen Canfell, who can answer that question far better.

DR. KATZ: My question is, does the self-calibration--

DR. HARVEY: Excuse me. I'm sorry to interrupt. We don't want to get too far into the specifics of each individual device because we're trying to keep this on a generic level. Sorry.

CHAIRMAN EGLINTON: I mean, we can assume from the standpoint of marketing expertise that nothing beyond the console is going to require it for any particular implement. No manufacturer is going to require that you have to buy a \$75,000 cart to come in and plug your console into.

[--- Unable To Translate Graphic ---]

DR. KATZ: I was thinking more of some sort of standard material that was tested with the device as part of a self-calibration procedure rather than any physical--

CHAIRMAN EGLINTON: Are you talking about a phantom?

DR. KATZ: Precisely.

DR. LEVY: Mr. Chairman, I had a couple more issues in the guidance document that I just thought we should at least address. One is the safety for the operator or the clinician doing the procedure, particularly our eyes, and that that needs to be addressed with some of these devices, more than likely.

A second issue that I didn't see addressed in some of the papers was the potential for some of these applications of energy to change the natural history of the way the virus interacts with the cell. Just given that we are applying energy to cells that have a viral load, at least in some cases, I feel a little bit uncomfortable that we may not be changing the natural history of the disease by applying energy and would like to see some reassurance on that point.

DR. SCHIFFMAN: Studies on X-ray of--you would expect therapeutic X-ray of the cervix plus HPV would be a

[--- Unable To Translate Graphic ---]

major risk factor given that one's genotoxin, you know, high level therapeutic radiation, and in a field that is infected throughout the entire female genital tract, and yet you really don't see much additional vaginal or vulvar cancer risk in women with irradiation versus surgery. So sometimes people are talking about--it sounded like this was visible light with just infrared, so--I mean, some of them. I don't know what the other company is but--

DR. LEVY: I wasn't saying that there was a problem, but simply that this is something, as we draft a guidance document, that should be addressed in a PMA. It may be one paragraph that says exactly what you said. It's just something that I think in our guidance document should be in there. We have to assume that somebody may come up with some totally unique something that we've never seen before five years from now.

CHAIRMAN EGLINTON: Okay. We'll adjourn here for today and reconvene at 8:30 in the morning. Thank you.

[Whereupon, at 4:35 p.m., the meeting was adjourned, to reconvene at 8:30 a.m., Tuesday, February 15, 1997.]