TRANSCRIPT OF PROCEEDINGS

DEPARTMENT OF HEALTH AND HUMAN SERVICES

FOOD AND DRUG ADMINISTRATION

CENTER FOR DEVICES AND RADIOLOGICAL HEALTH

RADIOLOGICAL DEVICES PANEL MEETING

Volume II

This transcript has not been edited and FDA makes no representation regarding its accuracy.

Pages 1 thru 259

Rockville, Maryland
August 18, 1998

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

'98 AUG 26 A9:13
2103

DEPARTMENT OF HEALTH AND HUMAN SERVICES

FOOD AND DRUG ADMINISTRATION

CENTER FOR DEVICES AND RADIOLOGICAL HEALTH

RADIOLOGICAL DEVICES PANEL MEETING

Volume II

Tuesday, August 18, 1998

9:00 a.m.

Room 020B
9200 Corporate Boulevard
Rockville, Maryland

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

PARTICIPANTS

Naomi P. Alazraki, M.D., Chairperson
Robert J. Doyle, Executive Secretary

MEMBERS

Judy M. Destouet, M.D
Brian S. Garra, M.D.
Melvin L. Griem, M.D.
Arnold W. Malcolm, M.D.
A. Patricia Romilly-Harper, M.D.
James B. Smathers, Ph.D.

TEMPORARY VOTING MEMBERS

Alicia Y. Toledano, Sc.D.

NON-VOTING MEMBERS

Edward S. Sternick, Ph.D. (industry rep)
Patricia Whelan, M.S. (consumer rep)

GUEST SPEAKERS

Wendie Berg, M.D., Ph.D.
John Gohagan, M.D.

GUESTS

Craig Beam, Ph.D.
Barbara Monsees, M.D.
Edward Sickles, M.D.

FDA

Susan Alpert, M.D., Ph.D.
Robert A. Phillips, Ph.D.
William Sacks, Ph.D., M.D.
Lillian Yin, Ph.D.

C O N T E N T S

Call to Order: Naomi P. Alazraki, M.D. 4

FDA Introductory Remarks: Robert J. Doyle 6

Radiological Health Reengineering Overview:
Joanne Barron 10

Open Public Hearing 17

Charge to the Panel: Naomi P. Alazraki, M.D. 19

Presentation of Digital Mammography Issues

Introduction: Susan Alpert, M.D., Ph.D. 21
Background: William Sacks, Ph.D., M.D. 25
NCI Perspectives: John Gohagan, M.D. 33
Image Analysis: Robert F. Wagner, Ph.D. 36
Statistical Issues: Gregory Campbell, Ph.D. 56
Reader Variability: Wendie Berg, M.D., Ph.D 69

Industry Presentations

Trex Medical Corporation: Richard Bird 86
GE Medical Systems: Edward Hendrick, Ph.D. 101
Schick Technologies: David Schick 114
Fischer Imaging: Morgan Nields 121
Fuji Medical Systems: Robert A. Uzenoff 131

Panel Member Discussion 142

Open Public Hearing 208

Panel Wrap-up 217

1 P R O C E E D I N G S

2 **Call to Order**

3 DR. ALAZRAKI: Good morning and welcome to this
4 second session of this Radiological Devices Panel meeting.
5 This session will be devoted to issues related to digital
6 mammography. I also want to request everyone in attendance
7 to sign in on the attendance sheets which are available at
8 the door.

9 I would like to ask each panel member and special
10 guests at the table to introduce him- or herself and state
11 his or her specialty, position, title, institution and
12 status on the panel. I will start with myself. I am Naomi
13 Alazraki. I am Professor of Radiology at Emory University
14 School of Medicine and Chief of Nuclear Medicine at the V.A.
15 Medical Center in Atlanta.

16 I chair the panel and, as such, I am a voting
17 member but I vote only in the event of a tie.

18 MR. DOYLE: My name is Bob Doyle. I am the
19 executive secretary of the Radiological Devices Panel and a
20 scientific reviewer in the Radiology Branch of the Office of
21 Device Evaluation.

22 DR. SICKLES: My name is Ed Sickles. I am
23 Professor of Radiology at UCSF in San Francisco. I am a
24 mammographer and I am here as a guest of the panel.

25 DR. DESTOUET: I am Judy Destouet, Chief of

1 Mammography for Advanced Radiology in Baltimore, Maryland.
2 I am a voting member of the panel.
3 DR. ROMILLY-HARPER: I am Pat Romilly-Harper,
4 Medical Director, Indianapolis Breast Center. I am a voting
5 member of the panel.
6 DR. BEAM: My name is Craig Beam. I am an
7 Associate Professor at the Medical College of Wisconsin. My
8 area is in statistics. I am a guest on the panel.
9 DR. BERG: Wendie Berg. I am Associate Professor
10 of Radiology and Director of Breast Imaging at University of
11 Maryland. I am a guest speaker on the panel.
12 DR. YIN: Lillian Yin, FDA, Center for Devices and
13 Radiological Health. I am the Director of the Division of
14 Reproductive, Abdominal, Ear, Nose and Throat, and
15 Radiological Devices.
16 DR. SACKS: I am Bill Sacks. I am a radiologist
17 with the Office of Device Evaluation in the Radiological
18 Branch.
19 DR. MONSEES: I am Barbara Monsees. I am
20 Professor of Radiology at Washington University and Chief of
21 Breast Imaging. I am here as a guest.
22 DR. SMATHERS: Jim Smathers, Professor of
23 Radiation Oncology, UCLA. I am a voting member of the
24 panel.
25 MS. WHELAN: I am Patti Whelan. I am a clinical

1 social worker at St. Vincent's Hospital in New York City. I
2 am the consumer representative to the panel and a non-voting
3 member.
4 DR. STERNICK: Ned Sternick, Vice President of
5 Clinical Affairs at NOMOS Corporation outside of Pittsburgh.
6 I am the industry representative to the panel and a non-
7 voting member.
8 DR. GRIEM: I am Melvin Griem, Emeritus Professor,
9 University of Chicago, a radiologist and recently President
10 of Great Lakes Nanotechnology, a microscopic imaging firm.
11 I am a voting member.
12 DR. TOLEDANO: I am Alicia Toledano, Assistant
13 Professor of Biostatistics at the University of Chicago. My
14 specialty is evaluating diagnostic technologies. I am a
15 consultant to the panel with voting status.
16 DR. MALCOLM: I am Arnold Malcolm, a radiation
17 oncologist, Medical Director of Provident St. Joseph Medical
18 Center, Burbank, California. Voting member.
19 **FDA Introductory Remarks**
20 MR. DOYLE: The following announcement addresses
21 conflict of interest issues associated with this meeting and
22 is made a part of the record to preclude even the appearance
23 of any impropriety.
24 To determine if any conflict existed, the agency
25 reviewed the submitted agenda and all financial-interest

1 reports by the committee participants. The conflict of
2 interest statutes prohibit special government employees from
3 participating in matters that could affect their or their
4 employers' financial interest.
5 However, under the final rule, interpretation,
6 exemptions and waiver guidance concerning 18 USC 208 Title
7 5, CFR Part 2640, a special government employee may
8 participate in any particular matter of general
9 applicability where the disqualifying financial interests
10 arise from his non-federal employment.
11 Since the agenda item for this session involves a
12 particular matter of general applicability, the agency has
13 determined that Dr. Barbara Monsees may participate fully in
14 today's deliberations.
15 The agency has also determined that participation
16 of certain members and consultants, the need for whose
17 services outweighs the potential conflict of interest
18 involved, is in the best interest of the government.
19 A waiver has been granted for Dr. Ada Romilly-
20 Harper for her interest in a firm that could potentially be
21 affected by the panel's deliberations. This waiver allows
22 her to participate fully in all matters before the panel
23 today. Copies of this waiver may be obtained from the
24 agency's Freedom of Information Office, Room 12A-15, of the
25 Parklawn Building.

1 We would also like to note for the record that the
2 agency took into consideration other matters regarding Drs.
3 Naomi Alazraki, Wendie Berg, Edward Sickles, and Brian
4 Garra. Each of these panelists reported interest in firms
5 at issue but in matters not related to the agenda for
6 today's session. Therefore, the agency has determined that
7 they may participate fully in all discussions.
8 In the event that the discussions involve any
9 other products or firms not already on the agenda for which
10 an FDA participant has a financial interest, the
11 participants should excuse him- or herself from such
12 involvement and the exclusion will be noted for the record.
13 With respect to all other participants, we ask, in
14 the interest of fairness, that all persons making statements
15 or presentations disclose any current or previous financial
16 involvement with any firm whose products they may wish to
17 comment upon.
18 If anyone has anything to discuss concerning these
19 matters, please advise me now and we can leave the room to
20 discuss them.
21 No one seems to have any.
22 FDA seeks communication with industry and the
23 clinical community in a number of different ways. First,
24 FDA welcomes and encourages premeetings with sponsors prior
25 to all IDE and PMA submissions. This affords the sponsor an

1 opportunity to discuss issues that could impact the review
2 process.
3 Second, the FDA communicates through the use of
4 guidance documents. Towards this end, FDA develops two
5 types of guidance documents for manufacturers to follow when
6 submitting a premarket application. One type is simply a
7 summary of the information that has historically been
8 requested on devices that are well-understood in order to
9 determine substantial equivalence.
10 The second type of guidance document is one that
11 develops as we learn about new technology. FDA welcomes and
12 encourages the panel and industry to provide comments
13 concerning our guidance documents.
14 Finally, I would like to remind you of the
15 meetings of the Radiological Devices Panel tentatively
16 scheduled for the remainder of this year and next. They are
17 November 16, 1998, and then, next year, on February 8,
18 May 17, August 16 and November 8.
19 You may wish to pencil in these dates on your
20 calendars but please recognize that these dates are
21 tentative at this time.
22 DR. ALAZRAKI: I would like to ask Dr. Brian Garra
23 to introduce himself as a panel member.
24 DR. GARRA: My name is Brian Garra. I am
25 Professor and Vice Chairman of Radiology at the University

1 of Vermont Medical Center and voting panel member.

2 DR. ALAZRAKI: Thank you.

3 I would like to introduce Ms. Joanne Barron of the

4 Office of Compliance who will give a brief overview of

5 radiological health reengineering.

6 **Radiological Health Reengineering Overview**

7 MS. BARRON: Good morning.

8 [Slide.]

9 Our team has been in existence since February. I

10 am sure you are familiar with the reengineering efforts that

11 have been going on in the center. This one is a little bit

12 different from the others because, with this one, we are

13 tackling radiological health issues instead of the medical-

14 device issues.

15 [Slide.]

16 As you are aware, we have a number of issues that

17 have initiated the whole reengineering issue. We have

18 shrinking resources and the Radiological Health Program we

19 have had approximately a 70 percent drop in FTEs over the

20 past ten or twelve years.

21 That has left us with a lot of issues of trying to

22 figure out what is the best thing that we can be doing for

23 the resources that we have left. We have a lot to cover and

24 so we are trying to figure out how to do that.

25 There is a perception, at least, if not a reality,

1 of a lack of coordination across the center of all of the
2 efforts because there has been so much reduction of FTEs.
3 We have new products, new technologies, we have difficulty
4 keeping up with. There is a difficulty in trying to balance
5 the radiological-health work compared to all the device
6 issues and trying to figure out how to deal with that issue.
7 We have problems with things that we have resolved
8 and the people have left who did it and there was no
9 documentation. We are trying to figure out how to deal with
10 that issue as well.
11 Of course, the last two questions up here are just
12 simply reengineering; should someone else be doing it and
13 what happens if we stop, which are real issues for us in
14 this case.
15 [Slide.]
16 We found an old definition in an old department
17 book that defined radiological health and went back and
18 looked at it, realized that it was actually one of our core
19 values. We have been told that that is something that you
20 don't reengineer. So we are trying to figure out how to
21 work around this definition as well of both protecting and
22 promoting.
23 [Slide.]
24 This slide is our best effort to try to show the
25 scope of what is involved here. We have three laws that we

1 are trying to enforce. We have mandates specifically on FDA
2 on each one of them that I will go over in just a moment.
3 But we have to cover all those products.
4 As you can see in the top one, under the Radiation
5 Control Act, there are a lot of non-medical products that we
6 have responsibilities for. It is not just the products. It
7 is all the different kinds of radiations.
8 It is all of those up there; ionizing, non-
9 ionizing radiations, acoustic radiations are all included.
10 A lot of those, we don't have a whole lot of information on.
11 The products do not have a registration requirement. There
12 is not a mandatory inspection requirement so we don't even
13 always know who the manufacturers are or what all the
14 products are that we need to be involved with.
15 These down here that have overlap between the
16 device law and the Radiation Control Act, we think we are
17 covering fairly well but there are a lot of issues there
18 that are strictly radiological-health issues that we are not
19 spending a lot of time on.
20 The device issues, I think, are being covered
21 fairly well. These down here that are only device issues
22 are the ones that are nuclear materials that we don't have
23 responsibility under Radiation Control because they are not
24 electronic products.
25 [Slide.]

1 What I want to show you is a comparison between
2 the three laws because this kind of gives the overview of
3 where it is I think we are going to have to do the
4 reengineering. In the device law, we have very clear
5 mandates of what we are supposed to do for premarket. We
6 are supposed to do a clearance of some of the products,
7 approvals of others; very clear. We have special control
8 options if we want to use them and we have requirements to
9 do inspections.
10 Mammography acts, I am not going to cover at the
11 moment. The Radiation Control Act, if you will notice,
12 there is no mandate for premarket approval or clearance.
13 There are some authorities on the premarket side but I have
14 not listed those and have put them down here on the bottom.
15 We have requirements for standards to both develop
16 and enforce. We have not done much of that in the last,
17 probably, ten years or so. We are starting to do a little
18 bit more with the TEPRSSC committee working--that is the
19 Technical Electronic Product Radiation Safety Standards
20 Committee, very similar to this panel that is mandated under
21 this law.
22 We have a requirement to monitor industry testing.
23 The two authorities given to us are testing products and
24 inspections. We also have the authority to require reports
25 from manufacturers but they are basically a notification of

1 how they met the standard or how they are meeting any type
2 of safety issues.
3 But, it basically leaves it wide open. So we can
4 do any kind of industry testing, monitoring of the industry
5 testing, that is appropriate for a particular industry.
6 That leaves it wide open in terms of reengineering because
7 that gives us a lot of latitude.
8 There are four things that are very distinct;
9 testing products--that is, FDA testing them; doing research;
10 evaluating exposure and emissions from products; and
11 training information. Those are not in any of the other
12 laws. But we haven't spent a whole lot of time there. The
13 question is, do we need to.
14 What we want to particularly point out is the
15 intent of that law. It is the control of exposure. It is
16 not control of products. So that also means that we need to
17 take a more global look at what we are trying to do with the
18 reengineering.
19 [Slide.]
20 One of the issues here is population exposure, not
21 just an individual but the entire population that is being
22 exposed to radiation needs to be considered. When we
23 evaluate a medical device, for example, we may not be
24 looking at that issue. We may be looking at one
25 manufacturer's one specific model of a product and not

1 looking at the whole issue across the board.

2 This gives us that ability. If we looked at what

3 people were being exposed to a number of years ago from

4 medical diagnostic X-ray, we could say something on that

5 order for the numbers of exams that were being conducted in

6 this country. Now, there are many more exams being done.

7 They are being done for longer periods of time. They are

8 being done with much higher exposures in some cases.

9 We have got non-medical products, like the

10 television VDTs that used to be just a television receiver

11 in your home. You sat six feet away and you watched for

12 four hours at day. Now, we sit at VDTs 18 inches away, ten

13 hours a day, and we put them on every piece of medical and

14 non-medical equipment you can imagine.

15 That has something rather unique that we also have

16 to consider and that is there is a radiation there that is

17 intended. It has to be there for the purpose. That is the

18 visible radiation. But there is also ultraviolet, X-ray,

19 and so forth. When those products emit all those different

20 kinds of radiations, that is what we have to take into

21 account.

22 We have also got new products we don't know

23 anything about. We have practically no information on the

24 bioeffects for cellular phones and all those other new

25 technologies that keep coming up.

1 [Slide.]

2 So what we are doing is we are taking a real quick
3 look at the products that we are trying to give to our
4 customers and we are going to be trying to evaluate and
5 reengineer where we put the effort that we are doing. These
6 are only some of the examples we have come up with. We are
7 still trying to work on an analysis of the program and we
8 will probably have within two months some kind of a proposed
9 direction.

10 At that point, we are going to want a lot of
11 feedback. So we are asking this panel would you participate
12 with us in that case.

13 [Slide.]

14 Questions. Who is currently doing the work? What
15 would happen if we don't do it? What is ineffective? Those
16 are the kinds of questions that we are looking at. Those
17 are the kinds of feedback that we would like to get from
18 you.

19 [Slide.]

20 We have an e-mail address for REGO, which is
21 reinventing government, at cdrh.fda.gov. At any time during
22 the process, you can send us an e-mail and we will take it
23 into consideration. We will take it to our group. You can
24 also e-mail to me directly or you can mail to me directly.

25 [Slide.]

1 You can put that up. This is the team. While he
2 is doing that, what I have here is a questionnaire that we
3 have set out to the TEPRSSC committee and we have gotten
4 many of them to respond. We have also sent it out to the
5 states through the Conference of Radiation Control Program
6 directors and we have gotten a few back from them.
7 What I would like to do is leave with Bob Doyle
8 these questionnaires and people on the panel, in particular,
9 we would be interested if you would be willing, in the next
10 two weeks, to fill this out and send it back in to us.
11 The first is just simply ranking what you believe
12 are the most important customer needs and the second is
13 taking a look at those outputs that I put up there. Do you
14 think we are looking at the ones that are most effective and
15 the ones that are least effective, and then a few questions
16 of where you would give us some suggestions.
17 That's all I have. Thank you.
18 DR. ALAZRAKI: Thank you, Ms. Barron.
19 Do any of the panel members have any questions for
20 Ms. Barron?
21 If not, we will now proceed to the first of the
22 two half-hour open public hearing sessions.
23 **Open Public Hearing**
24 For this meeting, we have two of these open public
25 hearing sessions, the first of which occurs now, the second

1 of which occurs following the panel discussion and before
2 the panel wrap-up.
3 At these times, public attendees are given an
4 opportunity to address the panel, to present data or views
5 relevant to the panel's activities. If there are any
6 individuals wishing to address the panel, please raise your
7 hands and identify yourselves now.
8 [No response.]
9 I would like to remind public observers at this
10 meeting that, while this portion of the meeting is open to
11 public observation, public attendees may not participate
12 except at the specific request of the chair.
13 Since there are no individuals who have indicated
14 that they want to address the panel at this first open
15 public hearing session, we will go on. I do have some
16 instructions for anyone who would be addressing the panel.
17 If you are to address the panel, please come forward to the
18 microphone and speak clearly as the transcriptionist is
19 dependent on that for providing an accurate transcription of
20 the proceedings at the meeting.
21 If you have a hard copy of your talk available,
22 please provide it to the executive secretary for use by the
23 transcriptionist to help provide an accurate record of the
24 proceedings.
25 We also request that all persons making statements

1 either during open public hearings or the open committee
2 discussion portions of the meeting disclose whether they
3 have financial interests in any medical-device company.
4 Before making your presentation to the panel, in addition to
5 stating your name and affiliation, please state the nature
6 of your financial interest in the company. Of course, no
7 statement is necessary from employees of that company.
8 Definitions of financial interests in the sponsor
9 company may include compensation for time and services of
10 clinical investigators, their assistants and staff, in
11 conducting the study and in appearing at the panel meeting
12 on behalf of the applicant. Direct stake in the product
13 under review--for example, inventor of the product, patent
14 holder, owner of shares of stocks, et cetera--owner, or
15 part-owner of a company.
16 Since there were no individuals indicating that
17 they wanted to speak during the open public portion of the
18 meeting, I am going to close that portion of the meeting and
19 go on.
20 **Charge to the Panel**
21 DR. ALAZRAKI: A number of manufacturers have
22 developed digital mammography devices and are about to
23 evaluate them clinically, and more are expected in the
24 future. In June of 1996, the FDA published a guidance
25 document entitled Information for Manufacturers Seeking

1 Marketing Clearance of Digital Mammography Systems.

2 Since that time, significant developments have

3 occurred in the technology and understanding of how these

4 devices work and are used. The FDA is now asking this panel

5 and the experts invited for today's meeting to provide input

6 on improvements that can be made to this original guidance

7 to reflect the knowledge gained since it was originally

8 published. In particular, the agency is interested in input

9 in clinical-study design parameters.

10 We will now proceed with presentations for this

11 agenda item. It will be introduced by Dr. Susan Alpert who

12 is the Director of the Office of Device Evaluation.

13 Following Dr. Alpert's presentation, Dr. William Sacks of

14 the Office of Device Evaluation will provide overall

15 background information. Then Dr. John Gohagan of the

16 National Cancer Institute will present the NCI perspectives

17 followed by Dr. Robert Wagner of the FDA Office of Science

18 and Technology who will address the fundamentals of image

19 analysis.

20 Dr. Gregory Campbell, Director of the Center's

21 Division of Biostatistics will then cover statistical issues

22 and, in conclusion, Dr. Wendie Berg, Director of Breast

23 Imaging at University of Maryland Hospital will address

24 inter- and intrareader variabilities.

25 Dr. Alpert?

1 **PRESENTATION OF DIGITAL MAMMOGRAPHY ISSUES**

2 **Introduction**

3 DR. ALPERT: Thank you, Dr. Alazraki.

4 Good morning. First, let me thank all of you on
5 the panel and all of the speakers and guests for coming
6 together today to assist us in addressing issues related to
7 introduction of digital technology to mammography. If you
8 are having deja vu, we all are.

9 Several years ago, we sat in this very room, on
10 the other side of the room but in the very room, and
11 discussed the very same topic. At that point, we had only
12 theory to work with and no actual clinical data upon which
13 to build, no actual experience to guide us in our
14 deliberations.

15 Today, we return to the topic a wiser group of
16 scientists. But before we begin to discuss the new data,
17 the issues that have arisen and the work that resulted from
18 our last meeting, let me just take a few moments to remind
19 all of us why this is of such importance that we have
20 returned after almost three years to a public forum for this
21 discussion.

22 Breast cancer remains the most commonly occurring
23 cancer in women in the United States. One out of eight
24 women will have such a diagnosis in their lifetime. All of
25 us are related to women with breast cancer or know women

1 with breast cancer.

2 Also, let's not forget that about 1 percent of

3 people with breast cancer are men. On the other side of

4 this statistic is the fact that life-saving potential of

5 early detection of breast cancer from screening mammography

6 has been continually improving.

7 Earlier diagnosis of more cancers is saving lives.

8 None of us wants to impact this statistic adversely. We

9 would be doing the women of this country a great disservice

10 should that occur. New technology that does not support the

11 continued gains we have seen in screening mammography or

12 provide a real benefit as just newer but may not be better

13 or even as good as the old.

14 Our job is to assure ourselves and then the women

15 who rely on screening mammography that the technology used

16 is the right technology.

17 That assessment has at least two goals; to assure

18 that more cancers are not missed than with the current

19 technology and that we don't simply increase the number of

20 women without cancer that are referred for more invasive

21 procedures.

22 The ease of handling information acquired

23 digitally is not, in itself, sufficient to support marketing

24 unless we are assured of the two clinical outcomes I just

25 described; no loss in sensitivity, no clinically important

1 loss in specificity.

2 Complicating our discussion is the fact that the
3 technology and the devices being currently developed do not
4 yet appear to take full advantage of all of the potential of
5 digital technology. We are told by the industry that the
6 time when digital mammographic images will be read on-screen
7 where they can be manipulated for contrast and highlighting
8 is not yet upon us.

9 We are at an earlier stage but there is a great
10 desire to take advantage of what can be done today with
11 digital acquisition of images.

12 Other complicating factors have also come to
13 light. There is a well-known inter- and intrareader
14 variability due, of course, to the fact that the human eye
15 and the human brain are some of the discerning pieces in
16 making these diagnoses and, as we all are aware, machines
17 are much better at reproducibility than we humans.

18 But the radiologist remains a critical part of
19 this process. The judgment of such professionals
20 contributes greatly to the success of mammography. Teasing
21 out the contributions of the device from the reader has
22 turned out to be very difficult.

23 Another complication in studying new technologies
24 in this particular arena is that, although the number of
25 women diagnosed with cancer is increasing, the actual number

1 of positives among the total number of women screened
2 remains low. This is good for the women and difficult for
3 the studies.
4 Lastly, at least for my introductory remarks, is
5 the complicated analyses that result from these realities.
6 The challenge of identifying the appropriate number of
7 locations, films, readers and the right statistical analyses
8 to provide the assurances that we need to move forward with
9 this technology has proven to be a greater problem than we
10 anticipated when we last met and we simply didn't get it
11 right the first time.
12 You will hear this morning from a knowledgeable
13 group of individuals regarding all of these issues. After
14 listening to the presentations and digesting them over
15 lunch, we will sit together to discuss ways to move forward
16 in this high-impact area.
17 Again, let me thank you all for joining us. Let
18 me remind you, again, that this is a discussion that we hope
19 will be open and complete and candid on the part of all the
20 participants. There are a lot of issues here. This is not
21 a voting situation for one way to do things but a desire on
22 our part to have a full and open discussion with all of the
23 experts in the field to understand the technology and to
24 understand the best way for us, the industry, the clinical
25 community and the regulators, to move forward in providing

1 advances in mammography and, in this particular setting, for
2 the introduction of digital technology to the mammography
3 suite.
4 Again, thank you for participating and I look
5 forward to participating with you this afternoon in the
6 discussion.
7 DR. ALAZRAKI: Dr. Sacks?
8 **Background**
9 DR. SACKS: Good morning, again.
10 [Slide.]
11 Yesterday, I took you through some recent history
12 going back as far as last November's panel meeting.
13 [Slide.]
14 Today, we are going to search the archeological
15 records and go all the way back to March of 1995 when the
16 panel first met to discuss this issue of digital
17 mammography.
18 [Slide.]
19 At that time, there were basically two
20 alternatives considered by the panel as ways of designing
21 clinical trials with which to evaluate the contribution of
22 digital-mammography systems to screening mammography. The
23 first was a full screening trial and the second was an
24 agreement study.
25 This shows certain features and how they differ.

1 First of all, a full screening trial would require the
2 double exposure--that is, once with an analog or film-screen
3 mammography and secondly with a digital mammogram--numbering
4 between 10,000 and 20,000 women.
5 Full screening trials would require a gold
6 standard which we would look to biopsy and/or follow up of a
7 year or so to prove the absence of cancer. We would have
8 used ROC analysis and the sensitivity and specificity would
9 be the two relevant clinical parameters.
10 The other alternative that was discussed at that
11 time was an agreement study which would involve the double
12 exposure of approximately 500 women, give or take, namely in
13 the hundreds rather than in the tens of thousands.
14 Here, the clinically relevant parameters would be
15 the conditional sensitivity and conditional specificity by
16 which I mean the sensitivity or the probability that the
17 digital mammogram would be positive given that the analog
18 mammogram on that woman was positive. Similarly, the
19 conditional specificity would be the probability that the
20 digital mammogram would be negative given that the analog
21 mammogram was negative.
22 These would have to be set at some minimum
23 requirements. By the way, we will hear from Drs. Gohagan
24 and Wagner shortly on the issues of the full screening
25 trial.

1 [Slide.]

2 As a result of that panel discussion, the FDA, in
3 June of 1996, issued a guidance to industry for agreement
4 studies. Guidances that are issues by the FDA are not
5 mandatory. They are suggestions.

6 [Slide.]

7 This guidance set the conditional sensitivity
8 criterion at 95 percent and the conditional specificity
9 criterion at 98 percent. Furthermore, this guidance defined
10 the negative mammograms--that is, a negative analog or a
11 negative digital--as falling into the BIRADS categories 1 or
12 2 and defined the positives as BIRADS 3, 4 or 5.

13 One of the questions that we have posed in the set
14 of questions that were sent to the panel in advance
15 concerned the question of the assignment of the BIRADS 3
16 category, whether it should be included with the positives
17 or included with the negatives and that should be part of
18 the discussion later today.

19 [Slide.]

20 Since June of 1996, when the guidance was issued,
21 the FDA has not only received some data but we have had
22 numerous discussions with a number of companies who seek to
23 develop digital-mammography systems. All of this plus
24 searches of the radiological and diagnostic imaging
25 literature have revealed certain potential flaws in the

1 guidance that was issued in 1996.

2 [Slide.]

3 In fact, as a result of these discussions and some
4 of this data, a third alternative has been promoted and that
5 is study called a non-inferiority study in which the goal
6 would be to demonstrate not the agreement or equivalence of
7 digital to analog but that digital was not inferior to
8 analog by more than a certain specified small quantity; that
9 is, we would compare the sensitivity of the analog and the
10 sensitivity of the digital and require that that difference
11 not be bigger than some pre-set delta, and similarly for the
12 specificities.

13 This, too, of course, would require the gold
14 standard because you can't measure sensitivity and
15 specificity without one. The conditional sensitivities and
16 specificities in an agreement study would not require a gold
17 standard because there, in essence, we were taking the
18 analog mammogram as the standard to which we were comparing
19 the digital.

20 Such a study, a non-inferiority study, would
21 double expose women numbering somewhere in the hundreds to
22 the thousands, short, generally, of the tens of thousands
23 that a full screening trial would require. This number
24 would be highly dependent not only on the choice of the
25 deltas but it would be dependent on just exactly what the

1 difference between digital and analog turned out to be in
2 terms of their sensitivities and specificities.
3 [Slide.]
4 The problems that came to light can be lumped into
5 three main groups. First is the variability between and
6 within readers. Dr. Wendie Berg will be talking about this
7 later on. Because it is difficult to get analog in a trial
8 to agree with analog, let alone with digital, this tends to
9 offer an obstacle to agreement studies, although there may
10 be suggestions for ways to solve this problem and that,
11 certainly, is among the questions that we are asking the
12 panel.
13 A second problem that comes to light is that the
14 difference, as Dr. Alpert was saying, between hard copy and
15 soft copy. One the one hand, there is the acquisition of
16 the digital image and, on the other hand, there is the
17 display of the digital image. Display can be either on film
18 or it can be on a monitor where the radiologist can actually
19 adjust the windows and levels.
20 This is very similar, of course, to CT where that
21 can be done either way. Trials that are based on hard copy
22 such as the ones that we have been discussing with most of
23 the companies, do not, in fact, test the full potential of
24 the device which really involves the ability to go to the
25 monitor and change windows and levels.

1 A third problem is that with non-inferiority
2 studies, there is a selection bias which may be able to be
3 overcome mathematically but that consists of the following:
4 the enrollment of the women in the trial is based on the
5 results of their analog mammograms. Given that the
6 prevalence of breast cancer is so low in the population,
7 less than 1 percent, the enrollment in a trial is usually
8 based on taking all of the analog positives and exposing,
9 again, all of those women who, after all, sign a consent, an
10 informed consent, to a digital mammogram.
11 But of the analog negatives which will far
12 outnumber the analog positives, only a random selection of
13 those would be chosen to double expose and give a digital
14 mammogram to.
15 Because that selection is based on the analog
16 negatives as opposed to the digital negatives, this
17 introduces a bias which, in fact, in the first instance, is
18 against digital sensitivity, as it turns out--I am not going
19 to go through the mathematics--and in favor of digital
20 specificity.
21 Those issues, by the way, the statistical issues,
22 will be discussed later by Drs. Campbell and Bushar.
23 [Slide.]
24 I want to discuss why it is that we need clinical
25 trials in the first place. First of all, digital

1 mammography is not like CT was to old plain X-rays or MRI
2 was to either CT or X-rays or ultrasound and so on, which
3 were all completely new technologies. MRI and ultrasound
4 even more so because at least CT involves radiation
5 interaction with human tissue just as plain films or
6 fluoroscopy do.
7 Digital mammography is, after all, the same
8 exposure of human tissue to ionizing radiation. In that
9 sense, it is identical to analog mammography. It is the
10 acquisition and the display after this interaction, the
11 physiological interaction, with the patient takes place that
12 digital mammography differs from analog.
13 So this is only, therefore, an incremental
14 technological change from analog mammography and it uses the
15 same, as I said, physiological interaction of radiation with
16 the patient.
17 [Slide.]
18 Secondly, in the laboratory, one can, and the
19 Office of Science and Technology here at FDA has, measured
20 changes or the differences in the dynamic range of digital
21 and analog and in the signal-to-noise ratio and in
22 resolution. Dr. Wagner will be talking about these issues.
23 [Slide.]
24 So even if it turns out that digital has superior
25 dynamic range and has superior signal-to-noise ratio over

1 most or all of that range--

2 [Slide.]

3 --it is still the case that conceivably,

4 sensitivity could be lowered when you get to the

5 radiologist's perceiving and speaking into a microphone to

6 dictate a report. For example, the better viewing of normal

7 surrounding tissue could, conceivably, offer camouflage to

8 cancer.

9 [Slide.]

10 Similarly, specificity, conceivably, could be

11 lowered despite an improvement in dynamic range and signal-

12 to-noise ratio, for example, by improved visibility of

13 microcalcifications or of margin irregularities in masses.

14 [Slide.]

15 Therefore, the real question for the panel, the

16 core of the question that we are asking to be discussed

17 today, is what level of clinical evidence is necessary to

18 demonstrate the effectiveness of digital-mammography

19 systems. Safety is, of course, not an issue here because it

20 is identical to that of analog mammography. Of course,

21 there is a range of exposures, but it is in the same range.

22 When we ask what level of clinical evidence, we

23 mean agreement studies, non-inferiority studies, full

24 screening studies, perhaps evaluation of images from

25 selected subgroups such as the extremely dense breast and so

1 on. Perhaps, there are others which the panel may wish to
2 suggest.
3 [Slide.]
4 Finally, how can those trials be designed to
5 obviate the problems that were described earlier.
6 Thank you.
7 DR. ALAZRAKI: Thank you, Dr. Sacks.
8 Dr. Gohagan?
9 **NCI Perspectives**
10 DR. GOHAGAN: Thank you very much.
11 [Slide.]
12 I think everybody here, at least the people whose
13 faces I recognize, won't see anything new in anything I have
14 to say today. We all know that the NCI and the FDA have
15 somewhat different perspectives on the new technologies and
16 existing technologies and how they are to be evaluated.
17 But what I have to say is not what I would say is
18 "the" NCI perspective. I am just going to give an overview
19 of what one might think of as "an" NCI perspective which is
20 really the point of view of early detection.
21 [Slide.]
22 The mission, basically, is to reduce cancer-
23 specific mortality and morbidity and to reduce cancer
24 incidence as one of the things, the last not being early
25 detection but one of the things which is important to NCI.

1 [Slide.]

2 Of course, the approaches go from basic research
3 down to surveillance. We happen to be early detection in
4 the prevention arm. Early detection is called secondary
5 prevention.

6 [Slide.]

7 The things that we do are technology development,
8 to some degree, not necessarily all new technology
9 development but modifications of technologies, perhaps,
10 assessment and transfer of those technologies through
11 publication and principles of operation for early detection.

12 [Slide.]

13 The levels of evidence that we consider are in two
14 groups, basically. I think the one that we stand by from
15 the longest history probably is the PDQ levels of evidence
16 for judging technologies. I think that we have got them
17 ranked there in an order which most people are familiar
18 with; randomized clinical trials for mortality reduction are
19 considered to be the definitive approach.

20 Then, at the bottom of the list, ecologic studies
21 where we are looking for patterns. I know that with the
22 technology that is being discussed today, there is not going
23 to be any discussion about mortality endpoints. I think
24 there is good reason for not discussing that. We already
25 know what goes on with analog mammography. The question is

1 how to compare digital mammography to analog mammography and
2 to avoid such huge trials.
3 The six-tier effects model which has been around
4 for some time, some of the members of FDA and NCI panel have
5 been looking at as a way of characterizing kind of the work
6 that goes on in both institutions and where we fit together.
7 I think the first two items there are primarily
8 ones which we have agreed are FDA levels of activity more
9 than they are NCI although there is some overlap, at least
10 with sensitivity and specificity. But the lower group on
11 the right is primarily where the NCI fits into things.
12 So, as you might guess, we have figured out that
13 the NCI worries more about the end uses of the technologies,
14 the impact on mortality. We wouldn't have technologies to
15 look at if the FDA weren't approving technologies that can
16 be used in the clinic.
17 [Slide.]
18 Then, finally, a couple of cases in point, I
19 guess. Some years ago, I was here for the hearings on PSA.
20 It was my first time here, having come from a university. I
21 could see then that there was an entirely different point of
22 view that was being taken at the FDA rather than the NCI.
23 We could approve a technology for a specific
24 purpose but it might not have anything to do with the
25 mortality impact at that point of proof. Rituximab is

1 another example of that. I think it has been approved as an
2 immunotherapy for non-Hodgkins lymphoma and we have clinical
3 trials going to evaluate it to find out what the survival
4 impact is.
5 Digital mammography is the one you are talking
6 about today. We have already heard from Dr. Sacks the
7 various approaches that could be taken to evaluating it.
8 So there really is an interesting small overlap, I
9 think, that we have at the NCI with what goes on at the FDA
10 at this level. On the other hand, in the long run, the
11 technologies that get approved by the FDA, we end up looking
12 at sometimes for long periods of time trying to determine
13 whether they have the effect that we are looking for.
14 Thank you.
15 DR. ALAZRAKI: Thank you, Dr. Gohagan.
16 Dr. Robert Wagner?
17 **Image Analysis**
18 DR. WAGNER: Good morning. Members of the panel,
19 thank you very much for this opportunity to speak with you
20 this morning.
21 [Slide.]
22 I would like to give an overview of a model of the
23 sources of variability that occur in the assessment of
24 diagnostic-imaging modalities. If you were to go up to the
25 person on the street in your neighborhood and ask them to

1 list for you some of the sources of variability, I am sure
2 the first thing your neighbor would say to you is, "Well,
3 patients vary from patient to patient."
4 If you went to someone with a little bit more
5 experience, perhaps who had been reading the newspapers,
6 they would say, "Well, radiologists are different one from
7 another." But if an engineer or physicist was overhearing
8 this conversation, they would say, "Well, what about the
9 hardware? Are there differences and does that contribute to
10 the variability?"
11 What I would like to do today is to present a
12 model that has been evolving in several communities, the
13 academic radiology community, the Society for Medical
14 Decisionmaking, and several of the image-sciences
15 communities to put these things into context and see what
16 these words really mean.
17 [Slide.]
18 Here is the background of people who have studied
19 this field. This is about the first set of datapoints that
20 you encounter. You find out that in the '50's, there were
21 several large studies of chest radiography and the names
22 Yerushalmy and Garland come up in connection with those
23 studies.
24 It was reported that the variation seen in readers
25 of chest films was of the "disturbing magnitude."

1 Physicians disagreed with the diagnosis of a colleague on an
2 average of one out of three times. Physicians disagreed
3 with their own previous diagnoses on average of one out of
4 five times.
5 [Slide.]
6 Those of you who are familiar with the literature
7 and what happened in the next ten years will recall that Lee
8 Lusted, whom we lost a few years ago, in the late 1960's,
9 helped us to understand what was going on there by bringing
10 the concept of the receiver-operating-characteristic curve
11 over from several fields of communication science and
12 psychology and bringing it into medical imaging
13 The ROC curve is the plot of the true-positive
14 fraction versus the false-positive fraction but there is a
15 variable that you don't see in the ROC curve, and it is
16 called the latent variable, referred to as the "reader
17 mindset."
18 The understanding in this paradigm is that, as you
19 walk up the ROC curve, this corresponds to the image reader
20 taking a more and more aggressive stance toward calling the
21 presence of disease. The mindset is determined by the
22 reader's sense of the disease prevalence and the assessment
23 of the relative costs of the two kinds of errors, the false
24 positives and the false negatives.
25 Lusted and others pointed out there are two

1 basically different directions that ROCs face. What is the
2 direction along the ROC curve which they just spoke of a
3 minute ago which corresponds to the movement in the reader's
4 mindset, and the other is a direction in the other
5 direction, if you will, across the ROC curve which
6 corresponds back and forth, which corresponds to differences
7 in the level of technology or the reader's skill.
8 So this is an important distinction to keep in
9 mind and we will return to this in the course of this little
10 summary.
11 [Slide.]
12 You know what they say about not studying history.
13 History has recently repeated itself in mammography. A
14 famous celebrated paper by Joanne Elmore and colleagues in
15 The New England Journal in 1994; they studied ten
16 radiologists, whose name are here, A, B, C and through J.
17 These ten radiologists were studied with their
18 recommendations for biopsy in 150 patients. Here are their
19 recommendations in the patients with biopsy-proven cancer.
20 Here are the recommendations in patients without cancer.
21 You can see that the patients with cancer, the range was
22 from 33 percent to 82 percent were recommended for biopsy.
23 The patients without cancer, the range was from 3
24 to 20 percent. Again, the authors commented on the
25 remarkable difference in these recommendations.

1 [Slide.]

2 I think many of you know Carl D'Orsi and John
3 Swets. They came back to the New England Journal with a
4 letter to the editor. They plotted the recommendations of
5 Elmore and company in ROC space. They noted that the data
6 of these ten radiologists was consistent with one and the
7 same ROC curve.

8 They noted that the variation above and below the
9 curve was small relative to the variation along the curve,
10 the two directions that I called your attention to earlier.
11 We say, in the technical jargon of this field, that the ROC
12 curve controls for the variability due to mindset, the
13 variability to this swing in the level of aggressiveness in
14 the call.

15 An agreement study does not control for this.
16 Many pairs of points on this curve--many pairs of points--
17 would fail an agreement study. And yet all of these points
18 are consistent with this model ROC curve. They are
19 consistent with an equivalent, a homogenous, level of reader
20 skill and level of technology.

21 [Slide.]

22 I would like to go a little bit deeper. The next
23 two slides will be the most technical and then I will try to
24 give you some relief as we come out of this. The next two
25 slides are an attempt to get inside a model of what happens

1 when readers read and they are faced with the various
2 components of variability that they encounter.
3 I will be showing a lot of Gaussian curves, but
4 nothing that I have to say today in any way depends on an
5 assumption of normality. It is not essential at all. It is
6 just nice and easy to draw such curves.
7 The paradigm consists of a decision axis which is
8 a radiologist's rating of their confidence in the presence
9 of absence of disease. Then we have these two spreads, the
10 spread of the actually normal cases and the spread of the
11 actually abnormal cases.
12 Of course, I pulled them apart. If the curves
13 really were that far apart, this field would be not so
14 interesting and challenging. These two files actually
15 overlap significantly. I have taken the liberty of pulling
16 them apart in the interest of clarity and I have also used a
17 rather arbitrary normalization, again, to make it easy to
18 explain what I would like to say today.
19 These clouds of variability among the normals and
20 the abnormals have at least three pieces. There is always
21 an interior piece which physicists call the physical noise
22 floor. Mother nature provides us with this and it is
23 irreducible.
24 In X-ray imaging, it is from the finite photo
25 flux. In MRI, there is thermal noise. In ultrasonography,

1 there is speckle noise. Mother nature will not take this
2 away. This is always with us. So all the other variability
3 has to sit on top of this. We will talk more about this in
4 just a moment.
5 On top of the physical noise floor sits the
6 patient variability which is, really, the stuff of images.
7 It is the heterogeneity, the variability, of the image, the
8 parenchymal structure that borders the margins, the
9 spiculations, the calcifications, and so on, that is
10 different from patient to patient and even with a region of
11 the breast within a patient.
12 Statisticians refer to this combination of the
13 physical noise floor and the patient variability as the
14 case-sample variability.
15 On top of the case-sample variability sits the
16 component that we will hear more about today, the reader
17 variability. Again, if you go up to this person in your
18 neighborhood and say, "Well, what is reader variability?" it
19 is likely that the person in your neighborhood would say,
20 "Reader variability; isn't that the range of the reader's
21 skills?"
22 Well, there is a little bit more to it than that.
23 I would like to explain one of the more subtle aspects. The
24 reader variability does, in fact, include the range of
25 reader skills but there is a much more subtle effect and I

1 will describe that over the next slide.

2 The ROC paradigm, itself, is not shown here. This

3 is an underlying model. But the mindset is the setting of a

4 threshold to dichotomize this data. This would be less

5 aggressive and this would be more aggressive. And you would

6 move up the ROC curve.

7 So I actually don't show that variation over

8 mindset here. I am showing other underlying causes of

9 variability in reading out and assessing an imaging

10 modality.

11 [Slide.]

12 Perhaps my most subtle overhead, and then we will

13 get the relief I promised you, the way people read out, the

14 way people do experiments in diagnostic imaging is to think

15 in terms of this decision axis, the confidence in the

16 positive disease status, and in the high end of this rating

17 scale corresponding to more confidence in the presence of

18 disease and the other end being more confidence in the

19 absence of a disease.

20 If you are familiar with the literature, you will

21 know that for decades, the ROC analysis used a five- to

22 seven-category scale, these being these discrete categories

23 here. The leaders in this field now are taking us down a

24 direction toward a continuous hundred-point scale, quasi-

25 continuous scale. Then, if you follow the literature in the

1 field of psychology, you will know that we are sort of
2 returning to the roots there because people 30 years or so
3 actually used continuous scales.
4 Today, what people are actually doing is using a
5 little mouse and moving a little marker and have a
6 continuous scale on the computer and reading out a
7 continuous value. You will see why this might be important.
8 For all the importance of the BIRADS document and
9 the BIRADS categories--this little diagram explains to us
10 why some of us think one should not use the BIRADS
11 categories to do ROC analysis. By the way, this is a
12 description of the paradigm underlying ROC analysis.
13 I am not necessarily taking a position one way or
14 the other whether one has to use ROC analysis to do the task
15 at hand. All I saying is that whether you do an agreement
16 study, an ROC analysis or related studies, we don't think
17 you can understand what is going on without getting inside
18 the paradigm at the level that I am taking you to at the
19 moment.
20 If you went to use the BIRADS categories to
21 generate an ROC curve, you get the kind of results I have
22 taken here from a paper by Orel, Sullivan and Dambrow. This
23 is a rather rich study. There were a lot of patients in
24 this study. There were over 500 total and there were over
25 200 biopsy-proven cancers.

1 If you were to make the dichotomy, the cut,
2 between just below the BIRADS 5 so you have 5 versus 1
3 through 4, if you do that, you would get this point in ROC
4 space. If you dichotomize 4 and 5, you would set the
5 biopsy, say, at 1, 2 and 3, and would not, you would get
6 this point.
7 If you dichotomize 1, 2 versus 3, 4, 5, you would
8 get this point.
9 Now, there are at least three points, two of which
10 are glaring and one of which is very subtle on this display
11 here. One, of course, is the well-known, very large size of
12 this bin, this quantized bin, category 4, which is the first
13 thing the glares us in the face which says that if one does
14 use the BIRADS categories or an agreement scheme that uses
15 these categories, you will not sample, perhaps, the most
16 interesting and richest part of ROC space, which is here.
17 If you make the dichotomy 1, 2 versus 3, 4, 5, you
18 will be up here. All ROC curves converge in that corner.
19 So, even if you proved agreement, it is likely that you have
20 not a lot of power because things want to agree up in that
21 corner anyway.
22 It is more subtle what happens if you make the
23 dichotomy 1, 2, 3 versus 4 and 5 and have a point here, or
24 you make a dichotomy here, which you wouldn't be tempted to
25 do, but the same thing up there as here.

1 What happens is this is characterized by very
2 broad categories. What happens has been found out for
3 decades that especially if anybody has done such an
4 experiment, if you try to hold these category boundaries
5 stable, it is difficult to do. The inability to hold the
6 category stable is referred to as category jitter. If fact,
7 it shouldn't be called jitter because jitter makes you think
8 it is just some little movement. It should be like a
9 category rumble. This is not trivial.
10 So what happens is that there is the interaction
11 of this coarse category with this category rumble. So, if
12 you were to calculate with classical sample statistics the
13 uncertainties associated with points like this, you would be
14 ignoring a very important source of variability which is
15 this category size and its interaction with the jitter.
16 [Slide.]
17 Let me give you a little recap of what I said up
18 until this moment. This is from top down, a recap of the
19 model I have been presenting. I would claim that this model
20 has a lot of credence. It is still evolving. I think that
21 it is getting very sophisticated today with some work of
22 people, some of whom are in the room at the moment.
23 The first component of variability was the mindset
24 that I spoke of which is only controllable--as far as I, or
25 people that I know, know--only controllable by the ROC

1 paradigm.

2 The reader variability which you assume to be the
3 range of reader skills but which also includes the
4 interaction of the category size and the boundary jitter.

5 Let's move on to patient variability. As I said,
6 patient variability is the stuff of images, the
7 heterogeneity seen on films. I just stop here for one
8 moment to point out that, at first glance, superficially, it
9 looks like digital and analog mammography would have the
10 same kind, or the same level, of patient variability because
11 when you look at the digital mammogram and an analog
12 mammogram, superficially, they look similar.

13 If one uses hard-copy, high-quality written film
14 as the display. I am not talking about soft-copy display.
15 So if we go through one hoop at a time and we have a high-
16 quality, hard-copy display, then I think the first order of
17 this statement is true. Clearly, there are second-order
18 effects and this concerns just about everybody in the room.

19 The guy on the street has forgotten that patient
20 variability and sample variability includes the physical
21 noise floor, the mother nature contribution that she will
22 not allow us to forget about. So I would like to turn, now,
23 to a discussion for that physicist in the neighborhood who
24 overheard this conversation, the kinds of things that the
25 physics and engineering community are talking about when

1 they try to measure that physical noise floor.

2 I and some members of our group here at the center

3 have been working in this field for over twenty-five years.

4 This field of measurements, of which I will just give you a

5 very short synopsis, has reached a lot of maturity. There

6 is a lot of consensus and I would be prepared to back that

7 up if people would like to ask about the status of consensus

8 of these measurements.

9 The peculiar thing is that it is this very field

10 of the comparison of digital with analog mammography where

11 those measurements have reached their highest points, the

12 most significant level of measurement is in that very field.

13 So let's talk about these measurements.

14 [Slide.]

15 To do this, we use this intuitive introduction

16 which is well-known to any of you who have hung out with any

17 of the physical communities, this is a picture due to a

18 famous inventor of the TB/DB technologies. His name is

19 Albert Rose. He passed away a few years ago. This image is

20 almost fifty years old.

21 It simply shows what you can do with 1,000

22 photons, and 10,000 and 100,000 and some billions and

23 30 million or so photons. Clearly, the more photons in an

24 image, the lower the noise. The more photons, the more

25 information. You have to get into the millions before you

1 know that not only is this a woman but she is looking at a
2 flower of some type of flower of some kind here.
3 [Slide.]
4 Instead of talking about actual photons--actual
5 photons expose cassettes, but the image is not always worth
6 the number of photons it was exposed with. There is a set
7 of measurements that describe what an image is worth. It is
8 very elaborate to get these measurement. At least five very
9 sophisticated sets of measurements have to be done including
10 a signal transfer, a noise transfer, exposure and an X-ray
11 spectroscopy.
12 This is a sophisticated set of measurements to do.
13 But we put out on the X axis the actual exposure quanta and
14 hit the cassette. It is not 10^5 per image, as I showed you
15 with Albert Rose's image, it is 10^5 per mm^2. So there are a
16 lot of photons in X-ray mammography.
17 This is the number of photons mm^2 that hit the
18 cassette. On this axis is the number that the image is
19 worth based on a set of measurements that I don't have time
20 to tell you about that I just tried to sketch.
21 If you had a perfect image detector, the number of
22 noise-equivalent quanta, the worth quanta, would be equal to
23 the number of input quanta. So 10^5 would come out 10^5, 10^6
24 would come out 10^6, and so that is a perfect detector.
25 We don't have these detectors.

1 [Slide.]

2 But we do have an available technology and several
3 technologies that we know from the professional meetings
4 that we all attend. This is data from about five or six
5 years ago taken in our labs by someone in the group here.
6 This is an available digital technology. It came out in the
7 late 1980s. It is still being perfected and this is data
8 from about five years ago, a system five years old.

9 The system tracks a perfect detector in the sense
10 that it is parallel to the perfect detector, but it is short
11 by about a factor of 4 or 5. It still stays in sight of a
12 perfect detector over one decade, two decades, perhaps two-
13 and-a-half decades, of dynamic range.

14 I promise you, if you go up to somebody in the
15 Metro and say, "What is the dynamic range of analog
16 mammography?" they will tell you two decades. That is not
17 correct. That is absolutely wrong because this is based on
18 the fact that they know that you walk in the hospital, you
19 pick up a Macbeth densitometer, it beads out units of 0.01.
20 And they say, at 300 units, it has got a couple hundred
21 units of dynamic range.

22 When you do these signal-to-noise measurements, it
23 is absolutely incorrect. The dynamic range of the tunnel-
24 screen system, if I were to walk up to the screen, which I
25 would if I were adventurous, this is about one decade or

1 less. There is a threshold situation that applies whenever
2 silver-headlight technology is used. And then there is the
3 saturation phenomenon.
4 This limits the dynamic range of a film-screen
5 system to less than one a decade of dynamic range.
6 [Slide.]
7 I have three more overheads. Bear with me; one
8 more technical one and two conclusions so we are winding
9 down. But I think a very interesting point should be made
10 in connection with the comparison between digital and analog
11 mammography.
12 If you wanted to cover that dynamic range of at
13 least two decades or more that the digital technologies can
14 offer us, if you wanted to do that with film screen, you
15 would have to make--this is conservative. This is a
16 conservative estimate. You would have to make at least four
17 exposures with a conventional cassette and you would have to
18 use a very fast film for the low-exposure end.
19 You can do this. You could have a cassette and
20 you could use a fast film and you could cover this part of
21 the dynamic range. You could use a very slow film to cover
22 this end of the dynamic range and you can make several other
23 images. You could thus cover the dynamic range of over two
24 decades.
25 Up here, where a lot of exposure gets through,

1 this corresponds to the skin line. Where not a lot of
2 exposure gets through, this would correspond to the pectoral
3 muscle and the chest wall. Of, if you had a heterogeneous
4 breast, you could walk back and forth and have demands on
5 the detector over this broad dynamic range.
6 With the digital technology, you would get all
7 this with one shot, without any extra exposure. So is this
8 important? It is not important if you have a uniform
9 phantom, you have a uniform test object. Then you may just
10 challenge some factor 8, like the film-screen system does,
11 less than a decade.
12 If you have a pretty homogenous breast, you
13 wouldn't challenge this dynamic range. But if you have a
14 heterogenous breast and some of the situations we just
15 discussed a moment ago, you can very easily challenge the
16 two decades of dynamic range, and then there is a strip in
17 payback by this technology.
18 The technology does not come without some burden
19 which we just allude to in just a moment.
20 [Slide.]
21 So these are my conclusions. Coming top down on
22 the model, historically, it has been found very difficult
23 for diagnostic imaging modalities to demonstrate equivalence
24 even with themselves, which is to say if you have identical
25 ROC curves, this mindset variability is capable of giving

1 you a failure in an agreement study, even for systems with
2 identical with identical ROC curves.
3 That is what I just said. It is not used. It is
4 just not controlled for agreement studies.
5 Beyond this, the ROC paradigm within the actual
6 detection, theoretic paradigm, if you will, there is
7 additional variance which everyone would guess. There is a
8 variability over reader skills, but this is much more subtle
9 variability that I referred to; the interaction of the large
10 category size and category boundaries.
11 So both of these things are going on in reader
12 skills. You might guess that these are related and they are
13 conceptually able to be stated separately. But, in fact,
14 there is some relation between these two things.
15 Thus, the estimation of the resources you require
16 to carry out a successful study with this power that you
17 would like to have, the estimation of resources, then,
18 resources being patients, readers and the replication of
19 readings, this requires a multivariate approach.
20 You can't use univariate statistics to analyze the
21 problem I have just described to you this morning. As a
22 matter of fact, there are experts on this subject in the
23 room at the moment and this field has moved very quickly, in
24 the last five years, especially. There is software becoming
25 available to attack this problem within a multivariate

1 paradigm and this software is at various stages of
2 accommodation and being prepared for user-friendliness and
3 things of that sort.
4 So that is my one little reference to a very
5 sophisticated field in multivariate aspects of this problem.
6 [Slide.]
7 The physical measurements in contrast to all these
8 other sources of variability are able to be made, and more
9 or less crisply, if you will, with accurate and precise
10 measurements. It is very difficult. It is expensive.
11 There are only a few labs in the country that can do this,
12 but we are asking, in the guidance, for physical
13 measurements on imaging systems and the sponsors are
14 providing them.
15 We will ask and hope to use it in labeling.
16 Whatever you think about physical measurements--twenty-five
17 years ago, people said, "Oh, well; we don't understand
18 physical measurements." The field has come a long way
19 today.
20 However you feel about them, everybody wants it.
21 It reminds me about an old joke about Listerine. "Nobody
22 likes it, but everybody needs to use it." That is the way
23 they feel about these measurements. These measurements,
24 everybody wants them.
25 The difference between the two technologies--it

1 may be difficult to assess clinically, especially in general
2 populations. The effects that we have talked about could
3 very easily average out. But you can easily imagine that
4 there will be special populations in which the kinds of
5 differences that we talk about would manifest themselves.
6 So, clearly, in the postmarket stage, it is going
7 to be of great interest to single out these special
8 populations and to look at the possibilities.
9 Digital mammography is not a monolith. There are
10 many technologies. Each technology has its own tradeoff.
11 How will you know if you have different technologies? I
12 just showed you. You will have to make a family of
13 measurements like I showed you.
14 In fact, you have to do a more elaborate set of
15 measurements. I didn't take you down the limiting
16 resolution road, but that is more to the story. Each
17 system, or some of these systems, will take a very small
18 learning resolution hit. Some of these systems also have
19 artifacts from discrete fabrication. There is a discrete
20 number of detectors, a finite number of samples.
21 These effects are almost certainly going to be
22 impossible to see when averaged over a population. But if
23 you have some family member who is going to have a mammogram
24 taken, you may be interested to know whether these artifacts
25 have any effect on individuals.

1 Some of us think that panels of experts should be
2 asked to review sets of films that, perhaps, challenge
3 BIRADS lexicon features of something of that sort, to see
4 how these artifacts play out and whether they are, in fact,
5 negligible.
6 I think you for your time and patience with me.
7 This is an attempt to give a model that has been evolving
8 over several decades in these communities and to show how
9 some of the things we have been talking about today fit
10 within that model.
11 Thank you very much.
12 DR. ALAZRAKI: Thank you, Dr. Wagner.
13 Dr. Gregory Campbell.
14 **Statistical Analysis**
15 DR. CAMPBELL: Good morning.
16 [Slide.]
17 I have done this joint work with my colleague,
18 Harry Bushar. Would I would like to do today is talk about
19 a number of different issues in terms of statistical issues.
20 [Slide.]
21 I will talk about screening trials where the true
22 state of each individual woman, whether they have breast
23 cancer, is known or not; some issues where you are able to
24 resolve the truth in some cases but not all of them; notions
25 of how to do agreement trials using a statistic that is

1 called kappa; and some issues related to multiple readers,
2 issues related to designing the study; and then a new
3 approach, which I will get to at the end.
4 [Slide.]
5 Suppose the state of each individual woman is, in
6 fact, known and that each woman gets both analog and digital
7 mammography with some kind of operational definition of
8 known truth; breast cancer based on either biopsy or long-
9 term follow up.
10 It is possible to compare the two modalities,
11 digital versus analog, in terms of sensitivity and
12 specificity. It is a little complicated because it is
13 really a two-dimensional problem. There is the issue of
14 sensitivity. There is specificity.
15 So one question is how to combine the information
16 from both of those in terms of either a question of
17 superiority of digital over analog or in terms of non-
18 inferiority.
19 One way to do that, a simple way here at the
20 bottom, is to show superiority by, if I can use the little
21 notations, SENS-D is the sensitivity for digital, SENS-A for
22 analog and SPEC for digital analog, then you can use a chi-
23 square test, then, based on McNemar's test. This is getting
24 a little technical here. But you can simultaneously compare
25 them both and answer the question about whether digital is

1 better than analog in sensitivity and specificity combined.

2 There is also another way to do this statistically

3 which is to use what is called Fisher's combination of two

4 test statistics, logarithm and p-values.

5 [Slide.]

6 One of the problems, though, is that because you

7 are really in a two-dimensional problem, because you want to

8 compare sensitivities and specificities at the same time,

9 another approach to this would be to convert it to a one-

10 dimensional problem.

11 One way to do that would be to determine a risk

12 ratio in terms of public-health losses. L-FP would be the

13 loss for making false-positive mistakes. L-FN would be the

14 loss for false-negative errors. If you can quantify what

15 that ratio would be in some public-health metric, then,

16 taking into account the prevalence, which in this particular

17 application is pretty well-known, you can define a slope m

18 as the ratio of 1 minus prevalence over prevalence times

19 this loss ratio.

20 Then you could form the test statistic which is

21 the difference in sensitivities, digital minus analog, minus

22 m times the difference in specificities, digital minus

23 analog. Now you convert the problem to a one-dimensional

24 problem and you can ask and answer questions about whether

25 digital is superior to analog, digital is non-inferior where

1 small risk connotes superiority.

2 In terms of designing the experiment, the sample

3 size depends on knowing the correlational structure between

4 digital and analog for women who are true positives as well

5 as the correlational structure for women who are negatives.

6 [Slide.]

7 Suppose that it is very difficult, as it usually

8 is, to find the true state of each individual person as to

9 whether they have breast cancer or not. What you might

10 consider is some kind of trial where you do some truth

11 resolution. In particular, suppose, for the moment, that

12 you have a random sample of all women that are tested with

13 both digital and analog.

14 And suppose you now look at, in the situation

15 where you have a cutoff between deciding based on BIRADS

16 scale whether the individual has cancer or not, you get a

17 2x2 table where the upper left is positive based on analog

18 and positive based on digital.

19 The off-diagonal elements is where everything is

20 happening. The off-diagonal elements are the individuals

21 where there is a discrepancy between the two tests. If you

22 then can figure out what is happening there, you may have a

23 great insight into what is occurring.

24 In particular, it is possible to do McNemar's test

25 in this case to answer the question of whether there is a

1 difference between the two modalities. In terms of the
2 testing, itself, you can compare sensitivities and compare
3 specificities even though you don't know the absolute
4 sensitivities and specificities by resolving only the off-
5 diagonal elements.
6 So we are back to, really, the same problem as the
7 previous case because McNemar's test depends only on the
8 off-diagonal elements. So there is no need, in that case,
9 to resolve the entire table to figure out what is the true
10 state for women who are positive for both analog and
11 digital.
12 [Slide.]
13 Having said all that, though, you still have the
14 problem of trying to figure out how to size the trial. How
15 big a trial do you need? The problem is that that depends
16 on the correlational structure between digital and analog
17 for both disease-positive women and disease-negative women.
18 You don't know that going into the trial and that
19 creates a problem. It turns out you can approximate the
20 size of the trial with McNemar's test but you probably need
21 to resize the trial along the way.
22 So it is possible to ask and answer the question
23 of whether sensitivities and specificities differ by looking
24 only at the off-diagonal elements but it doesn't really tell
25 you about the performance, in particular, of digital. It

1 doesn't tell you how sensitive and how specific it is.

2 So, in order to do that, what you need to do is

3 also look not just at the off-diagonal elements but sample,

4 at some rate, from the other two cells, the ones that are

5 double-positives, the ones that are double-negatives, for

6 analog and digital.

7 By doing that, you can then estimate the

8 performance of digital and analog as well. You can see how

9 well you are doing by a confidence-interval approach which

10 will depend on how much you are willing to sample.

11 [Slide.]

12 Another approach is to do an agreement study. And

13 agreement studies can be done using a statistic that is

14 called kappa. Kappa measures how much agreement there is

15 compared to average agreement due to chance.

16 So, for example, if you had the five-point BIRADS

17 scale, you can generate the 5x5 table. There is no truth

18 here. You don't know what the true state of affairs is.

19 You don't know which women are positive and which are

20 negative.

21 What you do know is that if all the women line up

22 on the diagonal, on the table of analog versus digital, that

23 then they are in agreement. So the extent to which they

24 differ or wander from that diagonal, that indicates

25 disagreement. A measure of that is called kappa.

1 In particular, a more sophisticated way to do that
2 is to look at what is called a weighted kappa. If one is a
3 2 and one is a 3, digital and analog, that may be a less
4 severe kind of disagreement than if one is a 4 and one is a
5 5. So one could come up with a weight structure that would
6 say, "This disagreement is more severe. I'll give it more
7 weight. This other disagreement, these are very close."
8 So, for example, the difference between 1 and 2
9 clinically may not make any difference in terms of BIRADS
10 here.
11 It is possible using kappa or weighted kappa to do
12 statistical inference, either confidence intervals or
13 hypothesis testing. An important point to make here though
14 is that the issue is not whether kappa is zero. We are not
15 interesting in knowing whether there is any agreement at all
16 between digital and analog.
17 What we are really interested in is what is the
18 level of agreement and is it sufficiently large. In
19 particular, if one goes back to the 1996 guidance document
20 and asks for the unweighted kappa, if you require 95 percent
21 conditional probability for the positives and 98 percent
22 conditional probability for the negatives, what does that
23 translate in terms of kappa?
24 This is the calculation. Kappa is 0.93 and the
25 confidence interval is plus or minus 0.3.

1 [Slide.]

2 People have known for a long time that if you read
3 mammograms more than once, you decrease the variability. So
4 double reads decrease the variability. In fact, it is easy
5 to show that if you have more readers, you can decrease the
6 variability even more. So, in terms of something like
7 median scores, you could use a panel of readers and develop
8 the median score for those readers and then use that in your
9 analysis.

10 Now, you have to be a little careful. If you
11 generate a table that has, for example, n, say, 1,000 women
12 and you use the median scores, suppose you had five readers.
13 The danger is that if you are not careful, you will think,
14 "Oh; there are 5,000 readings here. I will just put them
15 all in one grand table and have a total of 5,000."

16 The problem with that is that the statistical
17 analysis that most people do will get them into trouble
18 there because there are really only 1,000 women. What you
19 are trying to do is decrease the variability by using
20 multiple readers.

21 It is possible that kappa or McNemar's or these
22 chi-square tests can be made valid but it is not a simple or
23 straightforward effort. The point here with multiple
24 readers is really a proof-of-principle point. It is
25 probably not how digital mammography would be intended to be

1 used.

2 [Slide.]

3 In terms of study design, an important point is

4 the following. Suppose that all the women get analog and

5 all of the women who are analog-positive get digital. But

6 only 10 percent of the women who are analog-negative get

7 digital. If you are not careful, you are going to get

8 yourself into a predicament here because it is easy to

9 forget that you are differentially sampling.

10 It made sense to differentially sample but let me

11 do an illustration to sort of bring home this point a little

12 better.

13 [Slide]

14 On the left here, in blue, suppose this is based

15 on 10 percent analog negative, so you do all the analog

16 positives, and now if I compare these two, suppose this is

17 for women who are really true positives, what I would see is

18 a statistically significant difference between sensitivities

19 with analog picking up more than digital. But that is

20 before I forgot to adjust for the fact that I was only

21 looking at 10 percent of the analog negatives.

22 If I now inflate this number and write here the

23 expected number, I would expect that if I look at all the

24 women, instead of just 10 percent of the women who are

25 analog negative, that I would get something like 20 there.

1 So, that is for people who might tend to forget that when
2 you do some kind of inferential sampling here of the analog
3 negatives it needs some kind of adjustment.
4 [Slide]
5 It is possible to do those adjustments with some
6 kind f weighted analysis. It is just a little more
7 complicated but it is not impossible, and you have to still
8 worry about how you do the sample size. That is always
9 problematic.
10 [Slide]
11 In the time that remains, what I would like to do
12 is present some ideas, with permission of Nancy Obuchowski.
13 Nancy is a statistician at the Cleveland Clinic, and
14 presented these ideas at two meetings earlier this year.
15 Her problem was the following -- it is not the
16 same problem. Her problem was let's compare filmscreen
17 analog with its digitized version. So, you take the film,
18 you digitize it, and then you compare the digitized version
19 versus the original filmscreen.
20 One of the issues in terms of equivalence was to
21 figure out what does equivalence mean. In terms of the
22 screening situation, she argued that what you want is some
23 kind of population agreement. You want the screening to
24 agree on a population level and not necessarily on an
25 individual woman. So, individual women could have different

1 results; could not agree on this modality but if on a
2 population basis they did agree, then this would be a
3 methodology which could be used.
4 Now, her approach used receiver operating
5 characteristic area, ROC area, for each of six readers under
6 the two modalities, digitized versus filmscreen. But other
7 methods could, in fact, be employed. She used 30 women who
8 were positives and 30 women negative, and 6 readers.
9 [Slide]
10 She made 10 measurements on each woman, 5
11 locations on each of 2 breasts, and used her earlier work
12 from a 1997 article in Biometrics to do an ROC analysis for
13 each reader. So, each of the readers looked at all 16 of
14 the women's films under both modalities, and generated for
15 each modality a performance measure based on the ROC area.
16 What it did, it adjusted for the correlational structure
17 because of the fact that you are looking at multiple
18 locations in 2 breasts.
19 What she did was she compared variability of the
20 original filmscreen mammogram with itself and then compared
21 the variability of the filmscreen with its digitized cousin.
22 In particular, if you take two readers, two different
23 readers, and you ask what is the average of variability if
24 they looked at filmscreen and then compared that to the
25 average variability of two readers, one of whom is looking

1 at filmscreen and one of whom is looking at the digitized
2 image.
3 [Slide]
4 At the risk of being really technical, this was
5 the statistic she ended up with. I sort of apologize for
6 writing this up here, except that if you look at the second
7 term here, this is the term beta 1 jk. Beta 1 jk is the
8 original filmscreen, one per filmscreen, by reader j of
9 patient k. So, this E stands for expectation. So,
10 averaging over all readers and all occasions where the
11 readers are different and the occasions are different, and
12 you are asking what is the ROC area for one reader at one
13 occasion compared to a different reader at a different
14 occasion. Actually, we made a mistake here. There should
15 be a parenthesis here. Sorry about that. That, in fact, is
16 just a different way of figuring out the regular old sample
17 variance.
18 What you do then is compare it now to this thing.
19 What is this? Well, beta 2 jk is the accuracy of the
20 digitized image, read by reader j, occasion k. So, you
21 compare performance of that individual, reader j, on
22 digitized versus a different reader on the same occasion
23 with the original filmscreen. That is the idea; you compare
24 those two variabilities. In order to do that, what she used
25 was a statistical technique called a bootstrap to compare

1 whether the agreement was close enough or not.

2 In terms of hypothesis testing, you could

3 formulate that in terms of one-sided alternative. What you

4 want to know is, is the agreement within delta^2 some

5 positive number or is it worse. So, in hypothesis testing

6 if you want to show agreement you want to reject this null

7 hypothesis if they differ by more than delta^2 where delta is

8 the upper bound for equivalence. Can you figure out what

9 delta is? Well, it should reflect patient morbidity,

10 mortality, perhaps monetary losses for inaccurate screening

11 decisions.

12 [Slide]

13 She actually went on a little more and did a

14 calculation using sensitivity alone to figure out what a

15 reasonable delta might be. Primarily, I am not sure how she

16 got this but I am reporting it only because it is part of

17 the story. In particular, in her calculations 1 additional

18 death per year per 100,000 patients would give you a delta

19 of a little less than 0.02, which would give you a delta^2 or

20 0.0004.

21 [Slide]

22 So in conclusion, let me say that there are

23 different ways to do this. One is to do trials where you do

24 the sum to estimate the performance of digital sensitivity

25 and specificity, or to compare digital versus analog

1 sensitivity and specificity.

2 The second approach is to do some kind of

3 agreement study where you use statistics to measure how much

4 in disagreement the two modalities are.

5 The third approach would be an approach such as

6 the one that Nancy Obuchowski has done where you measure

7 equivalence of the digital analog relative to the variable

8 of the sample.

9 Thank you.

10 DR. ALAZRAKI: Thank you, Dr. Campbell. The final

11 presentation for this part of the meeting is by Dr. Wendie

12 Berg.

13 **Reader Variability**

14 DR. BERG: First I want to thank the panel for

15 inviting me, and I am going to be presenting results of

16 studies that we did that were presented both at the RSNA and

17 the FDA earlier this year.

18 [Slides]

19 This is work that was supported by the Cohan

20 Foundation, and I would like to thank my co-investigators

21 and particularly my readers, of which you will recognize at

22 least one on the panel.

23 The relevance of this work to digital evaluation,

24 looking at reader variability, is really twofold. One is

25 the detection of lesions, initial detection, and the second

1 is the characterization of lesions which then allows
2 appropriate management. Our study, and I will get into the
3 details, really focuses on the second of these sources of
4 variability, and that is the characterization of lesions.
5 Obviously, it is important, I believe, to quantify the
6 variability in each category to assess the scope of the
7 study needed to establish equivalence of the two
8 technologies.
9 As was already mentioned by Dr. Wagner, really the
10 results I am going to be presenting are in agreement with
11 results that were presented four years ago in chest
12 radiology. This is nothing too dramatically different. We
13 will get into some specifics pertinent to mammography.
14 [Slide]
15 A lot of the data I am going to be presenting is,
16 in fact, kappa values. We did do later kappas only in the
17 sense of equivalence of categories 1 and 2 in the BIRADS
18 Lexicon. But there has been some analysis of kappa as a
19 statistic and, of course, 1.0 is perfect agreement, and I
20 think in our study, unlike the 0.93 that was aiming for with
21 some of the analysis presented this morning, we are going to
22 see numbers down in the 0.4 to 0.2 range with most of the
23 variables looked at in our analysis.
24 [Slides]
25 Of course, you are all familiar with the work of

1 Joanne Elmore and colleagues and, as already mentioned,
2 there was dramatic variability in the rate of recommending
3 additional workup, let alone biopsy, in women with cancer.
4 Her study really was a measure of both variability and
5 detection and variability in management.
6 A study by Dr. Beam, who is on our panel, looked
7 at variability also in interpretation of screening
8 mammograms and, again, showed dramatic variability in
9 screening sensitivity of 40 percent, as well as high
10 variability in specificity with a variability of 45 percent
11 in the rate at which women without cancer were recommended
12 for biopsy. Again, his study looked at differences in
13 detection as well as variability in management.
14 [Slides]
15 There is also a nice study by Baker and colleagues
16 at Duke, very similar to our study that came right after we
17 initiated our evaluation, and that was looking at a series
18 of five radiologists, of whom I think four had been trained
19 in the same setting. One observer also read each case
20 twice, and the lesion was marked on the film. So, in that
21 study they only looked at variability in lesion management
22 and characterization, and they had a kappa for final
23 assessment interreader of 0.43, and within the same reader
24 higher, of course, of 0.65.
25 In our study we had five experienced mammographers

1 who really are only doing full-time mammography. There were
2 103 screening mammograms and 86 with additional evaluation.
3 I am going to concentrate on the results from the 86 with
4 diagnostic evaluation because it was even more complicated
5 looking at screening.
6 These were cases that were all either proven by
7 biopsy or four years of follow-up. In each case the lesion
8 was marked on the films and there were good quality copy
9 films given to the readers. The readers were then asked to
10 reevaluate 13 of the screening and 11 with diagnostic
11 evaluation 2 months later for each of the 5 readers. Again,
12 we were looking at variability in lesion description and
13 management, not in lesion detection.
14 [Slides]
15 We did use the BIRADS Lexicon, with which you are
16 all familiar, asking our readers to rate breast density,
17 lesion type, mass shape, border and density,
18 microcalcification, morphology, distribution and number of
19 associated findings and impression or recommendations.
20 We did make a few modifications, very minor I
21 hope, to the BIRADS Lexicon in that it was my belief that
22 circumscribed masses are easily dismissed as probably benign
23 when they are round, but it may be a little bit more
24 difficult with those that are lobulated. So, we actually
25 divided that into two separate terms. We also grouped

1 together a variety of terms that are used to describe
2 typically benign calcification and lumping them all as
3 coarse calcifications in an effort to minimize that source
4 of variation which we did not think was clinically
5 important. In all our cases, there were at least
6 microcalcifications. So, we did not specifically ask for
7 readers to record the number of calcifications.
8 [Slides]
9 We did ask them to use the standard BIRADS
10 assessment categories. We debated whether to actually
11 separate the impression from the final recommendations, and
12 I think there may be some merit to doing that, as I will
13 talk about later, but we did allow our readers, and
14 restricted them to using the specific recommendations that
15 are tied with the specific impression categories, and
16 particularly category 3. If they were to consider a lesion
17 probably benign, they were restricted to 6-month follow-up
18 on that lesion.
19 In general terms, there was significant
20 variability between our readers just on the most basic of
21 categories, such as lesion type and breast density where
22 there was only moderate agreement even on breast density,
23 with a kappa of 0.43.
24 Screening impression -- you can see there was very
25 little agreement and, again, there are grouped categories 1

1 and 2 together of only 0.21 and for final impression after
2 diagnostic evaluation the kappa was 0.37.
3 [Slides]
4 Just to go through a few cases, this was one case
5 where everyone actually agreed on the management, although
6 the description was variable. I think that was one of our
7 hypotheses as well, that perhaps the initial perception of a
8 lesion might be the problem and if lesions were described in
9 a similar fashion we would at least see similar management.
10 For example, if calcifications were called pleomorphic it
11 was my contention that most of those would then be biopsied,
12 and if a lesion was called circumscribed it would then be
13 followed. Unfortunately, we found that that was not, of
14 course, that simple.
15 In this case, this lesion was called indistinct by
16 several readers and lobulated by other readers, and was
17 always recommended for biopsy, and this was a small
18 infiltrating cancer.
19 [Slide]
20 In this case, this was described as circumscribed
21 round by one reader, circumscribed lobulated by three, and
22 obscured by another, and the management was highly variable.
23 There was one reader that followed this and the other four
24 biopsied. It was a collared cancer.
25 [Slides]

1 Looking at mass borders for example, and I will go
2 through some of the specifics, probably more than you want
3 to hear, but there was the best agreement between
4 circumscribed and spiculated. Those are the two terms where
5 we saw good agreement with relatively high kappa values. I
6 have listed other terms that were most commonly used on the
7 same case by other readers. Where it is listed in yellow,
8 the other term was actually used more frequently than the
9 term on the left-hand column. The one that stands out the
10 most, of course, is obscured masses where there was only 6
11 percent agreement between readers, and it was much more
12 often called an indistinct mass by another reader or no mass
13 at all by some readers. Again, for spiculated, fortunately,
14 I think we showed relatively good agreement. As has been
15 previously shown by Baker's study, focal asymmetric density
16 is a difficult area for radiologists, and there was very
17 little agreement there and most of the time it was called an
18 indistinct mass by another reader.
19 When we look at the correlation between the
20 description and final assessment categories -- I apologize,
21 there is a lot of information on these slides but I have
22 ranked the descriptor by the assessment and, again, based on
23 the elegant work by Dr. Sickles, we would expect
24 circumscribed masses that were so described to be considered
25 at least benign or probably benign by most readers, and that

1 was generally the case, with the exception that there were a
2 number of cases even after the diagnostic evaluation that
3 was provided to our readers where they still wanted more
4 diagnostic evaluation.
5 We did see, of course, that spiculated masses were
6 always considered category 4, 5 or 0, and that was
7 reassuring. It was a little bit problematic to me but we
8 had indistinct masses that were considered outright benign.
9 I think there were some problems in applying the BIRADS as
10 we expected. I have failed to mention that the benign
11 lesions are listed in white here and the yellow are the
12 malignancies. Indeed, all the masses were at least properly
13 biopsied that were malignant, or at least recommended for
14 additional evaluation.
15 [Slides]
16 For mass shape there was very little agreement.
17 It was basically random by all statistical analyses. I
18 guess that didn't surprise me too much, although I would
19 have thought that irregular versus round might have had a
20 little better agreement.
21 For mass density, these were only the diagnostic
22 cases. So, fat-containing masses, by and large, were
23 outright dismissed on the basis of screening views as being
24 benign and did not have additional diagnostic evaluation to
25 provide our readers. So, there were no cases in this group.

1 There were very few low density masses. So, these numbers
2 probably aren't too helpful.
3 [Slides]
4 For calcifications, the case on the left was
5 problematic to all the readers, and I think that is quite
6 understandable. This was a patient who was undergoing her
7 annual follow-up after having had a very remote breast
8 cancer, and we could argue whether this is truly a screening
9 case but she had calcifications that had developed at the
10 lumpectomy site. These were considered coarse by all of our
11 readers, and four of the readers dismissed them as outright
12 benign and one probably benign. Unfortunately, this was
13 recurrent cancer.
14 Another case with calcifications was considered
15 punctate by three readers and pleomorphic by others. It was
16 variable whether they considered these to have associated
17 asymmetric density or not. This was recommended for biopsy,
18 fortunately, by all our readers and it was a small case of
19 DCIS.
20 [Slides]
21 For calcification morphology, I was surprised that
22 there was very poor agreement on the use of the terms for
23 typically benign calcifications, even allowing lumping of
24 all of those terms together, and we had only 17 percent
25 agreement most of the time. Another reader would call these

1 punctate or even pleomorphic. So, again, this was a problem
2 because these are the ones that we would like to dismiss as
3 outright benign whereas pleomorphic calcifications, as a
4 rule, we would recommend for biopsy.
5 Milk of calcium was appropriately recognized but,
6 again, there was overlap between that and punctate and
7 pleomorphic calcifications. Again, this would have
8 implications for different management.
9 We are going to talk a little bit more about
10 punctate calcifications. At face value, the agreement was
11 not too dissimilar to other categories, but there was a lot
12 of variability in the management, as had been seen in the
13 study by Baker and colleagues.
14 Fortunately, branching of pleomorphic
15 calcifications, as I will talk about more, were recommended
16 for biopsy as a rule, but there were some problems even
17 there. Looking at management, and there was not terribly
18 good agreement even with these, and intuitively I would have
19 thought branching calcification would have been a good
20 candidate for good agreement.
21 When we look at the correlation between the
22 descriptor and the management categories, and obviously this
23 is simplistic, not accounting for distribution as well but,
24 by and large, these were mostly clustered cases. Coarse
25 calcifications in general were considered benign, including

1 that including that one problematic case that I listed.
2 These are just individual readings so that one problematic
3 case is counted as five separate readings of cancer up here.
4 Punctated calcifications, as I mentioned, were
5 very problematic, and I think are quite problematic in
6 general, at least trying to teach my colleagues, my
7 residents that this is a real problem area. I would like to
8 see Dr. Sickles help us out on this one some more.
9 Amorphous calcifications, again, are somewhat
10 problematic although mostly they were recommended for
11 biopsy. Branching calcifications at least were uniformly
12 recommended for biopsy, as we would hope, since these tend
13 to reflect early cancer.
14 Pleomorphic calcifications -- I was surprised that
15 people would actually surprised cases as being pleomorphic
16 and outright benign. I show at least one of these cases
17 where it was, in fact, a cancer that was described as being
18 pleomorphic but then followed.
19 [Slide]
20 Distribution of calcifications -- again, there was
21 good agreement on cluster, fortunately. I would hope so.
22 Segmented calcifications, also pretty good
23 agreement.
24 Linear, not such good agreement and, actually,
25 most of the time those were considered clustered or

1 segmental and, obviously, there is some overlap because
2 segmental implies a duct and its branches which, in a sense,
3 is a linear distribution. So, I can appreciate that
4 difficulty.
5 Looking at diffuse calcifications, there was very
6 little agreement but we had very few cases in that category.
7 [Slides]
8 Overall, looking at the kappa values for inter-
9 observer variability and, on the left-hand side are our
10 results and on the right-hand side are the results of Baker
11 and colleagues. By comparison, you can see that for mass
12 borders overall the kappa was 0.4 which is only moderate
13 agreement. Mass shape was 0.28. These are kind of far from
14 that 0.93 we were hoping for. Mass density, 0.4, and
15 microcalcifications, 0.36, etc.
16 Associated findings -- we weren't terribly strict
17 about forcing people to answer that, although one of my
18 colleagues, Dr. Campassi, sat there as the cases were
19 completed to make sure that every form was filled out
20 completely. So, I was a little surprised to see that that
21 was so low for associated findings.
22 In terms of variability with Dr. Baker's results,
23 you can see that each individual number is significantly
24 better than the results in our study, except for special
25 cases, and really I think this probably reflects the fact

1 that he had a more homogeneous group of readers that had all
2 been trained together. I think perhaps our results are more
3 reflective of the general radiology population. Of course,
4 these are mammography specific folks.
5 [Slides]
6 Final impression categories -- I thought this was
7 particularly interesting. We had one case that was
8 considered outright normal by one reader, even though the
9 lesion was marked and it was a cancer. I thought that was
10 interesting.
11 For benign lesions there was pretty good
12 agreement. Also, with highly suspicious lesions there was
13 pretty good agreement, with an overall kappa of 0.63. Where
14 we saw the problems, as has been alluded to, was with
15 category 3 lesions. There was only 8 percent agreement for
16 masses; 20 percent agreement for calcifications for even
17 suspicious lesions, and there was significant overlap for a
18 lesion that was called category 3 by one reader and had a
19 greater chance of being called benign, suspicious or needing
20 additional evaluation by another reader.
21 Interestingly, and this is actually perhaps
22 germane to the work that was proposed by Dr. Obuchowski's
23 study, if you look at the overall population of our
24 readings, if you will, or our cases, they actually
25 correlated reasonably well with what should be from the

1 standpoint of positive predictive values in that the lesions
2 that were considered benign, even after diagnostic
3 evaluation -- there were four, perhaps a little high but
4 that was mostly that one case of coarse calcifications, and
5 they were mostly, in fact, benign. Those that were
6 considered probably benign are not quite the 1.4 percent
7 that we would like to see but pretty close -- 5 percent of
8 those actually were cancer. Suspicious lesions that were
9 considered suspicious, 33 percent of those were cancer; and
10 78 percent that were highly suspicious. Actually, 20
11 percent of those needing additional evaluation was very
12 close to the work that was presented at RSNA by Dr. Orel and
13 colleagues from their experience as well.
14 So overall, looking at the whole population, even
15 though there was so much variability, we actually did pretty
16 well. Readers did pretty well.
17 [Slide]
18 Intrareader variability was interesting as well.
19 Again, there was a two-month time period between the initial
20 reading and the subsequent reading, and the cases were
21 picked randomly from our overall group but all readers saw
22 the same group of cases the second time. Thus, for density,
23 again, there was pretty substantial disagreement even on
24 breast density. Lesion type did better. Mass borders,
25 better than interreader but there was significant

1 variability with moderate agreement. Microcalcification
2 morphology, again, varied a lot across our readers, with
3 some readers being very consistent and others not.
4 Interestingly, final impression across the readers varied,
5 with only 45 percent agreement with one of our readers.
6 Again, this is not grouping 2,3 or things like this. So,
7 this exaggerates some of the variability. It is not
8 necessarily all clinically significant variability.
9 [Slides]
10 Getting into that a little bit more, what
11 constitutes agreement? I think this is going to be the
12 focus of some discussion this afternoon. Just looking at
13 our readers, grouping normal, benign and probably benign --
14 categories 1, 2 and 3 which actually is what has been
15 recommended in the audit for the BIRADS MQSA inspections
16 now, and then grouping suspicious with highly suspicious and
17 those needing additional evaluation, and then looking at our
18 final impression of our readers within themselves, there was
19 complete agreement. This was actually complete specific
20 agreement over half of the time but not always, an average
21 of 55 percent of the time. But there was major
22 disagreement, in other words, initially they said follow and
23 the subsequent time they said biopsy or at least needs
24 additional evaluation. There was significant disagreement,
25 from 0 percent to 27 percent of the time across individual

1 readers.

2 [Slides]

3 Finally, looking at the consensus assessment, in
4 other words, all readings where, grouping together, they
5 recommended for biopsy or immediate evaluation, there were
6 47 percent of the readings of benign lesions that were in
7 this category. So, not a terribly great specificity. But
8 94 percent readings of cancers were recommended for biopsy
9 or immediate workup. This is where all the readers agreed
10 in the sense of either biopsy or immediate workup. So,
11 pretty good agreement on cancers overall.

12 As I mentioned, there was good agreement even on
13 the one case of cancer that was recommended for routine or
14 short interval follow-up, but at least there was agreement.
15 They were wrong but at least they agreed.

16 Disagreement was really very minor on the cancers.
17 There was 26 percent of readings on benign lesions where at
18 least one reader recommended the lesion for biopsy and the
19 others were follow-up.

20 [Slides]

21 Just to go through two of the cases that were
22 problematic and, I apologize, these don't project too well
23 but on the left is the one case that one reader actually
24 called these pleomorphic, as did all the other readers, and,
25 yet, it was considered probably benign and followed. This

1 was a case of ductile carcinoma in situ.

2 On the right-hand side is the case that was

3 considered coarse by one of our readers and pleomorphic by

4 the others. It was described as coarse and then followed,

5 and this was another cancer.

6 So, there were errors both in perception

7 description, if you will, as in this case, and in management

8 even though the description agreed. So, there were both

9 sorts of errors we observed.

10 [Slides]

11 In summary, there was significant disagreement on

12 category 3, even more so for masses than calcifications,

13 with overlap for both benign and suspicious categories. One

14 problem came to light in talking about these results with

15 our readers. One of the readers commented, "but I biopsy

16 lesions that I consider probably benign." So, obviously

17 there is a difference between the impression and the

18 management but BIRADS does not allow that kind of leeway.

19 By way of background, as I mentioned, there have

20 been nice, elegant studies by Dr. Sickles and colleagues

21 looking at what is a lesion that should be considered

22 probably benign, and has described several categories with

23 circumscribed masses, punctate calcifications, asymmetries,

24 etc., and has shown less than a 2 percent risk of

25 malignancy. Unfortunately, the rest of us may be not so

1 good at recognizing what is a punctate calcification and
2 that is the problem.
3 [Slide]
4 So, perhaps if we could educate ourselves on the
5 lesions that can be appropriately followed, such as specific
6 examples of coarse calcifications and punctate
7 calcifications, that might improve our performance and
8 agreement between readers.
9 Thank you.
10 DR. ALAZRAKI: Thank you, Dr. Berg. At this point
11 we will take a ten-minute break and resume after the break
12 with the industry presentations.
13 [Brief break]
14 DR. ALAZRAKI: We will now proceed with the
15 industry presentations on digital mammography. The first
16 presentation will be made by Mr. Richard Bird, Director of
17 Clinical Development for the Trex Medical Corporation.
18 **INDUSTRY PRESENTATIONS**
19 **Trex Medical Corporation**
20 MR. BIRD: Good morning.
21 [Slide]
22 I am Richard Bird, Director of Clinical
23 Development for Trex Medical Corporation. I thank the panel
24 and the FDA for this opportunity to participate in today's
25 important decisions.

1 As you know, the Trex Medical Corporation has
2 worked over the past several years to bring full-field
3 digital mammography systems to market. Our purpose in being
4 here today is to propose different methods for the
5 evaluation of digital mammography.
6 We feel uniquely qualified in making these
7 recommendations for several reasons. First, we are the only
8 company to have completed a clinical study of digital
9 mammography. Second, we are the only company to have filed
10 a 510(k) notification for digital mammography with the FDA.
11 Our filing is currently under review. We have spent the
12 last two years applying the elements of the guidance to
13 real-life clinical situations, and we have spent the last
14 nine months interpreting the data generated by our clinical
15 study. We feel that the guidance does not adequately
16 address real-world clinical practice and variables.
17 [Slide]
18 There are four elements of the guidance which we
19 will make, and support specific recommendations. There is
20 the study design. An agreement type of study versus a
21 noninferiority type study; positive and negative
22 stratification; looking at pure BIRADS versus screening
23 decision processes; judging methods using 2 readers versus 6
24 readers from a pool of 16 radiologists; and data analysis,
25 agreement versus sensitivity and specificity.

1 [Slide]

2 To assist me in presenting this information, we
3 have assembled several participants in our digital
4 mammography program. Dr. Lawrence Bassett, Professor of
5 Radiological Sciences at UCLA School of Medicine and
6 Director of Breast Imaging at the Iris Canter Breast Center,
7 served as a principal investigator in the clinical study and
8 consulted on the clinical aspects of the study design. Dr.
9 Bassett will speak on the clinical significance of making
10 positive and negative determinations.

11 Dr. Richard Olshen, Ph.D., Professor and Chief,
12 Division of Biostatistics, Stanford University, who designed
13 the statistical elements of the study and oversaw the
14 analysis of the study data, will speak on the statistical
15 issues affecting proper evaluation of the two modalities.

16 Bradley Betts, a doctoral candidate at Stanford
17 University, under the direction of Prof. Olshen performed
18 the data analysis. Mr. Betts will speak on the statistics
19 and the models that provide support for our recommendations.

20 Both UCLA and Stanford have received financial
21 contributions from Trex Medical, however, none of the
22 investigators has received any direct funding.

23 I will then wrap up the presentation with a short
24 conclusion.

25 [Slide]

1 Let me begin the presentation by speaking about
2 the study design, starting with our recommendation for a
3 noninferiority study. Our frame of reference as a
4 recommendation is a clinical study we performed in which 520
5 patients having analog screening mammograms received digital
6 mammograms as well for purposes of comparison.
7 Early in the design of our experiment, we became
8 convinced that an agreement type study was not the best way
9 to evaluate digital mammography. The expectation that
10 digital mammography would most likely outperform analog
11 mammography was the reason for this decision. If assumed
12 that the digital modality gave better results in terms of
13 imaging, disagreement with the lower performing analog
14 modality would cause rejection in the hypothesis that
15 digital mammograph agrees on a case by case with the analog
16 mammography.
17 In light of the possibility that superior
18 performance would lead to failure, the noninferiority
19 paradigm appears more appropriate. The noninferiority
20 hypothesis can be stated as digital mammography performs no
21 worse than analog mammography, and accommodates better
22 digital performance while ensuring that the present high
23 standards demonstrated by conventional mammography are
24 maintained.
25 In the noninferiority paradigm both modalities are

1 compared against a separate gold standard. It is important
2 that the gold standard or the truth set involve the use of
3 information beyond what is available from the mammogram
4 alone or the test judges. Dr. Bassett will discuss the
5 development of the truth standard.
6 This slide shows digital as being compared by the
7 analog, and this is effectively the truth set that you have
8 to compare to here, and in a noninferiority the independent
9 truth of the analog is compared independently as the digital
10 is compared to that truth set.
11 Our goal of the study was to demonstrate the
12 hypothesis for the indicated use of screening. This is
13 complicated by such matters as intra-observer variability,
14 inter-observer variability and selection bias.
15 [Slide]
16 Because of these, agreement studies are unlikely
17 to succeed even assuming the devices are absolutely
18 equivalent. This will be addressed by Mr. Betts.
19 At this time, I would like to turn this over to
20 Dr. Bassett.
21 DR. BASSETT: Thank you, Richard.
22 [Slide]
23 I would like to start with some very basic
24 definitions that are relevant to our study. One is the
25 difference between screening mammography and diagnostic

1 mammography. These are basically from the Agency for
2 Healthcare Policy and Research. Screening mammography is
3 defined as an examination of an asymptomatic woman to find
4 unexpected breast cancer. We want to find it when it is
5 early and curable. Diagnostic mammography is examination of
6 a symptomatic patient to evaluate clinically or
7 mammographically a detected lesion.
8 [Slide]
9 Before we look at how this works in our study, let
10 me just review the BIRADS categories as there were a couple
11 of questions about them, and they are relevant to some of
12 our data. One, of course, is negative which means that
13 there was nothing to comment on in the mammogram. It was
14 totally negative.
15 Two is benign findings, which means there were
16 findings but we can say that they are typically benign and
17 definitively benign.
18 Three is probably benign. You have heard about
19 that. These are the ones that are likely enough to be
20 benign that they can be managed with surveillance.
21 Four is a suspicious abnormality which merits a
22 biopsy. Five is highly suggestible of a malignancy, which
23 is a classic carcinoma which can be managed by biopsy and in
24 some practices by surgery.
25 There is a whole other group here that we need to

1 also understand, and that is the category or the group where
2 they cannot be assigned to one of these categories until
3 there is additional workup. Some people would call this 0
4 or the additional workup needed group.
5 [Slide]
6 Our study actually looked at actually screening
7 practice. We tried to model it after a typical screening
8 practice, and in a screening practice we were comparing how
9 well the digital and the analog images did in determining
10 whether the screened patient mammogram was normal or
11 abnormal. So, it is a binary decision process. We were
12 only using the digital mammograms in that kind of a setting,
13 in the screening setting.
14 [Slide]
15 The whole process is shown here, and it is a
16 little more involved. After the woman is screened, if it is
17 normal she goes to routine follow-up with screening
18 mammography at the next recommended interval. If it is
19 abnormal, then she has to have a diagnostic workup, and that
20 is usually done for the cases that are called abnormal. So,
21 the patient may actually have to be recalled if she has
22 already gone home. In those cases, the diagnostic workup
23 would determine then whether the case should be classified
24 as benign or probably benign, or suspicious or highly
25 suggestive of malignancy. Those cases will actually go on

1 to biopsy to determine if they are benign or carcinoma. So,
2 this is the whole process.
3 In fact, our study, however, which used digital
4 mammography only for the screening part of the study, really
5 involved this part of the screening versus diagnostic
6 scenario. We felt that the guidelines that were provided by
7 the FDA really were stressing too much this part of the
8 workup, the diagnostic workup, and cases suspicious and
9 highly suggestive of a malignancy, the 4 and 5 type cases.
10 However, we do appreciate the importance of the diagnostic
11 workup in determining a truth set for the cases that were
12 called abnormal.
13 [Slide]
14 So, now I can tell you how we developed the truth
15 set. The truth set development was performed by the
16 principal investigators at each patient enrollment site, and
17 these were all very experienced interpreting physicians.
18 The truth was determined to be the best screening decision
19 making for each case. That was determined by the principal
20 investigators retrospectively based on the actual case
21 outcome, which could include any workup or biopsy that was
22 done.
23 This is very similar to the process of the medical
24 audit to determine mammography effectiveness which is
25 conducted at mammography facilities, and is recommended, and

1 perhaps even required, by the mammography quality standards.
2 I should also mention that in addition to this
3 study there was another issue we addressed. That was the
4 subjective assessment by the principal investigators as to
5 which of the two modalities, the analog screening or the
6 digital screening examination, gave them the best
7 information in terms of the final or the truth, or the gold
8 standard. In fact, this was overwhelmingly determined by
9 the principal investigators to be the digital imaging study.
10 Now Dr. Olshen will present some of the statistics
11 that were developed in the course of this examination.
12 DR. OLSHEN: Thanks, Richard and Larry.
13 [Slide]
14 Planning of our study and analyses of data have
15 been carried out by members of the Signal Compression
16 Classification Group in the Department of Electrical
17 Engineering at Stanford. It includes engineers,
18 radiologists, statisticians and students. The group was
19 headed by Bob Gray, who is Professor and Vice Chair of
20 Electrical Engineering, and me. I am Professor and Chief,
21 Division of Biostatistics in the Department of Health
22 Research and Policy in the Stanford University School of
23 Medicine.
24 The group has worked in CT, MR, image compression
25 and now mammography. We have been funded by the NIH, the

1 Olney Breast Cancer Initiative, and especially by the
2 National Science Foundation, and also private sector
3 sources, including Trex Medical who supplied gift funds for
4 the support of graduate students. We have worked
5 independently for many years, and in part the last several
6 with Trex Medical.
7 [Slide]
8 I am here to share with you some principles by
9 which we design and analyze clinical studies, and to share
10 qualitative aspects of some of our findings, especially as
11 they relate to revising the FDA guidance document.
12 Studies are designed to be faithful to how
13 medicine is practiced. It should enable understanding and
14 evaluation of real sources of variability and bias in
15 studies of screening, as has been addressed by previous
16 speakers.
17 [Slide]
18 The bases of our analyses are sample size, power,
19 and real data, 2 X 2 agreement tables for binary outcomes in
20 which columns correspond to analog, right or wrong, and rows
21 to digital.
22 We stratified the studies by findings in the truth
23 set, positive or negative. The former group is germane to
24 sensitivity. The probability of study is found positive
25 given it is positive in the truth set; and predictive value

1 positive, the probability the study is positive in the truth
2 set given it is found positive. The negative group is
3 germane to the analogous notions of specificity and
4 predictive value negative.
5 We have expertise in ROC analyses but do not
6 pursue traditional approaches here because they violate some
7 goals of experimental design and may not well model reality
8 of practice and data. They tend not to be well suited to
9 quantifying management. Their use can entail introduction
10 of unnecessary variability, and in the worst case
11 inappropriate distributional assumption, Gaussian or
12 Poisson.
13 The bottom line for our study has been this: no
14 matter how you view its data, one struggles to find
15 differences between analog and digital. Small statistically
16 insignificant differences in sensitivity and specificity
17 seem to owe to quantifiable selection bias, while predictive
18 value positive and predictive value negative, which are less
19 subject to bias, differ only in a third decimal place.
20 These findings come from analog and digital
21 studies of more than 500 women, each read by not fewer than
22 6 radiologists from among 16 at multiple institutions. The
23 findings have been validated by simulation. On the other
24 hand, differences among radiologists are patent, though
25 individual radiologists differ little in their performances

1 as modality varies.

2 Before turning over the floor to Brad Betts, who

3 will illustrate the work that led me to these conclusions, I

4 do want to respond to a previous request and say that I have

5 no financial interest in anything whatsoever to do with Trex

6 Medical or any of its affiliate companies, nor have I

7 personally received any compensation from them.

8 MR. BETTS: Thank you, and good morning.

9 [Slide]

10 My name is Brad Betts, and I will speak today on

11 statistical issues related to determining whether or not

12 digital and analog mammography are equivalent. A question

13 that is deceptively simple to state, in fact, proves

14 difficult to properly test. Sources of variability

15 intrinsic to a study of mammography, coupled with the

16 perception that analog and digital technologies are

17 relatively similar account for this difficulty.

18 The challenge is to design a clinical study that

19 mitigates the effects of these sources of variability while,

20 at the same time, keeping the study practical. If one

21 believes that, in fact, digital might truly be better than

22 analog, the paradigm of an agreement study, as discussed in

23 the original guidance document, no longer seems optimal.

24 In such a framework, the possible superiority of

25 digital would work to show disagreement, perhaps penalizing

1 an actually superior technology. Instead, the paradigm of a
2 nonsuperiority study seems useful. Here, both technologies
3 are compared to an independent notion of the truth. This is
4 the gold standard of which Drs. Bassett and Olshen have
5 already spoken. Agreement has now been replaced by judges
6 either being right or wrong. Results can be organized, as
7 shown by Dr. Olshen, in 2 X 2 tables. Differences between
8 sensitivity and specificity values for the two technologies
9 can be formalized into statistical hypotheses.
10 In any clinical mammography study three parameters
11 are under the direct control of the designers, as shown
12 here. They are the number of judges to be used; the number
13 of patients to be studied; and the prevalence of abnormality
14 in the study population. Some parameters are, instead,
15 imposed. Three important ones have already been identified
16 in today's earlier presentations. They are intrareader
17 variability, interreader variability, and selection bias.
18 [Slide]
19 Based on data collected at Stanford during the
20 course of our own clinical studies in mammography, the
21 degree of interreader variability can be assessed.
22 Depending on the definition of agreement, we have seen rates
23 of disagreement as low as 10 percent and as high as 32
24 percent. The same study also allows a degree of interreader
25 variability to be quantified.

1 Data collected from Trex's study can also be used
2 for this purpose. For example, on the analog modality the
3 16 judges in the Trex study had a mean sensitivity of 73.7
4 percent and a standard deviation of 15.7 percent.
5 Only the selection bias is difficult to quantify.
6 All that can be said with certainty is that it exists to
7 some degree, and complicates in a study that is practical
8 determination of whether or not the two technologies are
9 equivalent.
10 [Slide]
11 Quantifying these parameters is important to make
12 an accurate prediction of how many judges and patients need
13 to be studied. Based on models we have developed at
14 Stanford, models which explicitly take into account all the
15 listed sources of variability, it seems clear that using
16 only two judges, a tie-breaker, and 500 patients in an
17 agreement framework would make success unlikely other than
18 by chance.
19 Although the effects of intrareader variability
20 can be reduced in this scheme, intrareader variability,
21 acknowledged in the literature and supported by our
22 research, is not properly addressed. Not only are there
23 fundamental questions of statistical size and power in face
24 of the sources of variability, but one is left wondering
25 should such a study show agreement whether or not the two

1 technologies are really equivalent or just equivalent for
2 the two doctors. Using 16 judges, as Trex has done,
3 provides more statistically believable results. Just as the
4 sources of variability impact upon an agreement studies, so
5 do they upon a noninferiority one. Their affect must be
6 accounted for in the design, and will be felt in the
7 observed data.
8 Once again, modeling analysis at Stanford using a
9 variety of techniques, such as bootstrapping, Monte Carlo
10 simulations, McNemar's test, normal theory approximations,
11 binomial distributions, all have been completed. They
12 suggest that 16 radiologists, 520 patients, a rate of
13 abnormality of 50 percent, and not less than 6 separate
14 readings of each patient should suffice to make a suitable
15 determination of equivalence. As always in matters of
16 statistics, the term suitable can be made precise.
17 One particular such analysis, when applied to
18 Trex's data allows for the following precise statement,
19 which I have summarized down here, and it says that in their
20 study with size 0.05 and power 0.88 the hypothesis of analog
21 being more specific and more sensitive than digital by at
22 least 5.4 percent is rejected. That is, with 95 percent
23 confidence analog is not better than digital by 5.4 percent,
24 and is, in fact, not demonstrably better at all.
25 I will ask Richard Bird to make some concluding

1 remarks.

2 MR. BIRD: In conclusion, I hope we have been able
3 to come here with our experience from an actual clinical
4 trial to provide the panel with guidance that can be used
5 effectively by ourselves and by other manufacturers. We
6 recommend that we use a noninferiority study rather than an
7 agreement study. We recommend a clinical practice paradigm
8 with an emphasis on workup as a screening decision rather
9 than BIRADS Lexicon. We also emphasize the use of multiple
10 readers to impact the importance of reader variability,
11 which is a real-world scenario that has to be dealt with.
12 Finally, any study designed to detect differences should be
13 sensitive to the effects of these known sources of
14 variability and bias.

15 I thank you all for your attention.

16 DR. ALAZRAKI: Thank you, Mr. Bird. The next
17 speaker will be Dr. Edward Hendrick from General Electric
18 Medical Systems.

19 **GE Medical Systems**

20 DR. HENDRICK: Thank you. It is a pleasure to be
21 here this morning.

22 [Slide]

23 My name is Ed Hendrick, and I am Professor and
24 Chief of the Division of Radiological Sciences at the
25 University of Colorado Health Sciences Center. I am not an

1 employee of GE Medical Systems. The University of Colorado
2 Health Sciences Center has a research agreement with GE to
3 acquire data toward 510(k) approval, to analyze that data
4 and to help with the protocol design. I am here in that
5 regard. I don't own stock in GE or in any company related
6 to GE that I know of.
7 [Slide]
8 What I want to talk about is specifically the
9 difficulties with the guidance document, and the approach
10 recommended in the guidance document. I am not going to
11 present results from the GE equivalence study which is in
12 progress right now. I want to specifically talk about the
13 difficulties posed by the guidance document in trying to get
14 this technology approved by the FDA.
15 One of the basic problems, as Richard Bird just
16 pointed out, is that if you restrict yourself to an
17 equivalence study, the study may fail just because digital
18 is superior to filmscreen.
19 In addition, if you restrict yourself to an
20 equivalence sort of trial, once you have established
21 equivalence, assuming you could, marketing becomes difficult
22 because you can't make claims beyond what you demonstrated
23 to the FDA. So, you are trying to market a device that
24 costs four to five times that of a regular mammography unit
25 on the basis of it being equivalent to that more inexpensive

1 unit.

2 [Slide]

3 I want to go back and look in some detail at the
4 issue of reader variability, and I just want to point out
5 one of the sources of variability that comes in, in an
6 equivalence study. Reader variability, as was nicely
7 pointed out by Dr. Wagner, could be because readers have
8 different ROC curves, possibly different levels of
9 experience, or different readers operate on the same ROC
10 curve but at different points on that ROC curve because they
11 have different decision points for what they call positive
12 and negative.

13 Once you eliminate the truth about the cases and
14 only look at equivalence of filmscreen and digital, you lose
15 this information on whether you are talking about readers
16 with fundamentally different ROC performance or just
17 operating at different decision points. So, you are in the
18 area where reader variability looks quite large when you are
19 restricted to an equivalence study.

20 On the other hand, something that involves finding
21 out the truth delays proving equivalence or superiority of
22 digital for a large amount of time to do the appropriate
23 follow-up.

24 There is also though, between screenfilm and
25 digital mammography, a positioning variability due to the

1 repositioning of the patient, and the possibility of
2 different positioning leading to different false-positive
3 calls in the two modalities, which play a role in off-
4 diagonal elements in the equivalence study. Then, there are
5 true modality differences, which is what you would really
6 like to get at in these studies.
7 [Slide]
8 The best reference I know relevant to the issue of
9 an equivalence study and the data that would need to be
10 evaluated in an equivalence study is the reference by Joanne
11 Elmore on variability in radiologist interpretations, which
12 has already been mentioned, and here is the complete
13 reference.
14 [Slide]
15 I want to look in some detail at this study to
16 point out where it is similar and dissimilar to the kind of
17 equivalence study that the FDA is recommending. In Elmore's
18 study, instead of 500-600 cases they used 150 hand-selected
19 cases, all on screenfilm, and they were only looking at
20 radiologist variability in reading the same filmscreen
21 cases.
22 This is the distribution of cases, and 54 were
23 originally read as normal; 1 had a cancer; 61 were read as
24 abnormal, probably benign, and 7 had cancer; and 35 were
25 read as abnormal, suggestive of cancer and 19 did have

1 cancer.

2 [Slide]

3 The diagnostic categories or interpretation

4 categories were not BIRADS comparable. There were four

5 categories only. It looks like the BIRADS categories 4 and

6 5 were collapsed into one and there was no category 0

7 permitted.

8 [Slide]

9 The follow-up recommendations -- and this becomes

10 important because this is one of the ways that results were

11 analyzed. The Elmore study distinguished the interpretive

12 categories from the follow-up recommendations, and these

13 were the recommendations that were possible. One of the

14 analyses looked at these recommendations and separated

15 recommendations for biopsy from everything else.

16 [Slide]

17 There is a little bit of difficulty in the

18 methodology of Elmore's study in that 10 radiologists were

19 used to interpret each case. If you look at the maximum

20 number of possible agreements for each case, it is like the

21 maximum number of toasts you can have or people clinking

22 each other's glasses at a table. If you have 10 people at

23 the table, you can have 45 clinks of the glasses, that is,

24 pairs of readers. So, if you look at it in that regard in

25 terms of the number of possible agreements, there were 6750

1 possible agreements -- 45 pairs times these 150 cases read
2 by each of the 10 readers.
3 But if you look in terms of the number of possible
4 disagreements, it was fewer because the worst case scenario
5 was that on a given case among these 10 radiologists, 5
6 called it positive and 5 called it negative, and that led to
7 a maximum of 3750 possible disagreements. So, there is a
8 pretty big disparity in what you can use as the denominator
9 for saying what the level of agreement is.
10 [Slide]
11 These are the results looking in terms of
12 agreement. So, the denominator is 6750. When the question
13 was how often did they agree in terms of diagnostic
14 interpretation, looking at interreader agreement it was 78
15 percent. Looking at intrareader agreement, which was
16 studied in a smaller sample, it was 84 percent. This is
17 agreement of the same reader reading about 6 months later.
18 In terms of biopsy recommendation, the interreader
19 agreement was 85 percent, and the intrareader agreement was
20 91 percent.
21 Since we are talking about an equivalence study
22 where you would parse the reading of the ACR BIRADS
23 categories into basically a negative set and a positive set,
24 if you just take this biopsy recommendation, even though the
25 dividing line is somewhat different, this I think gives you

1 a reasonable level of what you could expect to be the biopsy
2 recommendation to give you a reasonable estimation of
3 interreader agreement and intrareader agreement in the study
4 that we are talking about as an equivalence study, with one
5 proviso. That is, here the denominator was that 6750. If
6 you look at it in terms of maximum number of possible
7 agreements, this is just 1 minus the 79 percent and 1 minus
8 the 85 percent.
9 [Slide]
10 But if you look at it in terms of the number of
11 possible disagreements, where the 3750 is the denominator,
12 then the diagnostic interpretations disagreed 40 percent of
13 the maximum number of possible disagreements, and the biopsy
14 recommendation disagreed 27 percent of the maximum number of
15 possible disagreements.
16 So, I would take these numbers as sort of the
17 range of likely values for disagreement, interreader
18 variability for biopsy recommendations, somewhere between 15
19 and 27 percent.
20 [Slide]
21 The equivalence study results I tested on is this
22 p1 value, which is the fraction of screenfilm positive cases
23 that were assessed as positive by digital mammography. So,
24 it is the conditional probability that digital is positive
25 given the screenfilm is positive. And, p2 is the other

1 decision point, which is the fraction of screenfilm negative
2 cases that are assessed as negative by digital whether they
3 were truly positive and negative or not.
4 [Slide]
5 This leads to the construction once you have your
6 results of a 2 X 2 table. Normally this would be a truth
7 table, but there is no truth in this agreement study as
8 posed. So, the vertical plus and minus are for filmscreen
9 assessment. If filmscreen called it positive it is in this
10 column; if filmscreen called it negative it is in this
11 column. Then, in the upper left-hand corner are the cases
12 that filmscreen called positive that digital also called
13 positive.
14 In this case, if you had 100 cases called positive
15 -- this is just an example with half filmscreen positive,
16 half filmscreen negative -- in this case, of the 100
17 filmscreen positive cases, 90 were called positive by
18 digital. These are not actual results, but just sort of a
19 simulation of results. Ten were called negative and p1
20 would be 0.9. In the cases called negative by filmscreen,
21 if digital called 85 of those negative and 15 of those
22 positive, p2 would be 0.85. So, this would be at the hairy
23 edge of passing the FDA criteria in terms of p1 and p2 would
24 fail.
25 [Slide]

1 The criteria set in the guidance document were p1
2 greater than 0.9, with a test point of 0.95; and p2 greater
3 than 0.95, with a test point of 0.98. So, these are very
4 tight criteria set in the guidance document.
5 [Slide]
6 What I would like to look at is, if the guidance
7 document were followed in an equivalence study trying to
8 show that the same filmscreen mammograms were equivalent to
9 themselves, what the result would be in terms of p1 and p2.
10 [Slide]
11 Based on Elmore's results, the most likely
12 outcomes for p1 and p2 for one independent reader of each
13 film, and this would be where you don't use the same reader
14 so you are looking at interreader variability, would be 0.73
15 to 0.85. If you used the same reader, say, six months
16 later, this might go up to 0.91 according to Elmore's
17 results, but with different readers it would be in this
18 range.
19 If you went to the best two out of three readers,
20 which was suggested in the guidance document, the range of
21 p1 and p2 values would be 0.81 to 0.89. Since the FDA set
22 the requirement of p1 being 0.9 or greater and p2 being 0.95
23 or greater, even using the best two out of three readers as
24 the final decision for filmscreen and the final decision for
25 digital, even with just the reader variability taken into

1 account you are still outside the range set for p1 and p2 to
2 demonstrate equivalence.
3 [Slide]
4 So, the expected result of even trying to show
5 that screenfilm mammography is equivalent to itself is that
6 screenfilm mammography would not be equivalent to itself
7 using the guidance document p1 and p2 recommendations.
8 [Slide]
9 This takes into account reader variability. It
10 does not take into account the additional variability due to
11 patient repositioning, and doesn't take into account any
12 actual variability due to the modalities themselves.
13 [Slide]
14 In fact, any study of digital mammography that met
15 these guidance document requirements of p1 being greater
16 than 0.9 and p2 greater than 0.95 would be highly suspicious
17 in terms of case selection, interpretation, or data
18 analysis. It would only occur in the statistical fluke that
19 you had that special sample of cases where you defied
20 probability and got these numbers. In that case, you are
21 better off with a fewer number of cases to have that
22 unlikely coincidence.
23 [Slide]
24 I would like to raise some additional questions
25 for the panel, in addition to setting reasonable p1 and p2

1 values in an equivalence study or redefining the study
2 altogether. One of them, as mentioned by Dr. Bassett, is
3 the issue of the ACR BIRADS categories, which he very nicely
4 described.
5 [Slide]
6 There are two questions here. One is should ACR
7 BIRADS category 3 be positive or negative? It was stated
8 one way in the guidance document and the other way later in
9 advice from the FDA, verbal advice. So, there is some
10 question about that.
11 Then, there is the other question, which I think
12 is probably more important, of whether an ACR category 0
13 should be included. Since category 0 is recommending
14 additional follow-up, and since these are cases that are
15 read basically with just the filmscreen images or just the
16 digital images without additional follow-up, this is a
17 situation that would occur when you are reading these
18 basically as screening cases. It is very likely that you
19 would have cases read as needing additional follow-up, and
20 if you eliminate that category and force things into 1
21 through 5, you are not practicing radiology as it is done in
22 the clinical situation. So, excluding category 0 is, I
23 think, a problem.
24 [Slide]
25 There is also the additional question of

1 demonstrating not only hard copy printed digital equivalent
2 to filmscreen, but there will be the continuing question of
3 soft copy display, and is that equivalent to either
4 filmscreen or to the digital hard copy that was generated on
5 that particular case. This is going to be a continuing
6 problem as more companies come out with soft copy devices,
7 not necessarily attached to digital acquisition systems.
8 So, the panel needs to consider the question of
9 how to have a separate process to approve soft copy devices,
10 some with the manufacturer's blessing and some without.
11 One of the things -- and I don't have any
12 additional slides on this, but one of the approaches I would
13 like to suggest is going back to the approach of technical
14 evaluation of these systems and taking that more seriously.
15 Then falling back to what the original questions are about
16 digital mammography, that is, does it detect breast cancer
17 as well as filmscreen mammography? Is it missing breast
18 cancers? And, does it incur any higher degree of false
19 positives than filmscreen mammography? And, try to focus on
20 those being the simple questions.
21 In that situation, it may be reasonable to do a
22 technical evaluation, especially to address the issues of
23 additional artifacts that Dr. Wagner brought up which, in
24 our experience, don't seem to be a problem with digital if
25 it is properly calibrated.

1 Then, go on to the questions, the clinical
2 questions in possibly a completely different way, which is
3 to have a test set of digital films with their filmscreen
4 counterparts, possibly a reasonably small test set generated
5 by each company, say, 30 cases that were cancer, 30 cases
6 that were benign findings, 30 cases that were completely
7 normal -- I am sorry, 30 cases in total: 10 cases with
8 cancer, 10 cases with lesions that were found to be benign,
9 and 10 cases that are completely normal, with the filmscreen
10 and digital side by side, and to have some independent panel
11 of radiologists, with a large enough number of them, compare
12 side by side the digital to the filmscreen and say, in the
13 cases where there is cancer, is the cancer seen as well on
14 digital as it is on filmscreen? In the cases of normals,
15 are there additional false-positive findings on digital that
16 weren't seen on filmscreen? To have a large number of
17 radiologists evaluate these cases and do a side by side
18 comparison with reasonable statistical analysis about the
19 results, and come up with a decision about is this
20 equivalent in terms of seeing the cancers that were present?
21 Is it equivalent in terms of not incurring a greater number
22 of false positives? But to think in terms of a much simpler
23 study than the current guidance document equivalence study.
24 Thank you very much.
25 DR. ALAZRAKI: Thank you, Dr. Hendrick. The next

1 speaker will be David Schick, of Schick Technologies.

2 **Schick Technologies**

3 MR. SCHICK: Thank you for the opportunity to

4 present this to the panel.

5 [Slide]

6 The comments that I am going to make are going to

7 be very brief. Our goal is very specifically to place in

8 front of the panel a number of ideas which we are

9 considering in our preparations to begin clinical trials on

10 our device, and they are generated both by the company

11 itself and by some of the clinicians that we have been

12 working with. The goal is simply to place some fresh ideas

13 in front of the panel and allow you to consider them along

14 with the rest of the data which is before you now.

15 By the way, I am President of Schick Technologies,

16 and I am on the payroll of Schick Technologies.

17 [Laughter]

18 [Slide]

19 First of all, very briefly and by way of

20 background, we are a technology company. The products which

21 we make are based on radiographic imaging devices, and we

22 currently have two products which are on the market in the

23 biomedical industry. One of them is a dental imaging system

24 which allows intraoral radiographs to be taken by dentists.

25 The other is a bone densitometer for use in primary care

1 physician offices. If nothing else, we have shipped about
2 15,000 imaging systems to dentists and physicians. I guess,
3 for whatever it is worth, it is probably more than has been
4 shipped by other companies in those industries.
5 [Slide]
6 The technology which we are preparing for the
7 mammography industry is based on a tiled CMOS imaging array.
8 It has the benefit of being very low cost. It is compatible
9 with existing mammography machines. By that, I mean that it
10 fits the form factor of standard form cassettes and,
11 therefore, can be slid into the form drawer of mammography
12 machines. It has the benefits of solid state imaging. In
13 other words, our resolution in this prototype is 60 micron
14 pixels. One of our goals, as we have done in other
15 industries, is to make this a very available technology, and
16 our goal is to market it for less than $100,000 at least.
17 [Slide]
18 Just to go to the basics, our goal is to prove
19 substantial equivalence with filmscreen combinations.
20 Specifically, the first two bullets here are the primary
21 points: the ability to differentiate cancers in a screening
22 scenario, and the ability to detect microcalcifications.
23 Then, the third bullet is a proposed indication which has to
24 be validated, which is the ability to do more precise
25 screening in situations where there is denser breast tissue.

1 So, those are the basic goals.

2 [Slide]

3 As has been discussed exhaustively today, there is
4 an issue with the guidance which has emerged, which is that
5 there is more of a disagreement than an agreement between
6 radiologists in evaluating mammograms even with film, as we
7 have shown.

8 [Slide]

9 So, the goals of our suggestions have to do with
10 going back to ROC studies. The study which we are putting
11 up here is by Nab, and it was done with 150 mammograms.
12 They were both evaluated on film and then the films were
13 digitized. The purpose of the study was actually to show
14 that the correlation between the same radiographs with
15 different presentation methods, if you will. The results
16 were extremely positive, and I think that this is just an
17 indication that good old-fashioned ROC studies, if they can
18 be constructed in a way which is appropriate for the
19 industry, have the ability to show some significant clinical
20 results.

21 [Slide]

22 The problem, of course, is how do you have an a
23 priori knowledge of the truth? The only thing the panel has
24 to really decide is what is the meaning of truth -- sort of
25 a philosophical idea. What we are going to do is provide

1 two alternatives which differ what has been discussed so
2 far. As I said, you know, we are neophytes to this industry
3 and these are some ideas which we have been kicking around
4 and we would like to place them before the panel.
5 [Slide]
6 The first proposal has to do with using live
7 patients and women who have been recommended for
8 mastectomies. Unfortunately, there is a relatively generous
9 pool of such candidates, and the idea is to radiograph these
10 patients prior to mastectomies, both using a film and
11 digital modality, and then to confirm lesions by means of a
12 pathology report.
13 Just in terms of details, we recommend that the
14 breast be subdivided into categories, either in quadrants
15 or, as we have seen in some of the previous studies, into
16 regions, and then an ROC analysis can be performed based on
17 the actual presence or absence of lesions as proven by
18 pathology reports, and that can be correlated with readings
19 or interpretations of the radiologists.
20 The obvious advantages are that normal tissue is
21 not being needlessly irradiated, and that the positioning
22 variability can be minimized because the digital and analog
23 results can be taken sequentially. I suppose just as
24 obvious a problem with this study is that women who are
25 candidates for mastectomy typically have relatively obvious

1 lesions, and that may tend to bias the test. That also may
2 be dealt with by selecting the candidates for the test, or
3 also by selecting the subregion analysis so that there can
4 be a wide variety of different types of tissue even within
5 the same subject.
6 [Slide]
7 Then, the extensions of this study, and one of
8 them that I want to focus on in particular, are, first of
9 all, as I said inclusion of younger women with the hope of
10 analyzing the effect of digital mammography on dense breast
11 tissue. I think that is a reasonable thing. Unfortunately,
12 many women who are candidates for mastectomy are below the
13 age of 40. So, I think that is a practical goal of this
14 type of a study. Also, I think it is important to analyze
15 various types of pathologies.
16 I want to mention the idea of viewing modalities
17 and analyzing viewing modalities with this type of an
18 analysis. Since it is an ROC analysis with, hopefully, an a
19 priori knowledge of what true and false readings are, I
20 believe that it is possible to do this both on hard copy and
21 on soft copy.
22 One thing that I would like to say is that, as Dr.
23 Wagner pointed out, there is a lot of data in these digital
24 images. There is a huge dynamic range capable by means of
25 digital imaging, and it exceeds the human factor of viewing

1 ability typically. One thing that you can do with windowing
2 and other types of viewing techniques is to actually take
3 advantage of that. The manufacturers, speaking for myself,
4 I think are somewhat frustrated by the fact that the design
5 of the study initially at least does not allow us to take
6 full advantage of the digital systems. So, since this is an
7 ROC analysis where we know, hopefully, what the pathologies
8 are, it is possible to simultaneously study hard copy and
9 soft copy viewing results.
10 [Slide]
11 I would like to very quickly mention another
12 alternative which I think is practical, although it raises a
13 whole other range of issues, which is to utilize
14 anthromorphic breast phantoms. This is a technique which
15 has been used successfully in a number of settings.
16 [Slide]
17 I am just going to put up an image of what one of
18 these phantoms looks like. This is made by Nuclear
19 Associates. It is an available industry standard phantom.
20 [Slide]
21 What we are proposing is that either the industry,
22 by means of a panel, or a single company generate a series
23 of anthromorphic phantoms which simulate different types of
24 breast tissue, and overlays, which can be radiographed
25 simultaneously, which simulate various types of anomalies.

1 Of course, the major question here is, first of all, can
2 cancers be modeled using these types of modalities in any
3 type of a meaningful way?
4 Perhaps one of the answers to that is that they
5 can't be but perhaps the criteria of success can be more the
6 analysis of morphology than whether the shape itself is
7 benign or not benign. So, perhaps we can identify
8 calcifications certainly, spiculations and other types of
9 morphologies. That might be a more meaningful thing to do
10 via a set of anthromorphic phantoms rather than saying
11 whether something is benign or not benign, which obviously
12 is not really relevant to this type of study.
13 I think also there is a practical problem, which
14 is that for this study to be statistically significant, a
15 fair amount of these phantoms have to be designed, and the
16 design of those phantoms very, very much impacts on the
17 study. Because of that, I think it would be important to
18 have participation from, if not the panel of the FDA, at
19 least an industry panel which can determine the validity and
20 the applicability of the phantoms that have been used.
21 Nevertheless, the benefits are obvious. One such set of
22 phantoms can be generated, and can be used for all future
23 candidate systems, and there would be a uniform method of
24 determining the veracity of these systems. In addition, of
25 course, it eliminates the irradiation hazard to the patients

1 that are being tested. So, that is a possibility that we
2 would like to put before you.
3 [Slide]
4 To summarize, I guess, where we are coming from, I
5 think it is important for us to point out that we are not
6 answering any questions really. There are still some very
7 difficult questions for the panel to consider. Our goal is
8 simply to place before the panel some possibilities of
9 getting around the problem of not having a priori knowledge
10 of where the lesions are. I think that perhaps, if nothing
11 else, this will at least begin the discussion of other
12 perhaps more relevant ways of doing that, which I think
13 might be important for the adoption of standards in the
14 future.
15 Thank you very much.
16 DR. ALAZRAKI: Thank you, Mr. Schick. The next
17 speaker is Mr. Morgan Nields of Fischer Imaging.
18 **Fischer Imaging**
19 MR. NIELDS: Good morning.
20 [Slide]
21 We appreciate the opportunity to share some
22 comments on this difficult process of trying to receive
23 premarket clearance for digital mammography. Our company
24 has completed about a 600-patient comparative agreement
25 study, only to find now that the ground rules are going to

1 change as to what we maybe should have studied.

2 [Slide]

3 Notwithstanding that, we need to come up with some

4 solutions to the conundrum. Three and a half to four years

5 ago we argued here, at the FDA, in a preliminary meeting

6 that substantial equivalence could be established for

7 digital mammography when you can show that you image the

8 entire breast in the field of view; that you present image

9 quality which is comparable to the state-of-the-art

10 filmscreen mammography; and that you demonstrate that the

11 mean gonadal dose rate is equal to or less than filmscreen

12 mammography. We were unsuccessful then, and I suspect we

13 will be unsuccessful now in making this argument, although I

14 think elements of this should be considered more seriously.

15 [Slide]

16 The FDA did mandate an agreement study. We know

17 now, I believe, from the presentations we have heard this

18 morning, and from those of us involved in doing these

19 trials, that the p1 and p2 rates are not achievable given

20 these sources of variability amongst the readers. However,

21 there may be some alternatives to consider approval of the

22 technology by analysis of the variance or analysis of

23 disagreement. One would expect that in a study of 500

24 patients, according to the data that Dr. Hendrick presented

25 earlier, one can expect maybe disagreements in the area of

1 20-30 percent, which would indicate 100-150 cases may well
2 disagree on a BIRADS scale of 1-5.
3 One could take a look at this data first with
4 biopsy when a gold standard was available. There will be
5 biopsy results in these 100-some cases that will disagree.
6 Paired comparisons have been mentioned already, providing a
7 relatively objective view of lesion conspicuity, knowing
8 what the case was and then analyzing and comparing the
9 quality of the image in both modalities.
10 Naturally, an important point is the receptor
11 design. If, indeed, one can't see the lesion at all because
12 it is close to the chest wall, then that can be a problem
13 and that certainly can provide an analysis of the image
14 receptor design and consider if it is comparable or proximal
15 to filmscreen.
16 [Slide]
17 Other than that, I believe premarket clearance is
18 reasonable if one can show current MQSA image quality
19 standards, as was mentioned earlier on phantom studies, so
20 we have ACR phantoms for quality control; if we can show
21 through some of this clinical data that we can visualize
22 breast tissue as well as filmscreen; and that the dose is
23 qual to or lower. I believe that we will be able to
24 premarket clear these devices based on these criteria.
25 However, really to determine the role of digital

1 mammography will require a large study. A large screening
2 study will be required, as was referenced earlier in the
3 morning, to determine the ultimate sensitivity and
4 specificity.
5 [Slide]
6 At this point, I would go into an area that is not
7 necessarily the subject of the FDA's issue, but a study like
8 this can be helpful to HCFA and the other agencies who will
9 be faced with the very difficult task of reimbursing for the
10 technology. By reimbursing, I mean reimbursing more
11 because, as you have heard, this technology is significantly
12 more, three to four times more than the current cost of
13 filmscreening technology. The reimbursement amount, if one
14 can show that it has improved sensitivity, would depend on
15 the amount of improvement in sensitivity and specificity.
16 [Slide]
17 For example, if you look at approximate costs in
18 managing breast disease in the United States -- this is a
19 few years out of date, but a few years ago it was 8.5
20 billion dollars, of which about a third was screening and
21 diagnosis, and more than a third was biopsy. Obviously, if
22 one can improve the specificity of digital mammography,
23 there are significant savings in the biopsy area for the
24 payers. But in order to prove that, we must have a large
25 study. We cannot present a study of 500 patients when the

1 cancer rate is 5/1000 patients in the population. We cannot
2 show any type of analysis relative to reimbursement
3 requirements.
4 [Slide]
5 So, this is not new to the National Cancer
6 Institute. I have been lobbying for the NCI to sponsor such
7 a study in the area of digital mammography, and I believe
8 this type of study would serve two purposes. It would help
9 the FDA substitute for what one might call a postmarket
10 surveillance, which are words that send a chill up my spine
11 as an industry person, for a technology like this. It has
12 to be done. Somebody has to do it to determine the role of
13 this technology in medicine.
14 The 100,000 women randomized to film or digital
15 that I have shown here is not a biostatistically determined
16 number, but it is an approximation just to give you a scale.
17 We are not talking about 10,000 women to do a study like
18 this. A study like this would allow multiple technologies
19 to be tested because there are different approaches and
20 different technologies. We don't know the impact of those
21 technologies. And, it could be done relatively quickly.
22 Given the volume of mammograms, in one year 20 institutions
23 in a study could accrue 100,000 women in a randomized trial
24 like this, estimated to cost in the ballpark of 20 million
25 dollars. This is well beyond the resources of any one

1 company or any one trade group.

2 Then at the end of just one year, one could

3 determine in a comparative way, and glean some data that

4 would be very beneficial, for example, what is your

5 detection rate per 1000 women screened for each modality?

6 What is the positive predictive value of biopsy for each

7 modality? Have you, indeed, indicated and shown a lower or

8 higher false-positive rate? How many additional procedures

9 are required to achieve the diagnosis? What were the costs

10 to do this, and what was the patient tolerance?

11 Ultimately, these women could be followed

12 longitudinally for two to three years to determine the

13 ultimate sensitivity. This type of study would also have

14 the benefit of answering the question that is not well

15 answered, I don't think, and that is the role of mammography

16 in younger women. I think everybody has recognized that the

17 number of cancers in premenopausal women is substantially

18 lower. However, this is confounded by the dense breast

19 tissue of younger women typically encountered and,

20 therefore, digital mammography has a potential to attack

21 this problem. But really if we attack it without a

22 comparative trial to filmscreen, we won't really have a

23 basis to determine the ultimate role and how it should be

24 used.

25 [Slide]

1 There are some questions that the FDA has posed to
2 the panel, and since we have had an opportunity with an
3 agreement study, I would like to give at least our view of
4 this. My comments come from discussions with our
5 investigators, and there are about five investigators out
6 there using this technology.
7 First of all, we are measuring in our study hard
8 copy. This is the issue that was raised earlier, how do we
9 deal with soft copy? First of all, the hard copy that we
10 are presenting to the reader is limiting the amount of
11 information actually in the digital image because we are
12 only picking about 128 levels of grey out of 4000 levels to
13 select and present to the observer. There is much more
14 information in the digital image.
15 Further, soft copy manipulation, such as those
16 techniques on the slide, is likely. as we have noticed from
17 our investigators, to improve the presentation of the lesion
18 and, indeed, that is an area that I think our investigators
19 have testified to.
20 I will read from a couple of them who have
21 submitted letters, just last night, to the panel. The
22 panel, I believe, has copies of these letters. I would just
23 like to read a couple of excerpts. This is all anecdotal
24 and, obviously, we don't have a study, or we don't have
25 enough data.

1 From Dr. Pisano, who is Chief of Breast Imaging at
2 the University of North Carolina -- and I will just read a
3 couple of paragraphs. She writes, I am the radiologist
4 currently testing the Fischer SensoScan digital mammography
5 system. So far, over 250 women have been imaged at UNC as
6 part of the FDA trial.
7 My strong impression is that the digital
8 mammograms give at least as much information as the
9 filmscreen mammograms in every case I have reviewed. In
10 addition, if any effect in false positives is likely to
11 occur with digital mammography, it is to reduce them not to
12 increase them. This has been the experience of radiologists
13 testing all three types of equipment. I believe that this
14 will happen because image manipulation using soft copy will
15 allow radiologists to figure out whether a lesion is real
16 without calling a patient back for extra views in many
17 cases.
18 Also, I would just like to read some excerpts from
19 a letter written by Dr. Stephen Parker, from the Sally Jobe
20 Breast Center in Denver, Colorado. He writes, the Sally
21 Jobe Breast Center is currently contributing data to the
22 510(k) agreement study with the Fischer Imaging SensoScan
23 digital mammography system. To date, we have imaged
24 approximately 100 patients with the SensoScan whole breast
25 digital mammography system.

1 My overall impression is that the images reviewed
2 on soft copy display contain more information than is
3 present on the filmscreen mammograms. In fact, the image
4 quality is so clearly superior that I really don't
5 understand why an agreement study is being undertaken. I do
6 not believe that the information contained in the digital
7 images will increase false positives. On the contrary, i
8 believe that the additional information gleaned from digital
9 images will help increase the level of certainty regarding
10 any questionable finding. Therefore, unnecessary additional
11 filmscreen views, ultrasounds and biopsies can be avoided.
12 Some of these comments are reflected on the slide,
13 that essentially by using this technology most of the
14 radiologists can see more clearly, not be confused by the
15 technique.
16 I did include a quote in here because my point is
17 that digital mammography will not improve the skills of an
18 inexperienced mammographer. As Mike Lyles has said, a fool
19 with a tool is still a fool.
20 [Laughter]
21 [Slide]
22 There was another question posed to the panel, and
23 that is what about all these other ancillary capabilities
24 beyond improving image quality? My comment on this is that
25 just these ancillary features, such as digital image

1 storage, archive retrieval, telemammography and so forth,
2 will not convince radiologists to use digital mammography.
3 Only improved image quality in certain patients or in all
4 patients will influence radiologists. We are not looking at
5 approving a test such as a blood test where you put in a
6 slide with blood, and out comes an answer and you want to
7 look at the sensitivity and specificity. We are dealing
8 with a piece of equipment that is interactively used with
9 trained radiologists who make a diagnosis on breast imaging,
10 and it is a very difficult task, as best, as the variability
11 studies have shown.
12 The other point I would make is that FDA clearance
13 to market of a product doesn't substitute for knowledge.
14 There has been at least one PMA approved for breast
15 ultrasound and clearances for stereotactic biopsy devices.
16 Those clearances in themselves and those products in
17 themselves do not improve necessarily the results of
18 diagnosing breast cancer. Those are tools to be used by
19 experienced radiologists.
20 The last point I would make, and this is not to be
21 sardonic because I think it is real, that malpractice
22 attorneys really do set the standard because no reasonable
23 radiologist is going to use a technology and rely on it
24 without first establishing its performance in his or her
25 clinic. I think that is the case that we have seen. Just

1 because we have said that the FDA has cleared it for sale
2 does not mean that responsible radiologists are going to
3 adopt this without further study.
4 And, under MQSA, this is the only area in
5 radiology where radiologists must track and audit their
6 results, and very quickly with the approval of digital
7 mammography the positive predictive value of biopsy within
8 one's own practice can be analyzed, the call-back rates,
9 etc. All that data is reported as required by MQSA.
10 So my view would be let's find a way to make
11 certain the equipment is safe; the dose is lower; the image
12 quality is at least equal to filmscreen on phantoms; and get
13 the technology on the marketplace so we really can determine
14 at that point its benefits.
15 Thank you.
16 DR. ALAZRAKI: Thank you, Mr. Nields. The last
17 speaker for this session is Mr. Robert Uzenoff of Fuji
18 Medical Systems.
19 **Fuji Medical Systems**
20 MR. UZENOFF: I am Robert Uzenoff, Executive
21 Assistant to the President for Corporate Development of Fuji
22 Medical Systems.
23 The discussion today has been informative and
24 underlines the difficulty in demonstrating agreement between
25 screenfilm and digital mammography. This morning, FDA's

1 guidance for digital mammography has been questioned because
2 of difficulty in demonstrating agreement between the two
3 mammographic technologies. Now we are revisiting the
4 appropriateness of any one approach to demonstrating
5 substantial equivalence for digital mammography.
6 [Slide]
7 Fuji has three points that I will make this
8 morning. First, FDA's guidance, dated June 19, 1996,
9 entitled "Information for Manufacturers Seeking Marketing
10 Clearance for Digital Mammography Systems," sets forth only
11 suggestions of the type of evidence needed to demonstrate
12 substantial equivalence of digital mammography to
13 screenfilm. It did not set forth binding requirements for
14 510(k) submitters.
15 Second, the FDA's inclination to require clinical
16 data for digital mammographic equipment was rational as it
17 relates to the first rendition of the companies' digital
18 mammography devices.
19 Third, the achievement of a successful premarket
20 agreement study eliminates the need for a postmarket study
21 or surveillance requirement.
22 [Slide]
23 I want to go back and talk some more about the
24 role of guidance documents. We have heard persons in
25 industry express concern that adhering even loosely to FDA's

1 guidance would condemn studies to failure. Fuji was never
2 an enamored of the guidance document, particularly
3 Attachment A that accompanied it. We thought the guidance
4 was too general and incomplete to provide a satisfactory
5 template for developing a successful agreement study.
6 Accordingly, we assembled a team of experts in
7 study design and protocol development, including well-known
8 biostatisticians. We charged them to review and critique
9 the guidance and propose a protocol for our consideration.
10 It is important here and elsewhere to appreciate
11 that FDA guidance documents are not legally binding on
12 anyone. We have heard, for instance, just this morning
13 mention of 0.9 and 0.95 as requirements when, in fact, that
14 was merely a suggestion of one approach that a sponsor might
15 consider. Indeed, the guidance document clearly states in
16 bold letters that this document is intended to provide
17 guidance in the preparation of a regulatory submission. It
18 does not bind FDA or the regulated industry in any manner.
19 Fuji appreciates that agency reviewers, despite
20 this admonition, at times treat a guidance as immutable,
21 thus, pushing regulated persons to often uncritically follow
22 it. How many times have 510(k) sponsors heard that their
23 submissions are incomplete because they didn't follow the
24 agency's guidance?
25 Obviously, to the extent a guidance describes an

1 efficient path to the marketplace following it is the best
2 thing to do. However, the guidance document offers
3 suggestions and, given the complexity of this particular
4 substantial equivalence evaluation, it could do no more than
5 that. The lesson learned for everyone is that guidance is
6 not a regulatory requirement.
7 FDA's interest in clinical data, as I will
8 discuss, was rational, yet its attachment A protocol failed
9 to provide industry with a useful approach. What it did do,
10 however, was to put issues on the table for persons
11 interested in submitted a 510(k) for digital mammography to
12 consider. In this respect, the guidance was helpful.
13 Guidance is now needed to further suggest to potential
14 510(k) submitters what kind of concern they must take into
15 consideration when designing investigations that fit their
16 imaging technologies.
17 It is important for this panel and the FDA to
18 appreciate the limits and functions of guidance in an area
19 as complex and difficult as demonstrating substantial
20 equivalence between digital and screenfilm mammography.
21 Fuji believes that the best guidance here will reaffirm the
22 rationale for a premarket clinical study and identify issues
23 that should be confronted in designing one.
24 [Slide]
25 I spoke before about the appropriateness of a

1 premarket clinical investigation. A clinical study to
2 demonstrate that digital mammography is equal to or better
3 than screenfilm mammography is reasonable because of the
4 expectation that digital technology will supplant screenfilm
5 in time. Current screenfilm mammography, as FDA states in
6 its guidance document, quote, has been proven to detect
7 breast lesions that are indicators of early breast cancer
8 resulting in a reduction in breast cancer mortality.
9 Because of this specific and improvement purpose,
10 and the documented performance of screenfilm devices,
11 supplanting them with other clinical studies showing at
12 least agreement between digital and analog technologies
13 would undermine the applicability to digital devices of the
14 bulk of data supporting the effectiveness of the screenfilm
15 devices. In other words, admitting a clinical agreement
16 study in the premarket phase would leave unanswered
17 questions about digital mammography that should never exist.
18 Additionally, a sound agreement study will
19 benchmark each of the various digital mammography
20 technologies brought to market, thus, creating a means to
21 correlate clinical findings to physical measurements of
22 imaging performance. After the clinical performance of each
23 manufacturer's first digital mammography is established,
24 subsequent enhanced versions of each such device could then
25 proceed to market based on bench test data alone which

1 showed substantial equivalence.

2 I have spoken here of digital imaging

3 technologies. It is important to recognize that there are

4 several different technical approaches to the implementation

5 of digital mammography. Each is likely to have imaging

6 properties related to spatial filtering, fill factor, or

7 some other factor. There is at present no validating bench

8 testing method for estimating the impact on clinical

9 performance of these unique samples of imaging properties.

10 Each technology must, therefore, have an independent

11 clinical evaluation. This complex problem is obviously not

12 amenable to a cookbook or one-size-fits-all approach.

13 In Fuji's efforts to design a protocol to

14 demonstrate substantial equivalence between Fuji computer

15 radiography and screenfilm mammography, we first examined

16 the particulars of the guidance document. Our consultants

17 quickly recognized fundamental problems with FDA's criteria

18 for substantial equivalence, where it defined agreement as

19 the total agreement of digital and analog screening

20 demonstrating more than a minimal level of both positive --

21 that is, the probability of digital positive given analog

22 positive of 0.9 and negative, that is, the probability of

23 negative digital given negative analog positive greater than

24 0.95. That was the criteria for agreement.

25 Unfortunately, this definition of substantial

1 equivalence would condemn screenfilm technology to failure.
2 Specifically, the medical literature plainly shows, and we
3 have heard others speak of this earlier this morning, that
4 because of interreader variability screenfilm compared to
5 itself would not demonstrate agreement under FDA's guidance
6 criteria. Estimates from credible studies of the agreement
7 between two radiologist reading the same screenfilm
8 mammogram range between 60-80 percent. As a result, this
9 interreader variability must be accounted for in the design
10 of any study seeking to establish agreement between
11 screenfilm and any other mammographic technology.
12 Because of the nature of the guidance document, it
13 is incumbent upon each sponsor to develop a definition of
14 substantial equivalence that is appropriate to its
15 technology. In fact, FDA's current guidance states that
16 each sponsor must justify his definition of clinical
17 agreement. This is an important concept because without it
18 a proponent of digital mammographic devices cannot
19 successfully demonstrate substantial equivalence to
20 screenfilm.
21 Without question, alternatives to FDA's suggested
22 protocol for agreement study are available for
23 consideration. In addition to the Obuchowski approach that
24 we heard of this morning from Greg Campbell, for example, at
25 this year's SPI Medical Imaging Symposium researchers from

1 the Department of Radiology at the University of
2 Pennsylvania presented papers on two new approaches to the
3 problem of comparing different imaging technologies for the
4 same task. Kundel and Polansky discussed comparing observer
5 performance with distribution analysis when there is no
6 external gold standard. Chakraborty, Kundel and others
7 discussed the differential receiver operating
8 characteristic, or DROC, method.
9 After all has been said and done, Fuji believes
10 that the agency was not unreasonable in calling for
11 premarket clinical agreement studies.
12 Number two, reasonable criteria for substantial
13 equivalence can be satisfied in a well-designed study of
14 practicable scale.
15 Number three, the failure of some to demonstrate
16 agreement between screenfilm and digital mammographic
17 technologies was a problem of hypothesis selection and study
18 design, and not inherent in the concept of an agreement
19 study.
20 [Slide]
21 The third area I want to speak to you about this
22 morning is the area of postmarket studies. We have become
23 aware of rumbling or comments even this morning in the area
24 pointing towards postmarket studies. When this panel met on
25 March 6, 1995, among other things the panel considered the

1 wisdom of a postmarket study or surveillance requirement.
2 The panel did not reach a consensus on such a requirement.
3 Subsequent to the panel meeting, FDA issued a draft guidance
4 document for comment. The draft document included a
5 postmarket screening study to further compare digital and
6 screenfilm mammography. The agency, after receiving
7 comments, chose not to impose a postmarket study component
8 in its guidance document.
9 Fuji believes that FDA's determination then is
10 appropriate today. Simply put, the proper designing of a
11 premarket clinical study does not justify the imposition of
12 a postmarket study requirement. If a properly designed
13 premarket study is done, then a postmarket study will be
14 superfluous.
15 In other words, a premarket clinical study that
16 demonstrates agreement between digital and screenfilm
17 mammography to an acceptable level of statistical
18 significance negates the need for a postmarket study to show
19 the correlation of comparability between the two
20 technologies. Such a requirement would only create costs
21 without benefit.
22 Fuji believes that a demonstration of substantial
23 equivalence between digital and screenfilm mammography will
24 permit the fair imputation of the clinical and histological
25 data applicable to analog imaging to digital mammography,

1 thus, obviating any need to create a postmarket burden on
2 digital mammography manufacturers.
3 [Slide]
4 In closing, I strongly suggest that this panel
5 continue to recommend a properly designed clinical agreement
6 study to demonstrate substantial equivalence of each
7 sponsor's technology. It is important to understand that an
8 agreement study is not necessarily as constrained a
9 definition as we have heard some take it this morning, taken
10 literally from an example in Attachment A of the 1996
11 guidance.
12 Also, I respectfully suggest the FDA's guidance
13 document be revised to, among other things, eliminate the
14 Attachment A protocol, and to include discussion of
15 interreader variability associated with screenfilm
16 mammography; and the critical need for both a hypothesis and
17 a design which are sensitive to this issue.
18 Last, the statistically strong design of an
19 agreement study is critical to success. From the
20 information or which we have become aware, a weak design
21 will be unforgiving in a mammography agreement study.
22 Thank you.
23 DR. ALAZRAKI: Thank you very much, Mr. Uzenoff.
24 With those presentations, we conclude the industry
25 presentation section of the meeting. We will now break for

1 lunch. We will reconvene at 12:30, leaving only 40 minutes
2 for lunch, because of the time constraints and what we have
3 to cover in the afternoon, and the limitations on
4 individuals making airline connections. So, we will
5 reconvene at 12:30.
6 [Whereupon, at 11:50 a.m., the proceedings were
7 recessed, to reconvene at 12:30 p.m.]

1 AFTERNOON PROCEEDINGS

2 **Panel Member Discussion**

3 DR. ALAZRAKI: I would like to call the meeting
4 back to order.

5 At this point, I would like to invite Dr. Susan
6 Alpert to address the panel once again and then she will be
7 followed by Dr. Dan Schultz of the Division.

8 DR. ALPERT: I just wanted to speak to the issues
9 that brought us here today and that we want you to focus on.
10 That includes making just a couple of definitions that I
11 think will help in your deliberations.

12 The first is that our mandate is to determine the
13 safety and effectiveness of the devices that go to market.
14 There are two mechanisms for doing that. One is to evaluate
15 the safety and effectiveness from scratch for each product.
16 The other is an abbreviated way of evaluating new product, a
17 new model, a new company's product, a changed product for
18 devices of a type where we have already determined that they
19 are safety and effective, that we understand them, and that
20 we can use an abbreviated approach, a substantial
21 equivalence approach to put those products on the
22 marketplace.

23 But substantial equivalence has more words: it is
24 substantially equivalent in safety and effectiveness. So,
25 it is not simply a matter of comparing technologies. It

1 involves understanding the safety and effectiveness of the
2 differences in those technologies. In other words, if you
3 are going to do an abbreviated, you have to be in the same
4 realm of safety and effectiveness to go to market through
5 substantial equivalence. I wanted to be sure that was
6 clear.
7 The second thing, as was pointed out this morning
8 by the last speaker, guidance documents are just that, they
9 are a guidance. They are to provide the thoughts and
10 understanding of the agency both to other people in the
11 agency, our reviewers, as well as to submitters. They are
12 not binding, and they are not mandated. They are simply
13 what it says, they are guidance. They are an opportunity to
14 follow a process with the hopes that if you follow that
15 process you will have established either the safety and
16 effectiveness or the substantial equivalence of your
17 product, and it will easily move it to the marketplace. But
18 it is not a mandate.
19 The guidance that was developed and that we all
20 clearly recognize was not perfect, that we developed four
21 years ago and we admit that is why we are here, was
22 developed actually at the request of industry who were
23 concerned three years ago that absent a clear path for
24 substantial equivalence there would be no opportunity for
25 digital to enter the marketplace. I found it very

1 informative this morning to hear the companies raise the
2 concern that equivalence may not be quite what they had
3 expected. I thought that was interesting.
4 I just wanted to, again, remind the panel of what
5 substantial equivalence and safety and effectiveness were;
6 what guidance documents do; and the fact that what we are
7 trying to do here is, in fact, reassess the guidance
8 document -- we have all acknowledged it is not where it
9 needs to be -- and provide more information, the information
10 we have learned over the last three years, to make this a
11 more useful guidance. Thank you.
12 DR. ALAZRAKI: Thank you. Dr. Schultz?
13 DR. SCHULTZ: Good afternoon. I hope you have had
14 time to digest your food as well as all the information that
15 was presented this morning.
16 [Slide]
17 I am a general surgeon and I promise you that I
18 will not provide you with any new information. My goal, as
19 a surgeon, is just to help you, hopefully, use the
20 information that you have already gotten more effectively.
21 So, with that, I am going to run through these slides and
22 perhaps some of them will be left up for you to use in your
23 discussion. This afternoon, we want to hear your ideas. We
24 have paid a lot of money to get you here, and we are paying
25 you a lot --

1 [Laughter]

2 -- and this morning all you had to do was just sit

3 and listen, and that is not good enough! This afternoon we

4 are going to make you work.

5 [Slide]

6 We had this meeting for one very important reason.

7 We believe that breast cancer and mammography are an

8 extremely important problem that affects a lot of people,

9 specifically the women in the United States and, therefore,

10 we wanted to have an open public meeting to get the

11 perspectives of the clinical community, industry, NIH and,

12 finally, the regulatory people altogether to make this

13 happen.

14 [Slide]

15 As Dr. Alpert just said, the role of FDA is to

16 establish the safety and effectiveness. Whether you call it

17 safety and effectiveness or you call it substantial

18 equivalence, the idea is that individual medical devices

19 need to be assessed prior to going to market to make sure

20 that they will work and that they will not hurt people. The

21 ultimate determination of the impact of this technology on

22 medical practice is beyond the scope of what we are going to

23 be doing here today, beyond the scope of FDA, and will

24 require many, many years and a lot of other work by a lot of

25 other individuals in order to finally make that

1 determination.

2 [Slide]

3 Again, just to remind you, this is not a PMA

4 evaluation but just to remind you want we mean by safety,

5 which is that safety as defined by the law is that based on

6 valid scientific evidence and the intended use of the

7 product, the benefits of using the device outweigh the

8 risks.

9 [Slide]

10 Effectiveness, again in terms of why do we care

11 about clinical trials, the definition of effectiveness tells

12 us that in a significant portion of the target population a

13 device will provide clinically significant results.

14 [Slide]

15 There are several premises that I want to just

16 outline prior to putting up the questions. The full

17 potential of digital mammography, and I think this point was

18 made well this morning so I am not going to belabor it but

19 the full potential of digital mammography can only be

20 evaluated by determining the sensitivity and specificity of

21 soft copy digital in a screening population with cancer/no

22 cancer as the endpoint. I think that that is fairly well

23 recognized.

24 This will require a large-scale studies involving

25 thousands of patients in order to get sufficient numbers of

1 patients with cancer and sufficient follow-up to determine
2 malignancy in patients who are not biopsied.
3 [Slide]
4 Soft copy isn't ready. I think that we understand
5 that and I think that what we don't want to do is wait until
6 everything is all in place. We want to move ahead. We feel
7 that digital acquisition, which can be evaluated using hard
8 copy display, offers inherent advantages over filmscreen,
9 and we ought to be able to understand what those advantages
10 are, and be able to take advantage of them as soon as
11 possible and, therefore, a staged approach to market is
12 justified.
13 [Slide]
14 In the meantime, using all the resources, and you
15 are part of our resources, we need to define the preclinical
16 and clinical data requirements, and clinical study designs
17 for digital acquisition which will (a) meet the regulatory
18 standard for valid scientific evidence and, (b) are within
19 the capabilities of individual companies to perform.
20 [Slide]
21 We should look at digital acquisition with hard
22 copy display first. We should look at ways to show that
23 digital will not lower the level of diagnostic accuracy for
24 mammography that clinicians, industry and government have
25 worked so hard to achieve and American women have come to

1 expect.

2 [Slide]

3 With respect to agreement studies -- I am going to
4 leave this slide up perhaps. I will go through this quickly
5 and then we will put this one back up. There were several
6 questions raised with respect to agreement studies which you
7 will be discussing this afternoon. One of them had to do
8 with variability, reader variability, and the other had to
9 do with how best to use the BIRADS Lexicon, and specifically
10 whether BIRADS-3 should be considered positive or negative,
11 or perhaps there may be other options which you we be
12 willing to discuss.

13 [Slide]

14 Again, there was talk about noninferiority
15 studies. If we are going to talk about noninferiority
16 studies we need to understand from you, the clinical
17 community, what is a clinically appropriate delta for
18 sensitivity and specificity and, again, is there an
19 acceptable silver standard alternative to histology, i.e., a
20 panel of experts, in order to perform these studies?

21 [Slide]

22 Again, this is sort of open-ended but we would
23 like you to use all of your imagination to help us here and
24 tell us if you think that there are other options that
25 nobody has brought up yet with respect to clinical trial

1 design and all of the statistical issues that were raised
2 this morning.
3 [Slide]
4 Finally, and I put time permitting because I think
5 the previous questions will probably keep you busy for the
6 next couple of hours, but time permitting, I think it would
7 help us to at least address the issue of soft copy
8 manipulation, discuss briefly how you would anticipate soft
9 copy work stations; what kind of requirements would be
10 necessary to bring soft copy work stations into use; and to
11 discuss the issues of hard copy to soft copy, how are they
12 different; how to evaluate those differences; and what kind
13 of follow-up we should have in terms of assessing those
14 questions.
15 With that, maybe we will go back and put up the
16 slide on agreement studies, and I will let you do your job.
17 Thank you.
18 DR. ALAZRAKI: Thank you, Dr. Schultz, and thank
19 you, Dr. Alpert for bringing us back to our charge for this
20 afternoon. Dr. Edward Sickles, Professor of Radiology at
21 University of California School of Medicine, San Francisco,
22 has agreed to take the lead for the panel discussions, and I
23 am very happy to turn this job over to Dr. Sickles.
24 [Laughter]
25 DR. SICKLES: Thank you. You should understand

1 that I come to this with some experience, having served in
2 the same position many years ago as Chair of the Radiologic
3 Devices Panel. So, I am quite familiar with the process
4 that is involved.
5 I think probably a good start, before we get into
6 specifics, would just be to invite any of the panel members
7 who have specific questions to ask them now so that we can
8 get everybody up to speed. Is anybody here lost completely
9 who wants to be updated, or is confused by something that
10 was said before? Don't be shy. It is better to say it now
11 than to be completely even further lost.
12 DR. GARRA: I am not lost on the technical
13 aspects, but I am a little lost as to the policy background
14 behind all this. Clearly, if, say, a film company changed
15 their changed their film it could introduce all the factors
16 that we were discussing today. It could change the
17 sensitivity. It could change the way people interpret it.
18 It could obscure pathology potentially. All those factors
19 could be introduced. Similarly, if a designer of a view box
20 enlarged the view box by 200 percent you could have the same
21 effects. And, I don't believe that the FDA has asked to
22 conduct equivalence trials on view boxes or films. I could
23 be wrong. That is the first question.
24 The second one is why was a policy decision made
25 to select digital mammography for additional scrutiny, and

1 why was it selected versus the other things that all figure
2 into changing the way we interpret the studies?
3 DR. ALPERT: In answer to your first question, as
4 I mentioned earlier, when a determination has been made
5 about the basics of a technology, of a claim in the
6 technology that supports that claim, and it is put in a
7 classification -- in this case, in Class II which allows for
8 substantial equivalence submissions to come in, we always
9 have the option in evaluating changes to the technology to
10 ask the following questions.
11 The first question that gets asked is does the
12 change impact safety and effectiveness? If the change
13 impacts safety and effectiveness, a submission of a new
14 510(k) is required. When we review the 510(k) we look at
15 several aspects. One aspect is are there new questions, new
16 types of questions that we need to ask relevant to the
17 change in technology that we need to understand, that we
18 need to have answers to in order to determine that this
19 product ought to go to market? If there are new types of
20 questions, then it is no longer a 510(k) and a new type of
21 question puts you into Class III. So, we always ask that
22 question, are there new kinds of questions that this change
23 raises?
24 The second thing that we always ask is does the
25 change impact safety and effectiveness? Does it remain

1 substantially equivalent in safety and effectiveness
2 secondary to the change? We have had several cases even
3 recently where the same type of product, not in the
4 radiology area but where a product that shares the same
5 claim as other products in the area naturally is very much
6 the same product. It was a vein graft. It had not reached
7 the threshold of being the same in safety and effectiveness.
8 This happens. And, when it does, again, the product no
9 longer can come through the 510(k) process, the substantial
10 equivalence process. It needs to establish on its own what
11 its safety and effectiveness is.
12 So, as the changes that you describe come up in
13 display or in acquisition and film, those questions are, in
14 fact, asked, and they are asked in the individual 510(k).
15 First they are asked by the company. The company addresses
16 it and decides whether it is a big enough change to impact
17 safety or effectiveness. Then it is asked, if a 510(k)
18 comes in, by our reviewers in their evaluation of the
19 technology. So, those questions do get asked.
20 I think what is underlying your question is the
21 fact that there are big changes that happen that way, and we
22 don't know what the impact of all of those changes going on
23 at the same time are one by one on the total impact on
24 screening mammography. You are absolutely right. But we
25 made a determination that going from filmscreen to digital,

1 changing the entire impact on screening, raises the concern
2 for us because we are then needing to understand -- because
3 screening is a one-shot deal for most patients. As was said
4 this morning, these are patients without signs or symptoms.
5 They come in; they get an evaluation and, based on that
6 evaluation you put them in a bin and determine what is going
7 to happen to them next. The significance of making that
8 decision, and the significance of the change raised the
9 question to us that we needed to be very careful as that
10 change was made.
11 We came to the table three years ago saying we
12 think this may be a PMA issue, and we were asked to
13 reconsider, by the industry, what steps could be taken to
14 establish substantial equivalence and move into the
15 marketplace. And, that is what resulted in the guidance
16 document.
17 DR. SICKLES: Just to amplify a little bit, the
18 practice of mammography has changed very substantially over
19 the last 20 or 30 years. In fact, there have been several
20 evolutionary, rather than revolutionary, changes in imaging
21 over that period of time. This is the only one that has
22 actually come to FDA regulation. But initially mammography
23 was done with direct exposure film. Then xerography came
24 along. That was not regulated although it was a similar
25 change. Then screenfilm came along. That was not regulated

1 although it was a similar change. The FDA, I think with
2 reason, three years ago or three and a half years ago --
3 nothing was known about digital; it was this big, black box
4 -- concluded that this could be a big difference. That is
5 why they set up this whole process. And, I think that was a
6 reasonable decision.
7 DR. ALPERT: I would only point out two other
8 things. One is that we are still very new at the law.
9 Device regulation only began in 1976. It actually began in
10 a more aggressive way because the first years were spent in
11 trying to figure out what was going on in the current
12 marketplace because of the way the law was written, to
13 classify the things that we already had out there, and then
14 figure out how to manage change. The program has evolved in
15 twenty-three and a half years. Our decisions today are not
16 necessarily the same as the decisions that were made in the
17 late '70s and early '80s. So, the medicine has changed as
18 well.
19 Secondly, the practice issues -- and I wanted to
20 pick up on the wording of Dr. Sickles comment because the
21 practice issues are not on the table. The issues of how
22 within the practice of radiology, specifically mammography,
23 the training and the work, and so on, is not part of the
24 premarketing evaluation of the technology. I know it is
25 somewhat difficult to separate because the mammography

1 Quality Standards Act is regulated in the center but it is
2 not a premarket device program. It is quite a separate
3 program looking at practice but that is very separable for
4 us, and they are managed separately.
5 DR. SICKLES: There was another question here.
6 DR. SMATHERS: The current discussion answered it.
7 Thank you.
8 DR. SICKLES: All right. Yes?
9 DR. ALAZRAKI: I have another comment relevant to
10 what Dr. Garra brought up, and that is that in today's
11 practice we have the issue of HCFA and reimbursement, and I
12 think we are entering into a new technology which may
13 increase the costs of performing mammographic exams, and if
14 there is any change in the charges for those costs I think
15 HCFA or Medicare would not be paying any more unless FDA had
16 given its blessings and endorsement of that new technology.
17 DR. GARRA: Probably not.
18 DR. ALPERT: That is a good point. Our clearance
19 or approval of a technology is really only one step, and the
20 reimbursement community looks at other measures as well. It
21 is not a given that even if we clear or approve something
22 with a claim it will change the reimbursement patterns.
23 DR. SICKLES: If there are no other sort of basic
24 questions, I think the next thing I would like to do is
25 invite any of the panel members or invited guests who wish

1 to make comments on the presentation so far today to do so
2 at this time. I know some of us are quite knowledgeable
3 about mammography and maybe we can share our perspectives
4 about what we have heard and how we filter that information
5 -- just as a starting point. Anybody want to begin? I
6 would rather be last because I don't want to influence
7 people too much with my own opinions. Yes?
8 DR. ALAZRAKI: I have one other point which is
9 relevant. That is, one of the presenters brought up the
10 issue of phantom studies. I think it was Mr. Schick. While
11 there is no substitute for clinical effectiveness studies in
12 the clinical environment, and I don't know the practicality
13 and the difficulties of the phantom studies, but I think it
14 is important in something like digital as a way to show the
15 real superiority of digital over conventional mammography.
16 Even if the clinical environment just demonstrates
17 equivalency, the companies still have the phantom data to
18 show the superiority.
19 DR. SICKLES: I think we have to be really careful
20 with phantoms in two ways. First of all, phantoms are
21 clinically relevant only if they reproduce the clinical
22 situation. Although they are anthropomorphic phantoms that
23 are supposed to look like breasts, they don't reproduce the
24 range of parenchymal patterns which we see in clinical
25 mammography. That doesn't mean they could not be produced.

1 They could be, but that is something that does not now
2 exist. That would have to be created.
3 Secondly, the overlays that he talked about, which
4 we be reproducing disease states which would be superimposed
5 on a phantom, don't exist. They would have to be created.
6 Then they would have to be validated. In order to put that
7 into -- this is my opinion, in order to patient that into an
8 FDA guidance document would be putting the industry back at
9 least as long as it takes to create that infrastructure, and
10 that could be years. I would say that is a wonderful gold
11 and companies who want to develop these things, fine; go
12 right ahead and do it, but my recommendation to the FDA is
13 don't wait for that because if you do, you are going to be
14 putting everybody back by years.
15 DR. ALPERT: Those are very helpful comments. One
16 of the things that I neglected to say is that actually three
17 years ago we also discussed whether there were phantoms, and
18 we do also support and encourage the development of such
19 phantoms as a tool that can be used to get us a far piece in
20 the evaluation of the technologies. But you are absolutely
21 right. Validated phantoms that are understood and agreed to
22 by a wide group don't exist, at least not yet. And, we also
23 encourage the development of good phantoms as a means of
24 getting a long way in evaluating this kind of technology.
25 DR. GARRA: I would like to make a comment

1 relevant to that. I am very concerned about the lack of
2 phantoms. Although the FDA would like a phantom, I think we
3 have a problem in that there is a difference in liking one
4 and really going after one. There hasn't been any concerted
5 effort to develop these phantoms. If three years ago there
6 had been a concerted effort with major dollars poured into
7 it we might have that phantom today. It is always the issue
8 of we don't have one; it will take too long to validate it.
9 Then it stops. Then what happens is the patients get
10 exposed to double doses of radiation because we have to do
11 clinical studies again.
12 I think it may be too late here for digital
13 mammography but there has to be serious consideration given
14 to having an ongoing effort to develop the phantoms in
15 advance so that this thing can be done. It can't be done
16 cheaply. Perhaps a joint effort between the NIH and the
17 Food and Drug Administration to have significant funding to
18 direct it in this direction would alleviate this problem in
19 the future because it is something that has to be ongoing
20 because it is always too late to get it validated in time.
21 Then we end up doing the clinical studies which are very
22 costly. The FDA may not see that cost but the manufacturers
23 certainly do. Perhaps since it relates so directly to
24 public safety, it should be something that the government
25 does invest in monetarily.

1 DR. MONSEES: Before we get into a discussion
2 about what equivalency trials might consist of, I would like
3 to reiterate a point that was made by several speakers this
4 morning, and that reflects the say that we practice breast
5 imaging. Screening is a very different process from looking
6 at and evaluating which patients need to go to biopsy. I
7 think perhaps some of the failure of some of the different
8 studies that have been described this morning pertaining to
9 reader variability come from the fact that readers may have
10 been asked to judge whether somebody is going to biopsy
11 based on the screening films themselves rather than on the
12 workup films.
13 I think we need to think about breaking down the
14 trials into two parts. One would be the detection type of
15 trial, or some sort of equivalency for detection. It
16 doesn't necessarily have to be a full-scale screening trial.
17 Then, the second part would be whether we make the judgment
18 to go to biopsy based on workup films.
19 One of the reasons that that is important is
20 because mammographers are highly criticized for over-
21 biopsying. We want to make sure that we don't make that any
22 worse than the situation that currently exists, and
23 hopefully we would make that better.
24 So, I think I can't see any way around doing this
25 in two separate phases, and I just wanted to bring that to

1 the table before people started describing ways to perhaps
2 design trials.
3 DR. SICKLES: That is an important point. Whoever
4 is taking notes, this one goes right up there at the top.
5 In assessing whatever type of study that you are doing, I
6 think you should be making a distinction between whether the
7 technology, whether digital mammography is being used in the
8 screening setting or whether it is being used in the
9 diagnostic or workup setting. In the screening setting the
10 exercise is perception, and is the woman sent back for
11 routine screening in a year or does she need more tests?
12 That is all that needs to be done. That is the clinical
13 assessment. You don't have BIRADS. You just have recall/
14 no recall. Nothing more. That is the clinically relevant
15 endpoint.
16 In the diagnostic setting, since you raise the
17 point, the important endpoint here is biopsy/no biopsy;
18 tissue diagnosis/no tissue diagnosis. There are various
19 forms of biopsy. If you want to use BIRADS, then that would
20 be BIRADS 4 and 5 is positive, and BIRADS 1, 2 3 is
21 negative. BIRADS 0 does not exist because in the workup
22 setting there should be no need for additional imaging
23 assessment. You should complete it.
24 And, I think most of the breast imagers on this
25 panel would agree. I would hope we would agree that that is

1 a good way to consider the two approaches. Now, from the
2 FDA point of view, whether you are interested in two
3 separate assessments or you are more interested in the
4 screening assessment because it affects the population in
5 general rather than just those women who are abnormal, that
6 is a different issue. But to merge these two uses into a
7 single assessment will only confound the analysis of
8 results.
9 DR. ALPERT: As we pointed out earlier, the idea
10 here is screening mammography, and I think that that is
11 really our first focus, our main focus. We are interested
12 in the other, of course, because the devices are going to
13 claim all of the above and not just screening. But
14 screening is, in fact, the major focus..
15 I would ask a question for clarification, which is
16 one of the reasons why we are sitting with you at this point
17 in time, and that is, would you feel the same for soft
18 reading as for hard reading in what was just discussed for
19 the kinds of evaluations that ought to be made? Because it
20 seemed to me, and I am raising it as a question, from the
21 discussion this morning and from my understanding of soft
22 reading that you can obviate the call-back. So it is a bit
23 of this and a bit of that. It is not just the initial
24 screen anymore. It also is the enhancing, and focusing, and
25 so forth. So, I raise the question. If you want to make a

1 distinction between hard film and soft reading for digital?
2 DR. SICKLES: I am probably the only one here who
3 can answer your question based on experience because I don't
4 think anybody else here works with full-field digital
5 mammography, and has even worked with soft copy images. But
6 I can answer that for you because I have a lot of
7 experience. Over the last year we have done over 300 cases
8 and we have interpreted them soft copy. We have had access
9 to hard copy and pretty much rejected it because it is not
10 as good because of the way it limits one's interpretation.
11 But you should understand that there are strengths
12 -- not just limitations but also strengths to hard copy
13 imaging. Hard copy imaging allows you spatial resolution
14 which you don't have with current monitors, display
15 monitors. Although the display monitors give you that full
16 dynamic range that you heard about, they don't give you the
17 same level of resolution to see tiny structures.
18 The question is, and I think the FDA will have to
19 have two separate assessments, if you are first going to
20 look at hard copy and say a hard copy works at that one
21 window and level which was designated for that particular
22 case, that doesn't necessarily mean soft copy will work.
23 Soft copy probably will be better, but only -- only if the
24 user is careful enough to know how to work with it, and only
25 if the display is sufficient to show with moderately good

1 detail the very fine structures.

2 I mean, I can tell you from my own experience that
3 properly used, with an advanced work station, you can see on
4 soft copy full-field digital mammograms more calcifications
5 than you can see on a corresponding film. Now, that is
6 something I don't think we could have said three years ago.
7 Three years ago we would have said probably less but now I
8 think most people who have worked with this technology will
9 tell you that without a doubt you see more.

10 DR. ALPERT: Two things. One is thank you for the
11 comment on the fact that we probably are dealing with two
12 different things. The second is that I hear you echoing
13 what we stated as a premise this morning, and that was
14 echoed by some of the manufacturers but not all, that soft
15 copy is not ready yet; that we are really not talking soft
16 copy advance ready for every mammographer yet.

17 DR. SICKLES: That is a matter of opinion, and I
18 think in general soft copy is not ready in terms of
19 commercial product. That is a different statement than soft
20 copy is not ready for assessment by the FDA. I believe that
21 any of the manufacturers, if they chose to -- remember, they
22 were influenced by your previous guidance document which I
23 think emphasized the hard copy, but I think any of the
24 manufacturers, if they chose to, could mount a study now
25 with soft copy that would convince you that the technology

1 is ready for soft copy. That doesn't mean that users would
2 want to work with the work stations in terms of throughput
3 and other issues which haven't been overcome. But in a
4 study design where the people who are looking at the images
5 are willing to take ten minutes to look at a mammogram
6 instead of one minute, they can produce very good results.
7 DR. ALPERT: Thank you for the clarification.
8 That is very helpful for us.
9 DR. SICKLES: Good. Yes?
10 DR. MALCOLM: As a non-mammographer, what I have
11 heard you saying is we are talking about bias -- we are
12 talking about screening, and what you are saying with the
13 soft copy you are moving beyond a screening position -- that
14 is what I am hearing because you have a screening hard copy
15 mammogram, and if you compare that to the digital image you
16 are basically tuning it and looking at that as if you had
17 taken a patient, for instance, with a hard copy mammogram
18 and brought them back for additional studies and in a sense
19 you are in-house working a patient up. Am I listening to
20 that incorrectly?
21 DR. SICKLES: Yes and no. You get different
22 information at soft copy than you get at hard copy. At hard
23 copy you choose one window and level setting in which you
24 are going to look at that picture. It is a static picture
25 just like film imaging is. If you are clever in creating

1 the window and the level for which the hard copy is created,
2 you will have captured most of the information you want to
3 have there. You will never get it all on a hard copy image
4 because you lose the flexibility of tweaking it.
5 On the other hand, you get very fine resolution.
6 Digital mammograms are acquired with a lot more information
7 in the signal than can be displayed at soft copy. There is
8 more information there than you can display at soft copy
9 with conventional monitors, and monitors either would have
10 to be so huge that they are just not practical or even
11 available, or you would have to look at segments of the
12 breast, filling the monitor in order to display it at full
13 resolution. That is not practical.
14 That is the advantage of hard copy because you can
15 see it at full resolution in one image, and it is also
16 easier for the radiologist to look at because we are used to
17 looking at hard copy films. Soft copy is a stretch from
18 that. I mean, any radiologist looking at any digital image
19 soft copy has a big mental adjustment to make. It is a big
20 change, being able to tweak the image rather than just
21 looking at that one picture.
22 But that can be done at screening, and it can be
23 done at diagnostic. There is no reason why at screening
24 when you look at the image and you say, you know, I think
25 there may be some calcifications there and then you window

1 and level and you can bring them out. It has nothing to do
2 with workup. From my experience already with full-field
3 digital, I doubt that you will be able to completely
4 eliminate the workup mammography, the diagnostic
5 mammography, just by tweaking the screening image. If that
6 was your question. I think that is sort of where you were
7 going.
8 DR. MALCOLM: Yes. I think it is still kind of a
9 grey line.
10 DR. SACKS: I think maybe the way to resolve this
11 is that the use of soft copy for screening will obviate a
12 tremendous amount of call-backs -- well, maybe not
13 tremendous but it will allow you to do things that otherwise
14 would require call-backs. That is what I am trying to say.
15 Therefore, there will be some call-backs that will be
16 obviated.
17 DR. SICKLES: It will allow you to not call back
18 for certain things. It will not allow you to obviate call-
19 backs for some summation shadows, summation artifacts.
20 There is a whole bunch of call-backs that it won't obviate.
21 The second thing that you should understand, and
22 let's put this in perspective of other incremental changes
23 in mammography technology, is that I think we can expect,
24 despite what you have heard from the audience that in theory
25 this should decrease call-backs, I think that what you can

1 expect from past experience is that with any new technology
2 where people are learning how to use it is that you will
3 have increased call-backs in the beginning, and there will
4 be a learning curve. In the first two to three years of
5 general practice people will learn how to work with it, and
6 whatever transient increase in call-backs happens just
7 because you are starting to look at pictures in a different
8 way, we will learn how to deal with that and it will come
9 right back down to where it was before.
10 That leads me to another point, and that point
11 which I think is very important for the FDA to hear is that
12 in your assessment, in your premarket assessment sensitivity
13 is far, far more important than specificity --way far more
14 important than specificity. The bottom line of that is that
15 when you consider the difference between agreement trials
16 and what is called noninferiority trials, the noninferiority
17 trial in fact has an advantage. It is not just the
18 advantage that you heard from the presentation of Mr. Bird,
19 which is that it allows you to study cases with a better
20 endpoint of truth and take into understanding that digital
21 may already be better than film, but it also shifts the
22 biases, just by study design, to make it harder to prove
23 sensitivity if you design it the way you have proposed.
24 That is where you should be shifting the burden.
25 You should be shifting the burden to sensitivity and away

1 from specificity. I think you can expect specificity to go
2 up anyway. I think that is going to happen in general
3 practice, although it may not in the trials because people
4 in the trials will be trained to first work with the
5 interfaces and learn how to use it, and there will be more
6 expert type people who are less likely to show these
7 variances. But in the general population I think you can
8 expect it to go up because it did with xerox when we shifted
9 from direct exposure film to xerox, and it happened when we
10 shifted from xerox to screenfilm, and it will probably
11 happen here. And, I don't think that is bad. I think it is
12 to be expected, and I don't think you should be trying to
13 overregulate that.
14 DR. WAGNER: You said specificity would go up. I
15 think you meant to say that it would go down.
16 DR. SICKLES: Correct. I meant to say it would go
17 down.
18 DR. ALPERT: If I may, just one quick comment.
19 The issue that you raise about the focus that you were
20 suggesting we have on the initial market entry trial then
21 raises the question about how one can, in fact, evaluate
22 that what happened with changeovers in other technologies
23 actually happens with digital. I would like to leave a
24 place marker on the table regarding postmarket information
25 and necessary understanding of the impact. I am not stating

1 who should have control over that or how it would be done,
2 but I want to place a marker there, and your comment that an
3 expectation of what would happen in the marketplace would
4 need data at some point in time.
5 DR. SICKLES: By the way, what I gave you comments
6 on are anecdotal experience. Nobody documented ever in a
7 controlled way that specificity was affected when we shifted
8 technologies. We all believe it occurred but we don't know.
9 DR. ALPERT: And we, hopefully, will have the
10 opportunity to find out this time.
11 DR. DESTOUET: Dr. Sickles said that this was
12 evolutionary, but I think that if you deal with the general
13 radiologists out there, someone who does high volume
14 screening, this is really revolutionary because radiologists
15 in the past have looked at hard copy images and if you are
16 dealing in a high volume screening situation you can
17 actually look at 200, 300 cases in an 8-10 hour day.
18 Now, when we start talking about not just a
19 different method of acquiring images but a different way to
20 display that image, you really turn that whole scenario
21 upside down. And, I think the FDA has to recognize two very
22 different issues. I would suggest that what we need to
23 focus on today is something we did not know about in 1995,
24 and that is, it would seem that full-breast digital images
25 are equivalent to analog images, and is there a way, without

1 overburdening the industry, that we can show that with some
2 well-designed trials? In many cases that data may already
3 be there. Industry may already have the answer to that.
4 Then perhaps, maybe not at this sitting but at a
5 later date, we could deal with the display, whether it is
6 hard copy or a soft copy display. But I would say that
7 this, indeed, is revolutionary for the radiologists in
8 practice, not just because of what you pointed out, Dr.
9 Malcolm, that we may go from just a plain screening
10 situation to a situation where we can, indeed, decrease the
11 call-back rate by doing the image manipulation you talked
12 about, increasing our sensitivity and not really changing
13 our specificity and actually doing a better job for the
14 patient. But it is really going to change the way we
15 practice.
16 DR. SICKLES: I just have a question for you. The
17 revolutionary part of what Dr. Destouet is talking about
18 really relates to the way we look at the images, namely,
19 soft copy display. Does the FDA require this type of
20 assessment in looking at soft copy chest x-ray versus hard
21 copy chest x-ray digital, or soft copy anything versus hard
22 copy anything in other branches of imaging?
23 DR. ALPERT: No. But, again, we are dealing with
24 a unique environment with screening mammography, as we have
25 all talked about. That is what keeps coming back to a

1 difference between a screening situation and a diagnostic
2 tool. We have looked at them differently, but one of the
3 comments that was made earlier is that there are other
4 technologies that have been introduced in breast imaging as
5 a global, going from the screen all the way to whether or
6 not someone goes to biopsy, as was discussed in the panel
7 meeting yesterday, and we have asked the manufacturers of
8 those devices making a claim to have an impact on whether or
9 not someone goes to biopsy, to do a full-fledged safety and
10 effectiveness trial and demonstrate the impact on making
11 those decisions to not lose the sensitivity of the current
12 practice. So, we have done that.
13 So, when we get into whether we are making a
14 determination of whether biopsy is being recommended based
15 on a given modality, we have had situations where we have
16 asked for full-fledged trials. So, we haven't worked all of
17 that out yet. We are asking in part for your advice on how
18 to proceed in that area. That is where we would like to go.
19 We want to focus on the screening aspect for most of this
20 discussion.
21 MS. WHELAN: Is it impractical in a screening
22 environment to produce more than one hard copy of a digital
23 mammography?
24 DR. SICKLES: The answer is no.
25 MS. WHELAN: If you are manipulating the image and

1 you see several views that are of interest --

2 DR. SICKLES: Yes, there is no reason why one

3 could not, if one chose to, make eight different window and

4 level settings of hard copy and assess those as opposed to

5 just one. As a practical matter, the people who would be

6 looking at the images would then be taking eight times as

7 long to look at the pictures, with a diminishing level of

8 return of benefit.

9 The reason why I happen to believe -- having

10 worked with both hard copy very little and soft copy a lot

11 is that it is just much quicker to work with soft copy

12 because you see the dynamic changes as you adjust the image

13 in real time. But, yes, you could do just what you asked;

14 it just would be very time consuming and awkward.

15 MS. WHELAN: Yes, I was thinking of it in terms of

16 Dr. Malcolm's question, where he is not a mammographer but

17 would want to make an assessment for his patient based on

18 something that comes from an imaging studio, lab, or

19 whatever.

20 DR. SICKLES: I doubt, frankly, that he would want

21 to look at those images. I think much more likely, in a

22 given setting, you know, with a given parenchymal pattern of

23 the breast, one might do very well to have two different

24 hard copy images to fully image the parenchyma. Some women

25 have mostly dense breasts and some women have mostly fatty

1 breasts, and they would be imaged quite nicely with one
2 window and level setting, but there are other women who have
3 mixed patterns in a very focal distribution where it
4 probably would be very useful, even in a screening setting
5 if you were stuck with a hard copy, to have two images per
6 projection to look at, one at a dense window setting and one
7 at a fatty window setting.
8 DR. DESTOUET: I think you also have to recognize
9 that in the short term, as we change from one technology to
10 the other, that we will have to compare apples and apples,
11 otherwise it becomes very difficult to determine if there is
12 a significant change in a woman's breast. So, initially the
13 radiologists may have to deal with a hard copy display of a
14 digital image and compare that to a film. Then eventually,
15 as the technology improves, and the work stations improve,
16 and we learn how to use this to best advantage we may be
17 able to go fully to soft copy. But I don't anticipate that
18 will happen within the next year or two.
19 DR. ALAZRAKI: Along those lines, as you know,
20 many radiology departments are going to become filmless
21 departments, and that is happening. Now, I don't know
22 whether mammography is a hold-out here --
23 DR. SICKLES: It is; it is. It has been stated
24 that mammography will be the last --
25 DR. ALAZRAKI: The last?

1 DR. SICKLES: -- the last aspect of a department
2 to go filmless. With good reason.
3 DR. ALAZRAKI: Well, despite the fact that
4 mammography may be a hold-out, I expect it is going to go
5 the route of all --
6 DR. SICKLES: I agree with you.
7 DR. ALAZRAKI: -- and that may happen even sooner
8 than you might think. I don't think the FDA is going to be
9 evaluating all of those that go -- I mean, it has already
10 happened.
11 DR. SICKLES: My recommendation to the FDA is that
12 although you may expect to see initial applications with
13 hard copy, don't preclude or even discourage applications
14 that come to you with soft copy because I think the
15 manufacturers could produce that for you right now if they
16 wanted to.
17 DR. ALPERT: We have not discouraged --
18 DR. SICKLES: I know you haven't.
19 DR. ALPERT: I also want to make the comment that
20 we haven't discouraged anybody from establishing
21 superiority. It was implied this morning that we forced
22 people into agreement --
23 DR. SICKLES: Yes, I know.
24 DR. ALPERT: -- and our goal was to get
25 superiority tested three years ago --

1 DR. SICKLES: Well, my recommendation to the
2 manufacturers would be that you should be testing soft copy
3 right from the outset because that is what people will
4 eventually be using.
5 DR. ALPERT: I would like to comment on Dr.
6 Alazraki's comment. In general, the digital systems that
7 are out for diagnosis have general claims for their utility
8 and are well within their clearances and their introduction
9 in use in current radiology suites. In most cases they
10 don't carry a specific claim of the sort that we are dealing
11 with when we deal in mammography.
12 DR. MONSEES: I have a comment, but before I go to
13 that I would just say that if we do go to an all digital
14 department, it is more than likely that the mammograms won't
15 be read on the same stations that are used to read other
16 types of studies.
17 DR. SICKLES: True.
18 DR. MONSEES: There will be their own stations
19 that will probably be higher resolution.
20 My comment was just to revisit the terms that we
21 use pertaining to screening because I just wanted to make it
22 clear that I believe that screening has two phases. The
23 first is the detection phase, but screening is not done at
24 that phase. Then, to be making a terrible pun about the
25 detection phase, it is a digital decision, either 0 or 1;

1 you either call them back or you don't. But then part two
2 is to decide whether to biopsy. I think that is still part
3 of screening. I think that any screening program worth its
4 salt goes through that second phase and looks at its outcome
5 data, not just who gets called back and who doesn't with
6 that initial digital decision to get called back or not.
7 So, you know, if we are talking about screening,
8 are we talking about part A and part B of screening, or only
9 part A, only detection?
10 DR. SICKLES: That is another important point.
11 Although the interpretation and, therefore, the endpoint
12 that the manufacturer would demonstrate to you in an
13 observer performance study is call back/no call back, the
14 bottom line has to be truth and truth is cancer/no cancer.
15 Yes?
16 DR. TOLEDANO: I am going to expand a little bit
17 on Dr. Monsees' point and the point of everyone that spoke,
18 that our clinical trials need to be relevant to what is
19 going to happen in clinical practice, and for this whole
20 issue of whether we are going to be looking at a screening
21 setting, or the first part or the second part, or the
22 diagnostic setting we have to make sure that we design a
23 trial that can answer the appropriate question.
24 I would like to just remind the panel of another
25 very important fact of life that we need to include when we

1 are designing our trials, and that is the sources of
2 variability that Dr. Wagner mentioned, the intra-observer
3 and inter-observer variability. In the original guidance
4 document it was proposed that doing an agreement study one
5 should minimize that using the majority rule, using the
6 dichotomization of BIRADS categories. But if you do that
7 you can no longer estimate the impact of those
8 variabilities, and you can no longer tease out what is
9 happening with the agreement.
10 So, I think it is important that we design studies
11 that will enable us to tease out what is happening with
12 inter-observer variability because we will have that in
13 clinical practice, because we want to know that in clinical
14 practice; and what is happening with our accuracy in terms
15 of some truth standard, some gold standard that we
16 eventually find out cancer/no cancer.
17 DR. SACKS: Dr. Sickles, I am just looking for
18 clarification. It seems to me that I have heard two
19 definitions of call-back. One definition is a call-back
20 when you have a BIRADS 0, when you call back just to clarify
21 some technical anomaly on a film, versus the second type
22 which is call-back where the film is positive and you want
23 to do subsequent workup. Would you clarify that?
24 DR. SICKLES: Yes. Maybe others will amplify on
25 this. One can legitimately at screening make an

1 interpretation that it is already abnormal enough to fit
2 into BIRADS 4 or even BIRADS 5. This happens. In practice,
3 most screening exams that are recalled for additional
4 imaging are in BIRADS 0 because a substantial percentage of
5 those never get tissue diagnosis. They either are found to
6 be normal, or they go into surveillance, BIRADS 3, after
7 recall. So, both things happen.
8 There is still another type of recall, which we
9 call retake, which is a technology problem on the image.
10 That is likely to be minimized in a digital setting because
11 of the ability to tweak the image. If it is too light or
12 too dark you just change the window and level settings and
13 you probably have a diagnostic image anyway. On the other
14 hand, if you don't have enough breast on the film, or if the
15 patient moved during the exposure so that you can't read it,
16 then that type of retake would still need to be done.
17 So, there is a whole variety of reasons why you
18 may need another picture, but in clinical practice the most
19 common one is just to get more images to know whether you
20 even have something important.
21 DR. DESTOUET: I agree. Generally the woman comes
22 back for magnification views or ultrasound to determine
23 whether it is a cystic or solid lesion. So, it is really a
24 BIRADS code 0 for additional workup and then you determine
25 where she fits into the 1-5 BIRADS category.

1 DR. ALPERT: I would like to comment on Dr.
2 Toledano's approach and comment on the guidance. Our
3 attempt to control the variability in the trials is that our
4 goal in the initial trials is to evaluate the differences
5 between the technologies and to try to hold as constant as
6 we can all the other things that may impact. We recognize
7 that in most clinical trials to assess the initial
8 technology we are not getting a total real-world view of
9 what will happen when that technology is introduced into
10 clinical practice. We are attempting to evaluate the new
11 thing, and to understand it, and in order for us to tease
12 out in a reasonable sized trial, we have the tension between
13 mimicking what will happen in the real world and getting the
14 kind of data that will allow us to tease out the
15 contribution of the changing technology or the new
16 technology from standard practice, and being able to
17 understand the fact that as we get into the next phase, the
18 introduction into actual clinical practice, the issue of
19 what will be happening overall.
20 It is a problem we have with all medical devices.
21 Most of the things we look at have a treatment by user
22 interface. We do evaluate that in our clinical trials.
23 Greg is better at explaining exactly how we do that kind of
24 control and assessment, but it is a great difficult for us
25 to fight that tension of wanting to do both at the same time

1 and figuring out which one is really our necessity, and our
2 necessity is understanding the new technology or the new
3 device, and trying to leave some of that question for the
4 next stage because trying to do them all together is huge.
5 DR. TOLEDANO: Can I respond?
6 DR. SICKLES: Yes.
7 DR. TOLEDANO: You don't need a huge trial. I
8 just wanted to clarify that. Some of my colleagues said,
9 "well, wait; you are advocating a huge trial." You don't
10 need a huge trial. If you design your trial appropriately
11 you do not need 10,000 or 100,000 patients, and I am by no
12 means advocating a huge trial because no one would ever be
13 able to market the technology.
14 What I am saying is we need to sit back and think
15 about how we design our trial very, very carefully so that
16 we can minimize the number of patients, minimize the number
17 of exposures, and still obtain trial endpoints and
18 information from the trial that will be useful.
19 DR. SICKLES: If there aren't any other comments,
20 what I would like to do is try to focus in on some of the
21 specific questions that we have been asked to address.
22 The first one, does anybody on the panel have any
23 recommendation, serious recommendation to the point where
24 this technology should be assessed as being new, as opposed
25 to the substantial equivalency standard that you heard

1 earlier? It is my understanding that we can just go through
2 the substantial equivalency route and tell them to forget
3 about the other thing.
4 DR. ALPERT: Actually, that is not the real
5 question that is on the table at the moment.
6 DR. SICKLES: Okay.
7 DR. ALPERT: The decision about whether something
8 makes it through a substantial equivalence determination or
9 needs a PMA, as I said, is a decision we make when we see
10 the data and we evaluate it.
11 What we are asking is to make an assumption that
12 we can look at a comparison between the technologies and ask
13 ourselves the question can we get a valid assessment of this
14 new technology taking that approach. I think that is a more
15 useful way to look at it.
16 DR. SICKLES: Posed that way, does anyone on the
17 panel wish to answer that question in the negative? Do you
18 think it is reasonable to --
19 DR. GARRA: Can you repeat that --
20 DR. SICKLES: Okay, we will do it again. I have
21 to rephrase it. Can we recommend to the FDA that an
22 appropriately designed substantial equivalency set of data
23 be acceptable?
24 DR. ALPERT: Could be developed.
25 DR. SICKLES: Could be developed. Could somebody

1 come to them with data that shows substantial equivalency
2 and be meaningful? My answer would be yes. Does anybody
3 say otherwise?
4 [No response]
5 Well, you have a unanimous opinion there, unless
6 you want to refine it by asking us more questions.
7 DR. SACKS: Yes, I want to refine it by asking a
8 question. Substantial equivalency with respect to what is
9 something that we obviously have to decide on here.
10 DR. SICKLES: Well, you told us safety and
11 effectiveness.
12 DR. SACKS: Okay, and that is further defined in a
13 screening setting as no loss in sensitivity?
14 DR. SICKLES: I would define that in a screening
15 setting as no loss at all in sensitivity -- I wouldn't hole
16 them to that, that is an absolute --
17 DR. SACKS: In other words, that is where we talk
18 about the noninferiority studies where we set a delta. And,
19 one of the questions that we posed to the panel is what is a
20 reasonable delta for the sensitivity?
21 DR. SICKLES: I haven't gotten to that.
22 DR. SACKS: I just want to raise a question in
23 response to Dr. Toledano's point, and that is, in fact,
24 relevant to the exchange here that if we try to isolate the
25 technology and somehow control for all the variability that

1 will occur in clinical practice, have we, in fact, done
2 anything worth doing? I mean, this is a question. In other
3 words, given the variability among readers in actual
4 clinical practice, are we better off evaluating some
5 difference between the technologies or are we really trying
6 to evaluate the difference between one technology reader
7 interaction and result versus the other technology reader
8 interaction and result? In other words, the technology
9 reader is a unit. You don't get a mammography report out of
10 the technology. So, that is the setting that I think we
11 have to look at, and it isn't immediately obvious to me that
12 to have isolated the technology and shown that B is better
13 than A would necessarily imply that B and reader is better
14 than A and reader.
15 DR. TOLEDANO: Absolutely, and that is why if you
16 do a trial where you can figure out what is happening, what
17 the difference is between technology, for today's
18 discussion, of digital and analog, and then how large the
19 intrareader variance is, and how large the inter-observer
20 variance is. Perhaps we can even gain insight on how to
21 decrease those variances. If you design your trial
22 carefully you don't need 100,000 women; you don't need
23 10,000 women going for double exposure; but you can find out
24 is it going worth after. If you do the trial so you can
25 find out those components, then you know. If you do the

1 trial so that you are averaging in a majority rule, or a
2 consensus rule, or the highest number you will never know.
3 DR. ALPERT: I believe I have been a little bit
4 misunderstood. Can I clarify? Our job is to tease out, as
5 I said, what the impact of the technology is and not to
6 understand what will happen it gets into the clinic. What I
7 think both Dr. Toledano and Dr. Sacks are talking about is
8 how you figure out what is going on, not whether you figure
9 it out. So, I just want to make sure -- we are not talking
10 about different things; it is a matter of the semantics
11 used. The goal is to understand the impact of this changing
12 technology. As Bill points out, and I think Dr. Toledano as
13 well, there is an interface that we need to understand in
14 order to understand the contribution of the change that the
15 technology may make. I mean, our goal is to be able to
16 label a product so that the user understands what they get,
17 as well as how to use it, and understands what impact it may
18 have. It is not to measure that ultimate impact but to give
19 them an understanding of what we have seen in evaluating it
20 so they make a decision about choosing or not choosing and
21 understanding how to use it.
22 DR. SICKLES: I would make another suggestion to
23 the FDA. In addition to the designs that you have seen
24 proposed -- you have agreement studies which have problems,
25 especially if digital is really better than film. You have

1 the noninferiority studies, and at least one manufacturer
2 seems to feel that they can already meet reasonable criteria
3 for, if I heard the presentation correctly. But there are
4 still other approaches, other options.
5 Think about this one, ask a manufacturer to
6 collect cases in a screening setting with reasonably subtle
7 mammographic findings, not obvious findings; obvious
8 findings you will see with everything, but reasonably subtle
9 findings that are typically found at mammography screening,
10 such as clustered calcifications, masses, areas of
11 architectural distortion -- those are the big three. Ask
12 manufacturers to collect cases where these findings are
13 present and present cases where they are present at analog
14 mammography, and then also provide digital mammography
15 examples of the cases. Then submit those cases to an expert
16 panel of viewers -- this is a different design than you have
17 heard before and I would be interested how you would view
18 this as being valid or not. But blinded in terms of knowing
19 which manufacturer it is coming from so the manufacturer has
20 no control over who is going to be reading them. It could
21 even be a panel of expert readers that the FDA set up
22 independently. Then the expert readers simply looked at
23 these cases and agreed, first of all on the basis of the
24 analog mammograms that they sere subtle cases so what the
25 manufacturer produced doesn't necessarily have to be

1 accepted. But, you know, assuming that they were subtle
2 cases, just have the panel look at the analog cases, look at
3 the digital cases, and see whether the findings are seen
4 with equivalent visibility.
5 If you had calcification cases, mass cases,
6 architectural distortion cases -- and I would put two other
7 things in there besides those, which I will get to in a
8 minute -- and you had a panel of experts agree that you can
9 see the same findings on all these things, then I think that
10 is one way to demonstrate substantial equivalency. It
11 doesn't require any of the bells and whistles that you had.
12 Another thing is that it overcomes problems with inter- and
13 intra-observer error, and still allows you to get an
14 endpoint that the consuming public could feel comfortable
15 with. You know, if you can see all the findings with which
16 breast cancer presents, and you can see them equivalently
17 when you look at the images side by side, then that ought to
18 be equivalent.
19 I would add to those three that I gave you there,
20 the masses, calcifications and architectural distortion. I
21 would add two more things. The first thing I would do is I
22 would add breasts of unusually high density because this is
23 a problem, and they can be normal cases or they can be
24 abnormal cases. The second thing that I would add would be
25 lesions that are close to the chest wall. One of the

1 potential problems with digital imaging systems is that the
2 receptor is of a different physical nature, in some cases
3 thicker, than the standard film cassette. There can at
4 least potentially be problems in including the deepest parts
5 of the breast on the image with these devices. If
6 manufacturers had to include lesions that were very close to
7 the chest wall, they would have to show that these lesions
8 could be imaged at screening with digital.
9 DR. ALPERT: Let me ask a couple of clarifying
10 questions. One, I am not sure I understood how one would
11 create such a set. So, my question is was your suggestion
12 that each manufacturer select amongst the studies that they
13 have or subjects whose evaluations they have from an analog
14 evaluation difficult cases and then image them digitally
15 with the setting that they would recommend for their
16 product? Therefore, you would have both mammogram [sic] and
17 digital on a set of cases? Or, were you suggesting that
18 there be patients identified who have certain kinds of
19 presentations that might be identified by a routine
20 mammographer saying, well, I saw three of these last month,"
21 and bring those people for side by side evaluations? I am
22 not sure I really understood how you would identify the
23 population, and then what recommendation you were making
24 about how to have them read. I really wasn't clear.
25 DR. SICKLES: You could do it either way,

1 actually, but I suspect the manufacturers already have these
2 cases. I suspect they have them in what they have already
3 collected. They just have to pick the most subtle ones out
4 of whether they have.
5 The only problem with this approach is a selection
6 bias problem. That is, you know, when they look at a case
7 and they say, "aha, this is a subtle case but we didn't see
8 it on digital; we won't send that one in."
9 [Laughter]
10 That is a potential problem. You have to have a
11 way around that. You can develop design mechanisms to avoid
12 that situation. Yes?
13 DR. BEAM: I would like to raise another concern,
14 if not a problem, with your design, only because I think it
15 raises a bigger question. I would have a hard time
16 estimating sensitivity or specificity.
17 DR. SICKLES: I don't think you have to.
18 DR. BEAM: Okay. That is my point. Really this
19 is a level one technology efficacy kind of study. The
20 question that I would like to raise is whether or not the
21 FDA could focus on a level one proof and not go to level
22 two, which is looking at sensitivity and specificity.
23 DR. SICKLES: From my point of view, and I would
24 be interested in other people's around the table, if you
25 want to show is what we were told is substantial equivalence

1 and safety and effectiveness, and you pick subtle cases of
2 the way in which breast cancer presents and you don't have
3 an essentially selection bias sample, you can show that you
4 see the same things. I think that is enough. I don't think
5 you need to do sensitivity and specificity. By eliminating
6 that need you also eliminate all the problems that we have
7 heard about reader variability.
8 DR. BEAM: We have heard safety and efficacy
9 couched in terms of sensitivity and specificity.
10 DR. GARRA: I would like to make a comment on
11 that. I agree with Dr. Sickles, and I think that this
12 suggestion came up this morning. Dr. Hendrick mentioned
13 using subtle cases. You said that you were interested
14 primarily in looking at the technical aspects, and does the
15 technology produce images that are acceptable and high
16 quality. I am not sure that the FDA has a mandate to
17 determine how accurately radiologists can read various types
18 of studies. That falls within our province as doctors, to
19 learn how to use these technologies.
20 I like the idea of that type of study because it
21 focuses on the visibility of abnormalities, and it doesn't
22 focus on our interpretation of the visibility of
23 abnormalities.
24 DR. ALPERT: I think the issue of what you get for
25 that kind of study is very important. I think that there is

1 information though I cannot tell you that it is sufficient
2 for us to make a determination that the safety and
3 effectiveness of the two products will be the same. It
4 answers some of the questions.
5 I am interested that you commented that we are
6 only interested in the technology.
7 DR. GARRA: I didn't say that; I said primarily.
8 DR. ALPERT: Okay. We focus on technology and
9 then its impact on safety and effectiveness. Our mandate is
10 safety and effectiveness, not technology. Our mandate is to
11 figure out if things are safe and effective. Once we have
12 determined that something is safety and effective we can use
13 a technology assessment as an abbreviated way of expecting
14 that new technology will have the same safety and
15 effectiveness as something we have already evaluated. That
16 is what the 510(k) is intended to do.
17 I am not sure we are quite there yet. One of the
18 issues is how do we understand what that technology
19 assessment will give us and when it will reach the threshold
20 of establishing substantial equivalence and safety and
21 effectiveness. I think it is a piece of the puzzle. It is
22 like what you would get out of a phantom, if you will. Then
23 the question is "and," and we are struggling with that
24 "and," to understand what kind of comparisons, what kind of
25 information and what we will know from such a study.

1 That is why we are still asking questions. I know
2 Greg will have some and we shall probably have some
3 questions as well. How do we then understand how it is
4 substantially equivalent and safety and effectiveness based
5 on that assessment which is focused on certain aspects of
6 what the technology can provide? I think it is a question
7 we have to ask. I haven't heard all the information that I
8 think we would need. I can't say that would be sufficient.
9 DR. TOLEDANO: I have a question for Dr. Sickles.
10 So, you get your panel together and they look at that set of
11 cases, and you expect that there is going to be no intra-
12 observer or inter-observer variability?
13 DR. SICKLES: No, I didn't say that.
14 DR. TOLEDANO: How would we deal with that?
15 DR. SICKLES: I think what you do is you assemble
16 a panel of a sufficient number of individuals, not just one
17 but a sufficient number of individuals so that when there is
18 disagreement you can work through that disagreement. You
19 know better than I that there are methods to achieve that.
20 I am certainly not experienced in that kind of design but I
21 know that such designs exist, and I am sure that they could
22 be applied to this approach.
23 I would remind you that the selection bias problem
24 with this type of study is also shared by any other study,
25 short of a randomized, controlled trial, because the

1 manufacturers, even with a noninferiority study, could
2 choose to exclude a case where you didn't see it on the
3 digital and you did see it on the analog and just never show
4 it to you and never show it to their observers. So, the
5 same potential problem exists no matter what you accept.
6 DR. CAMPBELL: I guess I like the idea very much
7 of looking at a mix of cases and having a predominant number
8 perhaps that are very subtle. But I would probably suggest
9 -- and maybe this is based on your vast experience, but some
10 assurance that, you know, the very simple cases are also
11 easily picked out. One way to do that is to create a case
12 mix which you know about, which has maybe a higher
13 proportion of subtle ones for one or the other technology
14 but also include other kinds of cases in there as well, and
15 keep track of what the proportions of those are. Then, if
16 you can find the truth out you can evaluate sensitivity and
17 specificity.
18 DR. SICKLES: You would need a bigger sample,
19 wouldn't you though?
20 DR. CAMPBELL: Not necessarily, no. It is just
21 that the confidence intervals would be somewhat large. But
22 you would be able, in fact, to report point estimates for
23 sensitivity and specificity.
24 DR. SICKLES: Okay. I wouldn't object to that
25 approach, but I would emphasize the subtle cases. The

1 subtle cases are the ones where you are apt to see a
2 difference if there is any difference at all.
3 DR. MONSEES: Dr. Pisano also suggested in her
4 letter, that people here have a copy of, a similar type of
5 study but, rather than looking at the digital images in
6 isolation and the analog images in isolation and grading
7 them, she was suggesting a side by side comparison. I think
8 that would be an important part of an evaluation by experts,
9 to actually took at them side by side; on which one do you
10 actually see this better; and have them design a scale where
11 they can judge perhaps not only whether it is there but
12 maybe the extent of the lesion. You know that sometimes
13 these lesions extend in multiple areas of the breast.
14 DR. SICKLES: But screening extent isn't anywhere
15 near as important as detection.
16 DR. MONSEES: That is true. That is part of the
17 usual evaluation, but I think if you are looking at a
18 standard set you could probably include that perhaps as one
19 of the data point.
20 The other thing, when you were adding cases to the
21 test set, the other cases to the test set that might be
22 interesting would be lesions that presented as palpable
23 lesions that were mammographically occult on standard analog
24 images. So, that would be a very interesting group of
25 patients.

1 DR. SICKLES: This is screening, Barbara.

2 DR. MONSEES: Right. I realize that. But if you
3 are going to test it, you can parlay them out however you
4 want to when you do your analysis. But if you don't put BB
5 markers on and you have a patient who comes in and has a
6 palpable abnormality and it turns out to be cancer, and that
7 mammogram is read as normal, I think you can learn an awful
8 lot of digital were able to see those cases and analog
9 missed them.

10 DR. SICKLES: I think that is a very interesting
11 thing to do but I don't know that that is in the purview of
12 the FDA's mandate here.

13 DR. MONSEES: Well, it may help to improve on the
14 sensitivity for digital where you are talking about having
15 this terrible --

16 DR. SICKLES: I don't think digital would have a
17 problem.

18 DR. ALPERT: I do think you are giving an
19 interesting approach to looking at a different population of
20 women that may be a very important piece of information for
21 either digital or other new technologies that are used in
22 imaging the breast. I think that is very helpful. We are
23 focusing on the screening but I think that is a nice
24 comment.

25 DR. GARRA: I would like to make another comment,

1 and that has to do with this type of approach to evaluating
2 quality. It is a well accepted approach when we do
3 accreditation of practices in both the ACR and AIUL. It is
4 done by using a panel reviewing image quality. So, I think
5 it is well accepted by now by the profession as a way to
6 judge performance of a practice or an individual.
7 DR. SMATHERS: While I am fully aware of the bias
8 of physicians against physics phantoms, I would suggest that
9 to eliminate the bias on the part of the manufacturers that
10 the physics QA mammography phantoms that are out there be
11 used as well because that is one case where you can require
12 them to do the comparison.
13 DR. SICKLES: I can tell you that if you apply the
14 existing phantom to digital versus film, digital wins hands
15 down. You can have the worst possible digital system and it
16 will be far better than film. It is just the way the
17 phantom is set up.
18 DR. SMATHERS: Well, that speaks somewhat to the
19 system already.
20 DR. SICKLES: Yes.
21 DR. ALPERT: And also to the phantoms.
22 DR. SICKLES: Yes, also to the phantoms.
23 DR. GRIEM: As you move from hard copy to soft
24 copy, how do you change your methods of pattern recognition?
25 DR. SICKLES: I have no idea.

1 [Laughter]

2 A very good question. You learn with experience.
3 Everybody who is faced with this is going to have his or her
4 own approach, but with experience you learn.

5 DR. GRIEM: What is the final bottom line?

6 DR. SICKLES: The final bottom line is you get to
7 read with a different modality, whatever you want to call it
8 -- you just get used to it, and there is a learning curve.

9 DR. BERG: I would like to make a few comments on
10 the suggestion by Dr. Sickles. I think that this answers
11 the diagnostic half of the equation in large part, to my way
12 of thinking as well. That is, you have eliminated
13 variability within the readers themselves because they are
14 side by side, looking at the same case. I think in that it
15 would be critical to choose lesions of small size. I think
16 that would be the most objective criteria to choose a subtle
17 case. I think it would be important to record detailed
18 lesion features -- the number of calcifications visualized,
19 the borders of the lesion, and opinion of radiologists as to
20 conspicuity, and level of confidence in characterizing a
21 lesion by the two modalities. This is information that you
22 would get, and it makes a lot of sense.

23 I think what it doesn't answer and what is still
24 up for discussion is the screening half of the equation for
25 sensitivity. I am not an expert in this at all but I know

1 computer-assisted diagnosis was recently evaluated, and I
2 would just throw out for discussion is that something that
3 perhaps we should consider in looking at sensitivity of
4 digital mammography evaluation as being a more objective
5 method that eliminates intrareader variability?
6 DR. SICKLES: I noticed this in the handout that
7 we got. Are you suggesting comparing CAD with digital
8 versus human with film? Or, CAD digital versus CAD
9 undigitized film?
10 DR. BERG: I would suggest CAD digitized versus
11 CAD digital.
12 DR. SICKLES: If you completely eliminate the
13 radiologist, then the direct digital will win because there
14 is image loss when you digitize a film mammogram. That is
15 guarantied to work for the full-field digital. So, I don't
16 know that that is really what they are going to be wanting
17 to do because you know the answer right from the start.
18 Yes, Bob?
19 DR. PHILLIPS: I just might point out that the
20 algorithms for the computer-assisted reading were not
21 designed to cover all the possible contingencies you would
22 see on a piece of film.
23 DR. SICKLES: Absolutely. That is one of the
24 reasons why the CAD programs are not going to be replacing
25 radiologists for a long time to come because they only

1 detect some cancers; they don't detect others.

2 Also, at some point down the line you wanted to

3 know a set of endpoints, levels of various endpoints. Would

4 you direct the question again so we can address it head on?

5 DR. ALPERT: One of the issues that we are faced

6 with in doing these kinds of comparative analyses is

7 figuring out -- and one of the issues that has come from

8 industry to us is that the way the studies are being

9 conducted, the variabilities -- not the technologies --

10 overwhelm differences in technologies.

11 One of the questions we have to keep asking is

12 what is an acceptable difference, a clinically acceptable

13 difference in technology, not what is measurable but what is

14 a clinically meaningful acceptable definition of what is

15 within the same range of sensitivity and specificity that

16 would be appropriate for the clinical community. Because

17 that will, again, drive a lot of the design and the analyses

18 of these trials and size, of course.

19 We are looking for some comment from the clinical

20 community based on what we heard that there is a lot of

21 variation out there, but what is meaningful. Then I think

22 it comes back to that issue of whether we are measuring it

23 with a reader technology set or somehow otherwise

24 controlling the impact. We need to understand it.

25 DR. SICKLES: I think the bottom line there is are

1 you going to control for interreader and intrareader
2 variability? If you do the paired image comparison and
3 eliminate that issue to a great degree, then I think you can
4 look for levels of agreement that are very close. On the
5 other hand, if you are doing the kind of study that Richard
6 Bird produced, a noninferiority study where you do have that
7 variable even if you have multiple, multiple readers I think
8 you are going to have to allow a little bit more wiggle
9 room.
10 DR. ALPERT: We need advice on what is appropriate
11 wiggle room, sort of what the borders of appropriate wiggle
12 room are because we are going to see different studies. It
13 is a fertile industry and they will come back with
14 alternatives even to these new approaches. So,
15 understanding what the wiggle room ought to be clinically
16 would be very helpful to us.
17 DR. SICKLES: As a first guess, in a
18 noninferiority study I think a 0.95 level. You shouldn't
19 make it any harder to achieve than that because you are
20 already biased against digital by creating a noninferiority
21 study. So, I wouldn't make it any more stringent than that,
22 maybe even a little less but I believe they are already able
23 to achieve that. Is that right? Maybe he could answer that
24 question.
25 MR. BIRD: I am going to turn this question over

1 to our statistician, Dr. Olshen, and he will give you an
2 explanation on what confidence and size intervals mean and
3 how they relate to noninferiority.
4 DR. SICKLES: Good. That would be helpful.
5 DR. OLSHEN: At this point, I am afraid I am going
6 to have to turn the tables and throw the ball back in your
7 court, Dr. Sickles. What did you mean by 0.95?
8 DR. SICKLES: In terms of a clinically appropriate
9 delta for sensitivity in noninferiority.
10 DR. OLSHEN: Surely you mean 0.05.
11 DR. SICKLES: I am sorry, yes. I was thinking of
12 the p1 level that they had before. My background in
13 statistics is very poor. You have to excuse me.
14 DR. OLSHEN: You know, as the conversation here is
15 proceeding, it is difficult for me to follow it because
16 while we were admonished at its beginning to distinguish
17 between a study of managing patients and also detections and
18 things like that that come with further workup, I am afraid
19 that those distinctions have been muddied. So, I can only
20 return to where we were, that is, to where I came in to
21 this. In fact, I have an overhead that I would like to put
22 up.
23 [Slide]
24 Basically, where I came in to this is a study of
25 patient management where there was a gold standard. I

1 thought that Dr. Bassett described this morning how ours
2 evolved. I can't speak to the wisdom of others. I thought
3 ours was a good one. It did involve more than just the
4 screening mammographic data, as he described, and we talked
5 about a binary decision, as he emphasized, and there you
6 have it.
7 For us, the issue was workup as opposed
8 necessarily to BIRADS numbers. We have heard a lot about
9 BIRADS 3, and you will see that some BIRADS 3 involves
10 workup and some doesn't, and the same for BIRADS 2.
11 Having said all that by way of making precise what
12 I am talking about, we then defined sensitivity and
13 specificity in terms of our gold standards, and differences
14 of about 5 percent were the orders of magnitudes that seemed
15 reasonable to us, and that we have justified.
16 We have also done simulations, that we haven't
17 reported, on the extent to which selection bias could
18 influence those numbers. We have heard discussion among the
19 panel about that. We have information on it. I don't know
20 and I am not the person to answer to what extent it is
21 privileged, but I will tell you that it all fits together
22 beautifully and you will see it in time.
23 DR. SICKLES: Okay. I think we are on the same
24 wavelength in terms of design as what you have presented
25 because you are talking about using a full diagnostic workup

1 to achieve your gold standard, but you are still doing a
2 binary decision in terms of screening assessment. That is
3 really what we are talking about. So, we are really on the
4 same wavelength, not different wavelengths.
5 What I was trying to get from you is whether this
6 0.05 level is achievable at least in the study that you have
7 done.
8 DR. OLSHEN: Yes.
9 DR. SICKLES: That is all I wanted to know.
10 DR. OLSHEN: I am sorry if I overdid it.
11 DR. BEAM: Loss of sensitivity of 5 percent, isn't
12 that about 8000 to 10,000 missed cancers per year? Is that
13 cost acceptable? I think that is the issue.
14 DR. SICKLES: In terms of a study where --
15 DR. BEAM: If we are allowing a new technology to
16 perform less than the existing technology by 5 percent,
17 let's say, 5 percent less sensitivity, what we are saying is
18 that we are willing to pick up 5 percent missed cancers that
19 would have been picked up.
20 DR. OLSHEN: No, not at all --
21 DR. SICKLES: You will have to come to the
22 microphone.
23 DR. OLSHEN: First of all, we know that only a
24 small fraction of cases that are worked up have frank
25 cancer. So, we are certainly not talking about a 5 percent

1 flaw in picking up cancers. We are talking about confidence
2 intervals being intervals that before the experiment was
3 done had a chance of covering the putative parameter with
4 probability of 0.95 in admittedly fictitious frequentistic
5 replications of the experiment. So, those confidence
6 intervals don't cover the 0.05 deficit in the direction of
7 digital. It doesn't mean at all that we are willing to
8 tolerate 5 percent deficit, and in fact we haven't seen
9 anything like that. So, the answer is we don't have to
10 tolerate it because we don't see it, and we are not talking
11 about missing 5 percent of frank cancer by any manner or
12 means.
13 DR. TOLEDANO: That is where the confusion is
14 coming in because the delta in question is the proportion of
15 cancers that you missed, not the width of the confidence
16 interval, not the desired width of the confidence interval.
17 DR. OLSHEN: We are not talking about the width of
18 the confidence interval. We are talking about a confidence
19 interval that encompasses the minus 0.05, if you like, to
20 the deficit of digital.
21 DR. BEAM: The delta says how much you are willing
22 to give up in sensitivity, for example.
23 DR. OLSHEN: Well, we can run studies where we may
24 be willing to give that up. I am telling you, as a preview
25 of data that you will see and that some of you have already

1 seen, that we don't see it even with the patent selection
2 biases.
3 DR. BEAM: The question is how to select the
4 delta. My point is that if you are going to have a delta
5 for sensitivity, for example, then you must express a
6 willingness to give up a certain percentage of cancer
7 detection.
8 DR. OLSHEN: I am going to hark back now to
9 something Dr. Campbell talked about this morning. Again, I
10 feel like my hands are tied behind my back because I don't
11 have the overheads prepared, although we do have more
12 material than I presented. Greg talked about the issue of
13 sampling. In a 2 X 2 table, for example, there is this sort
14 of putative idea that all the information is in the off-
15 diagonal elements, and there is a sense in which that is
16 true in a technical sense that is made precise in Eric
17 Lehmann's book on nonparametrics. But he also said that
18 there is information to be gotten in the diagonal elements
19 because one of the key parameters is what he called the
20 correlation but, in some sense, the agreement between
21 digital and analog. Without specifying that parameter you
22 can't really do the analyses.
23 So, there are a couple of things that are key, and
24 one of them is the diagonal element of the degree of
25 correlation, for lack of a better word, and I would like

1 Greg to speak after me to address that, but between digital
2 and analog. That is one parameter. Another one is the
3 putative difference between the two technologies, which we
4 call H. And, a third one is owing to the selection bias and
5 now cases are gotten.
6 And, while the word selection bias has been used a
7 lot by people on the panel so far, we haven't talked
8 precisely about what it means. This can all be quantified
9 if Brad is willing to come up and raw a little picture of a
10 2 X 2 table, or Greg, and I would be happy to have either of
11 them follow me and speak. You really need to specify all
12 three of those things. It is not meaningful to talk about a
13 just a difference in H without talking about this
14 correlation, the extent to which digital and analog agree,
15 and also the selection bias.
16 DR. BEAM: Are you talking about an agreement
17 study?
18 DR. OLSHEN: We are talking about a management
19 study that is the first of the two kinds of studies that
20 Larry Bassett talked about this morning, and you really need
21 to specify all three of those parameters. Any one of them
22 is not meaningful without specifying the other two.
23 DR. BEAM: This is an agreement study.
24 DR. OLSHEN: it is not an agreement study; it is a
25 management study.

1 DR. CAMPBELL: Let me clarify. I think people are
2 talking about different things here. Let me attempt to sort
3 it out. If one is interested in the hard endpoint of
4 cancer, then a delta of 5 percent in sensitivity will result
5 in an admission that there may be a lot of cases that you
6 will miss.
7 DR. OLSHEN: That is not what we are talking
8 about.
9 DR. CAMPBELL: I know. That is what I trying to
10 help clarify here. If, on the other hand, what we are
11 interested in is a management study as in screening where
12 the question is should you go for further workup or not, and
13 you look back and say, yes, I should have or, no, I should
14 not have, then that is different. I think Richard's delta
15 of 5 percent is the 5 percent for the management study as
16 indicated on the slide.
17 DR. OLSHEN: But it has many components that
18 contributed to it. Many of them were addressed by Dr.
19 Toledano; others addressed by others of us. That 5 percent
20 incorporates all that can go wrong with all of those things.
21 Our analyses of size and power -- people keep saying, "gee,
22 you need to design a study that account for these sources of
23 variability and these sources of bias." Ladies and
24 gentlemen, we have done it. We have even carried it out.
25 DR. SICKLES: Dr. Alazraki, you have a comment?

1 DR. ALAZRAKI: Well, we will, hopefully, try to
2 conclude some of the things that have been discussed in the
3 conclusions section, but I just wanted to state that I think
4 it is very important ultimately -- as Dr. Sickles has
5 indicated, there is sort of a learning curve here with the
6 digital, and it would be important for all radiologists who
7 get into digital mammography to have some sort of a start-up
8 where they are doing it both ways, both by their
9 conventional way of doing it and with the digital, for some
10 number of cases or some period of time. But that is
11 something which perhaps could be addressed by both industry
12 and FDA. But that would be very, very important I would
13 think.
14 Whether FDA goes to some of these agreement or
15 noninferiority studies, or whatever, to conclude
16 equivalence, as Dr. Toledano indicated, I think there really
17 is no substitute for a prospective trial. She indicated
18 that it could be done without going into tens of thousands
19 of subjects. If it can be done with a thousand subjects, I
20 think that that would be the most conclusive kind of thing
21 that could be done. I think we ought to go back to looking
22 at that and seeing how practically that could be designed.
23 DR. TOLEDANO: On that question, when we are
24 talking about a noninferiority trial, then we go back to
25 this original question which got everyone so upset, which is

1 what is the acceptable delta. So, if we are looking at what
2 is the delta that includes all the sources of variability
3 and that is specific to this trial, we still do need to get
4 back to the question of once we have estimated these
5 relevant quantities, do we have some idea of the impact that
6 it will have when we use it larger screening?
7 I know that is not what we are specifically here
8 to do today, to talk about what happens once it goes
9 clinical, but I think it is very important for us, just as
10 people, to think about it in the back of our head. If we
11 allow the sensitivity to decrease by so-and-so on a study
12 that we do, what are we allowing to happen in the
13 population? So, I just wanted to say that.
14 DR. SICKLES: I think what we will do now is take
15 the break that we planned to do. I would urge the FDA
16 people to talk to us over the break and tell us which areas
17 you want further guidance on that we haven't hit on yet so
18 that we can do that after the open public forum. In this
19 open public forum, by the way, those of you in the audience
20 who want to make comments on what you have heard, that is
21 your opportunity to do it.
22 DR. ALAZRAKI: We will have a ten-minute coffee
23 beak and resume as close to 2:30 as we can.
24 [Brief recess]
25 **Open Public Hearing**

1 DR. ALAZRAKI: We are reconvening the panel
2 session and we are now moving to the second half-hour open
3 public hearing session. You are reminded that the same
4 identification processes and disclosure requirements
5 announced for the first open public hearing session apply to
6 this session. If anyone wishes to speak during this open
7 public hearing session, please raise your hand. Yes, Dr.
8 Hendrick? Is there anyone else who wishes to address the
9 panel during the open public hearing session? Mr. Bird?
10 Anyone else?
11 [No response]
12 In that case, we will move to this open public
13 hearing session. I am asked to inform the speakers that we
14 request that they take no more than five minutes to make
15 their presentations. There will be some time for
16 discussion, if that is appropriate, of their statements.
17 So, Dr. Hendrick, would you like to address the panel,
18 please?
19 DR. HENDRICK: This is quite ad lib as I didn't
20 expect to be coming up here first, but there are a number of
21 points that I wanted to address that the panel addressed,
22 and I just wanted to add a few things too.
23 One of them is the issue of soft copy read and
24 whether it is possible to do soft copy reader screening
25 studies. We have an Army-supported breast cancer grant,

1 along with the University of Massachusetts, and the
2 University of Colorado. Each of the sites has a full-field
3 digital prototype, and we have done side by side filmscreen
4 and full-field digital with soft copy read of approximately
5 3000 women now, starting exactly a year ago.
6 We have not had difficulties in doing soft copy
7 evaluation of those 3000 women, although we don't want to
8 make the case that it is faster or even equal in terms of
9 time to reading film, but it is doable, and we feel that the
10 results we are getting are quite satisfactory doing soft
11 copy interpretation of digital images.
12 There were a couple of other points we wanted to
13 make. One of them is something that appeared on an overhead
14 with regard to hard copy versus soft copy. It said soft
15 copy has greater potential; hard copy has higher resolution
16 and is more consistent. I would take exception to both of
17 those statements about hard copy. It has exactly the same
18 resolution as the digital data set normally, whatever that
19 happens to be. It is certainly not higher resolution on
20 film. It may, in some cases be higher resolution on soft
21 copy but that is not guaranteed either.
22 For the consistent part, it is consistent if you
23 show the same film to more than one radiologist. That makes
24 the image data more consistent. But in terms of printing
25 more consistent images on film from digital, the answer is

1 in the details of who does the printing and how well it is
2 done. So, I just wouldn't want to leave that unremarked on.
3 The other points I wanted to make are that I
4 thought you were getting towards something that was
5 reasonable when Dr. Sickles proposed the side by side
6 evaluation of filmscreen versus digital mammograms, and the
7 case selection sounded fine. It would be very easy to
8 compare what is missing on the chest wall or any other place
9 in the breast in those images in every images, filmscreen
10 versus digital. You wouldn't actually need to generate
11 cases with lesions close to the chest wall. You would get a
12 better picture of that from all the that were submitted.
13 And, I think that is a doable sort of study.
14 In addition, I think that kind of a study would
15 allow companies that have already generated a number of
16 cases of filmscreen and digital side by side to use those
17 cases, and I think most companies would already be there in
18 terms of having a sufficient case mix and number of cases to
19 submit that kind of data for some unbiased panel to
20 evaluate.
21 The part at which I thought things fell apart was
22 when you started asking the questions about what is the
23 tolerable level of sensitivity difference between the two
24 modalities because, once you start talking about
25 sensitivity, you are asking the manufacturers to find out

1 the truth about every case in the study. When you start
2 changing the vocabulary to that more outcome-oriented study
3 you are really talking about a study that will take a year
4 to two years at minimum to find out the truth about all the
5 cases. So, it is like switching the whole model by which
6 you want to prove digital equivalent to filmscreen.
7 So, I would just caution you that when you start
8 using those terms, sensitivity, specificity, and ROC, you
9 change the whole ball game in a very time-cost and patient-
10 intensive way.
11 The last point I would like to make is that
12 whatever you decide, please don't wait six months to tell us
13 about it because we will have submitted something within six
14 months, and most of the other vendors will have submitted
15 something, and if we are on the wrong track here we would
16 really like to know it within the next few weeks rather than
17 the next six months or a year.
18 Thank you very much.
19 DR. SICKLES: I have a question for you. In terms
20 of your comments on sensitivity and specificity, what about
21 if the endpoint weren't sensitivity and specificity but,
22 rather, management based, as the Trex people did their
23 study, and use as truth what happened in the cases which
24 already are proved? That would not impose the requirement
25 to follow-up patients for two years.

1 DR. HENDRICK: Except when you, again, pose the
2 question in terms of sensitivity you are assuming they know
3 something about the missed breast cancers, and if they don't
4 follow up every case, including the negatives, you may not
5 have that answer.
6 DR. SICKLES: The only ones that would have to be
7 followed up would be discrepancies. Correct?
8 DR. HENDRICK: No, not at all.
9 DR. SICKLES: You are looking for relative
10 differences. You are not looking for absolute numbers of
11 sensitivity and specificity. You are just looking for
12 relative differences between one and another. If you want
13 to show equivalency --
14 DR. HENDRICK: When you get into the detailed
15 statistical analysis you still need to know the absolute
16 values, 5 percent, or whatever, difference between --
17 DR. SICKLES: That is over my head. I will leave
18 that to the statisticians.
19 DR. HENDRICK: It is a good question. I think
20 that the absolute values come into play. I guess probably
21 my response to that would be that that is not a bad model
22 for a study to be done, but if that is really the model that
23 the FDA wanted, they should have put that in the guidance
24 document to start with.
25 DR. SICKLES: Leaving that one aside, I still

1 don't think there is a problem with a side by side
2 comparison model where cases are selected where outcomes are
3 known with biopsy proof, whether it is benign or whether it
4 is malignant. I go by the assumption that if the lesion,
5 whatever it is, is visible with equal amount of conspicuity,
6 whatever word you want to put on it, by the expert panel,
7 one system versus another, then one can make the presumption
8 that sensitivity would be equivalent, just even in those
9 terms.
10 DR. HENDRICK: I completely agree with you.
11 DR. SICKLES: That is all I was proposing. I was
12 not proposing anything beyond that.
13 DR. HENDRICK: The difficulty is that the side by
14 side study and the independent evaluation of each modality
15 are two completely separate studies --
16 DR. SICKLES: I agree.
17 DR. HENDRICK: -- and you can't mix the two, or
18 the language that describes them.
19 DR. SICKLES: I agree. They are two separate
20 approaches, completely separate.
21 DR. DESTOUET: I also have another comment about
22 your first remark, and that is the work station. I visited
23 University of Massachusetts and observed their full-field
24 digital operation. As it stands now, their hard copy versus
25 soft copy work station is not ideal by any stretch of the

1 imagination. I think Dr. D'Orsi will agree with that.

2 DR. HENDRICK: And we will agree with that.

3 DR. DESTOUET: As long as the panel recognizes

4 that it is not done yet.

5 DR. HENDRICK: No, I am not saying this is ready

6 for commercial sale, but it is at the point where you can

7 evaluate it and you can actually do a study with soft copy

8 read. It is certainly not optimized for digital mammograms

9 in the sense of being fast enough to display digital

10 mammograms with all the speed you need to make it

11 economically viable for radiologists.

12 DR. DESTOUET: Absolutely.

13 DR. HENDRICK: On the other hand, there is the

14 tradeoff of one view of the mammogram on film versus

15 multiple views of the mammogram when displayed on soft copy.

16 There are certainly advantages to that having the

17 flexibility, and that has been discussed.

18 DR. ALAZRAKI: Thank you very much. Mr. Bird?

19 MR. BIRD: One point that was brought out earlier

20 that I just wanted to add some information to was that of

21 taking the digital and the filmscreen mammograms and

22 comparing them to the two in a somewhat subjective fashion.

23 We did this in our study for all cases. For all cases we

24 asked the principal investigator to assess whether they

25 preferred the digital, the analog, or had no preference, and

1 in all cases with abnormality we asked them to assess, for
2 the abnormality, did they prefer the digital, the analog or
3 no preference. In both cases it was statistically
4 significant that the principal investigator preferred the
5 digital, and with the abnormalities that was with a p value
6 of 0.02 and for the normal breast structure it was a p value
7 of 10_{-4} or 0.01 percent, which is an extremely low p value.
8 So it was very significant in favor of the digital fir
9 subjective assessment.
10 DR. ALAZRAKI: I guess the problem with that is
11 the bias of the reader since they know what they are looking
12 at. They know what the digital is and they know what the
13 film is, and they undoubtedly come with some preconceived
14 prejudice.
15 MR. BIRD: Not necessarily. This was in addition
16 to a full-blown clinical study. This was one small subset
17 of the truth set development. This is now how we based what
18 the outcomes were. That was completely separate. If
19 anything, I would say due to familiarity of filmscreen it
20 probably would have been biased towards the filmscreen, if
21 anything, but I don't think there is any evidence to support
22 that one way or the other.
23 DR. SICKLES: I think if the FDA were to accept
24 such information they would probably be looking to an expert
25 panel that was apart from the one that you recruited

1 yourself. I think they would probably look to an expert
2 panel that they recruited to eliminate any biases of people
3 who are already affiliated with the manufacturer.
4 MR. BIRD: Yes. The one bias that we didn't have
5 at least was selection bias; it was all cases. But I
6 understand your point about the persons involved.
7 DR. ALAZRAKI: I am glad you mentioned that. What
8 you are describing is where the image itself is the
9 endpoint. It is a study where the image itself, comparing
10 one image to another, is the endpoint and, of course, you
11 want to have sufficient numbers so that you have a variety
12 of conditions that are evaluated or compared between the two
13 images. I don't know if FDA would consider that but I
14 personally think it is a very reasonable suggestion.
15 MR. BIRD: As I said, it was in addition.
16 DR. ALAZRAKI: Thank you. We will conclude then
17 the open public hearing portion of the meeting, and we now
18 will move to a wrap up of the panel's input on what
19 improvements or modifications to the guidance for the
20 clearance of digital mammography devices we can possibly
21 suggest at this point.
22 **Panel Wrap up**
23 DR. SACKS: What I sort of gleaned from the
24 muddying of the waters -- actually, it was good that the
25 waters got muddied -- I think we have been talking a little

1 bit at cross purposes, and I think there is a question that
2 we need to answer here.
3 First of all, we have been talking about several
4 different phases of the screening to biopsy process, the
5 first phase being a decision of does this woman need
6 anything further, or is she just a BIRADS 1 or 2.
7 Then there is another phase, after you have done
8 that, does she need ultrasound, say, and so on; does she
9 need a biopsy. I think Dr. Bassett's original slide this
10 morning showed these various levels, and Dr. Monsees has
11 referred to it as well.
12 Each of these phases has its own sensitivity, and
13 the difference between the sensitivity of digital at that
14 phase and the sensitivity of analog at that phase -- we
15 could set a different delta for every one of those phases.
16 Those deltas would not have any necessary relationship to
17 one another.
18 Having said that, the question for the panel is
19 which of these deltas -- is it just one of them or is it a
20 set of them that are the important ones, and what are
21 reasonable values? I think this gets at Dr. Beam's last
22 question.
23 DR. SICKLES: My understanding is that your
24 principal interest is in screening because you are not
25 really interested so much in looking at digital versus

1 analog in terms of diagnostic workup. So, you are really
2 concentrating on screening here because of the public health
3 implications of screening mammography. If that is truly
4 your interest, then you should concentrate on the first one,
5 which is workup/no workup, BIRADS 1/2 versus everything
6 else. That is a very clean endpoint that clinical
7 mammographers shouldn't have trouble with in no matter what
8 kind of study you are going to accept. That is easy because
9 we know what that is. There is going to be variability
10 between observers but at least the concept is reasonably
11 well known to radiologists.
12 What you judge to be truth, on the other hand, can
13 get very muddy. We have heard two different kinds of
14 studies proposed. One would be the noninferiority type of
15 study. The other one would be the more informal side by
16 side image comparison type of study. They are vastly
17 different studies. They have completely different designs
18 and they shouldn't be confused at all because one of them is
19 just comparing image versus image in terms of what you can
20 see, and the other one is much more in terms of what you
21 might measure with other endpoints that are clinically
22 related. Is that clear?
23 DR. TOLEDANO: Could we take a step back? We saw
24 this morning the six levels, the six-tier system, and I
25 think this is looking at the second level. What Dr. Beam

1 was saying is level one sufficient? Is the type of study
2 that Dr. Sickles is proposing, a level one study, sufficient
3 to prove substantial equivalence? I think that is probably
4 the first question we need to ask. What level of proof do
5 we need for substantial equivalence?
6 DR. SCHULTZ: I am not sure that there is a real
7 simple answer to the question. I think that in terms of the
8 FDA mandate, we look at safety and effectiveness. In terms
9 of those six levels, as Dr. Alpert pointed out earlier, in
10 most cases for a 510(k) when we are looking at substantial
11 equivalence we are able to determine equivalent safety an
12 effectiveness based on looking at the technology, if you
13 will, level one. But -- and I think that is why we are here
14 today, and this is what Dr. Sickles asked originally -- or
15 Dr. Garra -- somebody asked this question originally, why
16 did we pick digital to make all this fuss about? I think
17 the answer is that because of the screening potential for
18 digital, and the differences, and the expectations raised by
19 digital, we felt that in order to determine substantial
20 equivalence we needed something more than the level one or
21 the technological evaluation.
22 Now, the question is how much more? And, how can
23 we get that additional information in order to be able to
24 get this to market, you know, in our lifetime? So, that is
25 what we are sort of hoping we can get from you.

1 DR. ALPERT: Just to complete the issue, we have
2 made a determination internally that we need clinical
3 information in addition to an understanding of the
4 technology in order to make this substantial equivalence
5 determination. What we are asking you is what are the
6 options in that clinical information that we ought to be
7 focusing on.
8 I put these down because I think they are
9 important to remember. Digital is not one thing; it is a
10 type of technology and each company has its own device.
11 Given that each company also has the flexibility to design
12 its own study and determine how they are going to establish
13 to us using clinical data that they are substantially
14 equivalent in safety and effectiveness, the consensus I
15 heard is that as an approach a noninferiority study, a
16 clinically valid noninferiority study, properly sized with
17 the appropriate types of statistical analyses, is an
18 approach that the panel feels is a good alternative to the
19 equivalence because we talked about equivalence as being a
20 problem with this technology -- agreement, I am sorry. A
21 real agreement study doesn't get us there. A noninferiority
22 approach to comparing these from a safety and effectiveness
23 point of view made sense, and I just wanted to confirm that
24 that was the feeling of the panel.
25 DR. SICKLES: Yes.

1 DR. ALPERT: Okay. Second, given that the studies
2 will look different, given that many studies are already on
3 the way and that the studies that we will receive in
4 submissions will look different, what are the critical
5 things we need to know, where we need to look at the
6 analyses that they are presenting to us of their comparison
7 of digital to filmscreen? Where ought we to be putting our
8 attention?
9 The third thing is that we have talked about two
10 different model studies. We talked about one very much like
11 the one that Trex talked about, which is put them in this
12 bin; put them in that bin; and do enough establishment of
13 truth within that population to really know what the
14 sensitivity and specificity is in that practice type study,
15 but involved looking at truth for a portion -- I am not sure
16 exactly what portion of t he population but a significant
17 number of patients in their given study.
18 We have also talked about a different approach
19 being more a side by side comparison with an expert panel,
20 however that panel might be configured, to look at a
21 comparison, literally patient by patient, in digital and
22 filmscreen, feature by feature.
23 Those were two different approaches to looking at
24 clinical data to support this claim the manufacturers are
25 making that they are, in fact, substantially equivalent in

1 safety and effectiveness.

2 DR. TOLEDANO: What is your definition of

3 effective?

4 DR. ALPERT: That is a real hard one.

5 Effectiveness says it is a significant benefit to a

6 significant portion of the population. That is the

7 regulatory definition of effectiveness. When we look at

8 diagnostics, we tend to look at sensitivity and specificity

9 in whatever way we can understand it, and they are different

10 depending on what type of diagnostic you are looking at and

11 what the clinical environment is, and how tightly you can

12 establish that.

13 I want to make two more comments. The first is

14 that the question of where we are going to focus our

15 attention in understanding the bins does not mean that we

16 don't also need to understand what truth is, whether that is

17 for the off-diagonal patients and a portion of the on-

18 diagonals, as Greg presented this morning, or whatever the

19 right selection process is, because it is important. It may

20 not be the primary outcome but it is important to

21 understanding what we got in the first measure, and we can't

22 ignore it. So, that is in there.

23 The last comment I want to make is on a totally

24 different tack, and that was that we heard from the

25 manufacturers that they are concerned about the time in

1 which this revised guidance will be available. I need to
2 tell you that there is a good guidances practice process
3 which we must follow in issuing a new guidance in this area.
4 It requires us to publish such a guidance and to get
5 comments before it becomes a working guidance, and we will
6 have to go through that process.
7 Companies are not bound, as we are not bound, to
8 doing it this way. So, this discussion and even a
9 preliminary guidance helps companies understand where we are
10 going but there is a process that we need to go through, by
11 regulation, in order to get a working effective guidance,
12 active guidance. It will be a proposed guidance before
13 that.
14 Again, let me remind all of the manufacturers in
15 the room that these are guidance. They will be used as
16 guidance. They will not bind. If any reviewer binds you,
17 talk to Mr. Phillips, talk to Dr. Yin, talk to me.
18 DR. ALAZRAKI: Does any member of the panel want
19 to address any of Dr. Alpert's questions?
20 DR. SICKLES: Why don't we start from the top? To
21 answer the first question, I think that you have a consensus
22 and the answer is yes.
23 To answer the second question, what are the
24 critical questions should be asked, it depends on the type
25 of study. We have two types of studies that are being

1 looked at. In terms of the side by side images, what you
2 are looking for, in my opinion, and maybe we will poll
3 everybody around the room, is equivalence in imaging
4 features that are judged to be important in terms of cancer
5 detection -- mass detectability, mass margins, calcification
6 detectability -- just in terms of whether they are of the
7 type that would generate a recall or not.
8 Architectural distortion detectability, beyond
9 that you can't go at screening. Looking at the ability to
10 image the deep part of the breast, and I liked Dr.
11 Hendrick's suggestion about just using all the images and
12 assessing how far back you get in one versus the other.
13 Experts can do this readily just by a side by side
14 comparison -- this one got deeper than this one.
15 What mix you choose to have in there of abnormal
16 and normal cases I would leave to the manufacturer but you
17 can give them some guidance here too. I would emphasize the
18 subtle cases. I would emphasize a good mix of calcification
19 cases, mass cases, architectural distortion cases that
20 include cancers, known, biopsy proved cancers.
21 A simple way to minimize the chance for selection
22 bias is to find out from the people who are submitting cases
23 to you their approximate accrual rate of patients. We know
24 from screening statistics about how many abnormal cases
25 should be found. We know from screening statistics about

1 what percent of cases should be masses and calcifications
2 and architectural distortions. If you know their volume of
3 work and whether their expected outcomes are, you can ask
4 them to present information with reasonable numbers of cases
5 within a time frame, a designated time frame, so that you
6 are at least moderately comfortable that they are not
7 picking out only the ones where things do well. This is an
8 approach that the MQSA uses to a limited degree when
9 clinical image reviewers submit. They don't let you submit
10 images for three years; it is a one-month interval where you
11 have to pick your cases. That way there is less chance for
12 selection bias.
13 Let's have others comment on just this design
14 phase for now.
15 DR. BEAM: I would like to reemphasize that such a
16 study is not estimating sensitivity or specificity directly
17 but only indirectly by the principal presumption. So, in
18 many ways it is like a surrogate study. I think we need to
19 approach it in that regard, and give real hard thought as to
20 what the endpoints are that will be needed for such a study,
21 and what level of evidence would have to be mustered.
22 Again, I want to emphasize that this is a
23 substitute for doing a study that looks directly at
24 sensitivity and specificity.
25 DR. SICKLES: Well, I would suggest in that regard

1 since this is a clinical study and I have a little more
2 confidence than when you start asking about p numbers, that
3 overall it is at least equivalent for digital. In other
4 words, if you have 100 cases of whatever, then you ought to
5 be able to see things at least as well overall with digital
6 as you do with film. I think there will be no problem doing
7 that. I think you will see cases better with digital than
8 with film so I don't see that as a problem for the
9 manufacturers. But as a bottom line minimum, yes, there may
10 be one case out of 100 where you saw it better on film than
11 digital, but there may be 20 cases where you saw it better
12 on digital than film. I think overall, as a bottom line,
13 you ought to be able to see findings, important clinically
14 relevant findings at least as well with digital as you do
15 with film. That is the definition of noninferiority. I
16 don't know whether that is acceptable to everybody but that
17 would be acceptable to me as a surrogate for sensitivity.
18 DR. DESTOUET: I have a question. Is it hard copy
19 compared with hard copy, or is it hard copy because some of
20 the systems I think are not able to generate hard copy
21 images?
22 DR. SICKLES: The bottom line for film should be
23 hard copy of film. The bottom line for digital should be
24 whatever the manufacturer chooses to submit. If they choose
25 to submit using hard copy, that is fine. If they choose to

1 submit using soft copy, I think that is fine too. I think
2 the FDA has to seriously consider the big differences
3 between soft and hard copy interpretation so that if a
4 manufacturer gives you only hard copy data, at some point in
5 the future when they want to market it as a soft copy
6 device, you may want to ask them for the same thing using
7 soft copy because they are different approaches, and I think
8 they have to be evaluated separately. Maybe they will
9 choose to give you both right at the beginning, or maybe
10 they will only give you one. It is whatever they give you.
11 DR. TOLEDANO: I want to go back a stop again
12 because I like to do that, and get at the question that Dr.
13 Beam was getting at. We started off the first question with
14 trying to define what substantial equivalence is. Then I
15 asked another question, what is effectiveness. The two
16 types of studies that are on the table, as Dr. Beam brings
17 up, are looking at different definitions of effectiveness.
18 The feature analysis study is saying -- and this
19 is my understanding -- the feature analysis study is saying
20 that digital, in whatever form, hard copy, soft copy, has
21 the same effectiveness if it gives the same information to
22 the radiologist as determined by the expert panel of
23 radiologists. That is one definition of effectiveness.
24 The other definition of effectiveness has to do
25 with sensitivity. In screening that is what we want,

1 sensitivity. So, that would be a different study where we
2 would look at the sensitivity of the digital.
3 Two different studies, two different definitions
4 of effectiveness. And that, I think, is the more important
5 question, what do you mean by effectiveness?
6 DR. SCHULTZ: I am not sure we are really talking
7 about two different definitions of effectiveness. I think
8 what has been brought up is the fact that if you accept the
9 feature analysis option, really what you are saying is that
10 by analyzing feature by feature you are providing a
11 surrogate for sensitivity. I think that is what Dr. Sickles
12 is saying, at least as I interpret it. Much the same way as
13 the original agreement study was a surrogate for looking at
14 sensitivity and specificity because basically what we said
15 there was that if the two agreed to some very, very high
16 level, we would accept the sensitivity and specificity of
17 analog as being the same for digital.
18 So, I don't think that we are changing our
19 definition of effectiveness in a regulatory sense, but I do
20 think that there may be different ways to achieve that same
21 end. I think that is kind of what we are struggling with.
22 DR. TOLEDANO: I don't think you are changing your
23 definition of what makes something effectiveness; I think
24 you are changing your standard of proof for what makes
25 something effectiveness. Is it giving the appropriate

1 information to the radiologist? That is one standard of
2 proof for effectiveness. That is the feature analysis
3 study. Or, is it coming up with the same sensitivity? That
4 is another standard of proof. It is a different kind of
5 study.
6 DR. SACKS: I think if a feature comparison were
7 to come out identical, then it would necessarily imply the
8 other level of effectiveness. The problem, of course, is
9 that the likelihood that it will come out identical --
10 DR. BEAM: It won't.
11 DR. SACKS: That is right, it won't. Then we are
12 into the problem that you are raising. Put the slide back
13 up that I wrote.
14 [Slide]
15 Even if it were the case that in the first phase
16 digital pictures picked out the same percentage of women to
17 be worked up as the analog pictures did, it wouldn't
18 necessarily be the same ones, the same women. That is one
19 thing. But even if it were identical in numbers, that would
20 in no way imply that the outcome, as you follow it to the
21 final phase, would be the same or imply anything about
22 whether the digital would have higher sensitivity for cancer
23 or a lower sensitivity for cancer. That is the problem.
24 There is a disjunction between the first phase -- good; I
25 am glad you are shaking your head because the way I

1 understand it, there is a disjunction between that and/or
2 image comparison and effectiveness. In other words, what I
3 am saying is that I am not sure that these are effective
4 surrogates.
5 DR. SICKLES: Let me give you a clarification.
6 What I mean for manufacturers to put in their data sets are
7 cases of biopsy-proved cancer. So, you know they are
8 cancer. And, I want them to show that feature by feature,
9 on average, you see at least as much with digital as you do
10 with film. That is a surrogate for sensitivity because they
11 are known cancers. The cases are known cancers.
12 DR. SACKS: Aside from the retrospective bias that
13 Naomi mentioned.
14 DR. SICKLES: Yes. They can separately look at
15 the ones which are known cancers and come up with that
16 feature for feature analysis. You can put in other cases
17 that aren't cancer; fine. But you look at specifically at
18 the subtle cases of known cancer feature for feature, and I
19 think that if the same findings are visible feature for
20 feature, that is a reasonable clinical surrogate for
21 sensitivity.
22 DR. SACKS: That was my point. That is only if
23 they are identical.
24 DR. SICKLES: Or if it is better with digital. If
25 digital sees the features better, judged side by side, or if

1 digital sees the same features side by side it is
2 equivalent. Remember, this does not require every single
3 case to be identical. It means on average. If you have,
4 you know, 50 of these cases that are subtle and digital sees
5 48 better and film sees 2 better, that is in favor of
6 digital. It doesn't mean that digital has to see every
7 single case better, but it means that digital is functioning
8 at a better level than film. If you wind up with
9 equivalency, 25 for digital and 25 for film, that is good.
10 If it winds up even, that is good. If it winds up negative,
11 that is bad.
12 DR. CAMPBELL: I guess I have a question which I
13 would like to hear the panel's opinion on. I realize in a
14 screening study the sensitivity is very important. My
15 question is what happens if the false-positive rate doubles?
16 Even though the sensitivity is a little better but the
17 specificity declines by so much, at least in the short term
18 and maybe in the long term that won't be a problem as people
19 learn how to use the new technology, but the question that I
20 would like to pose to the panel -- I think you have to worry
21 a little about specificity even in a screening situation,
22 and I would kind of like some feeling for, you know, just
23 how tolerant the panel feels about this.
24 DR. SICKLES: I don't think any of us can answer
25 that question realistically because we don't know what has

1 been experienced with previous new technology introductions.
2 We just don't have the numbers.
3 DR. ALPERT: I think you are answering a different
4 question. You are answering the question about what happens
5 once it is introduced. That is a different question. What
6 we are trying to figure out is in the construct -- in the
7 construct, assuming that the construct has meaning and that
8 people get trained, what is an acceptable level for allowing
9 a product to enter the marketplace? What is the threshold
10 for loss that is acceptable for introducing the product into
11 the marketplace? Because that is the decision we are having
12 to make.
13 DR. SICKLES: Your question then is relevant only
14 for the outcome study, not for the feature analysis study.
15 How are you going to get false positives in feature
16 analysis?
17 DR. GARRA: You said 25 could be one way and 25
18 could be the other. So, if you take a simple analysis like,
19 say, a sign test they would be equal. So, if you had one
20 that had 28 and the other one had 22 -- so 28 versus 22, you
21 would have a difference, and then you have to test for the
22 significance of that difference based on the numbers.
23 DR. SICKLES: What I am getting at is that the key
24 cases in the side by side comparison are cases. We are
25 assuming in this model that we are taking the radiologists

1 out of the game. We are not looking at inter- or intra-
2 observer variability. We are just looking at imaging
3 features and their visibility to an expert panel. We have
4 to assume under those circumstances that if lesions which
5 are cancer are visible and seen as well or better, the
6 radiologists will use that information appropriately.
7 Similarly, we would come to the conclusion that lesions
8 which are benign which are visible and seen by radiologists
9 better would be interpreted at least as well as they would
10 be otherwise. I think that is reasonable.
11 DR. GARRA: You would still need a threshold.
12 DR. SICKLES: If film is better than digital, that
13 fails. If film is equal to or better than digital it
14 passes. I think that is a clear endpoint.
15 DR. GARRA: You would accept --
16 DR. SICKLES: I would accept anything -- it is not
17 going to come out that way, but I would think that if
18 digital winds up better than film on average -- you are
19 going to have a lot of ties and then you are going to have
20 some that are better than others. If digital comes out
21 better than film, I would suggest that digital shows
22 equivalency, noninferiority by your definition.
23 DR. GARRA: But you have to have the threshold --
24 DR. SICKLES: But I gave the threshold. What I
25 have trouble with, and I suspect most of the clinicians here

1 have trouble with is giving you specific numbers for your
2 outcomes analysis study. That is where we are failing you
3 because we just don't know what the right answers are. I
4 know you want them but I am not sure we are the right people
5 to tell you what they are.
6 DR. ALPERT: Maybe not absolutes, but reasonable
7 effectiveness, reasonable safety. We frequently have to
8 work in areas where the absolute is not known and we have to
9 go with what makes reasonable sense to a collection of
10 experts such as yourselves. That is one of our challenges
11 frequently.
12 DR. ALAZRAKI: In order to get the data that you
13 are suggesting be analyzed you have to have done digital and
14 conventional film mammography on a fairly large number of
15 patients to exact enough samples of people with positive
16 biopsies, or even benign biopsies, to analyze the features
17 as you are suggesting. I would think that, you know, you
18 would have a large sample, therefore, of films that could be
19 compared in addition to those with real findings. So, in an
20 analysis by an expert panel or group of all the films, there
21 would be lots of films there to compare which did not
22 necessarily have malignancy and you could get a good idea in
23 the readings of how man variances there are, I would think.
24 DR. SICKLES: Well, yes and no. You will get a
25 different reading but, as Dr. Hendrick said, in that

1 situation you won't have truth because you have to follow
2 them out to two years to find out whether the positive call
3 that wasn't acted on really was or wasn't cancer.
4 DR. TOLEDANO: Well, then I need to go back to the
5 question of effectiveness because a long time ago in one of
6 our handouts, and this is what Greg Campbell just brought
7 up, it said does not decrease rates of false negatives,
8 which is our emphasis for screening, but also does not
9 decrease the rates of false positives, and that is the
10 specificity. So, if we need to demonstrate anything about
11 that, then your feature analysis study needs to include
12 those kind of patients.
13 DR. SICKLES: Number one, as I said before, I
14 don't think that specificity is an immediate problem. I
15 think that could be something you might direct to the post-
16 approval process. That would be my opinion. I fully
17 expect, although I didn't hear it from the manufacturer
18 side, I fully expect not just in practice but actually in a
19 trial, if people are using it the way it normally happens
20 with new technology, that you are going to get more false
21 positives, not fewer. If you get fewer, wonderful; it is a
22 fantastic technology that gets introduced without a learning
23 curve. But I expect you will have more, and I think it will
24 be really hard to separate the learning curve from the
25 ultimate practice.

1 DR. ALAZRAKI: The expert panel reading the films
2 has to be a panel with experience reading digital. So you
3 get away from that learning curve.
4 DR. SICKLES: Well, it is going to be very hard to
5 get an expert panel who has experience reading digital who
6 aren't already affiliated with one of the manufacturers.
7 DR. DESTOUET: I would like to point out that
8 there are many radiologists who are exposed to digital
9 mammography. They are just not exposed to whole breast
10 digital. But most of the mammographers are radiologists who
11 do stereotactic biopsies, for instance, which have small-
12 field digital capability, and do all the window settings.
13 So, there is a group of people out there who are exposed to
14 digital.
15 DR. MONSEES: Plus, one could use a training set
16 to educate individuals so that they wouldn't necessarily
17 have to have vast experience but they could have enough
18 experience to get over that learning curve.
19 The other thing about the side by side studies
20 that could be added to it, probably without much additional
21 difficulty since many of them would be cases that would be
22 worked up, would be side by side comparison of the
23 magnification images. So, the panel could evaluate, say,
24 the screening views analog versus digital and the side by
25 side mag. views analog versus digital in terms of

1 recommendation for biopsy. I am talking about another thing
2 that could be added.
3 DR. SICKLES: It could be added but that is way
4 out of the realm of screening.
5 DR. MONSEES: Yes, it is. It is, but when we
6 started to talk about specificity, if people had any worries
7 about that, I just thought that might be another thing that
8 could be added in.
9 DR. TOLEDANO: I just wanted to go back to the
10 definition of effectiveness. In our utility we have to
11 include the false negatives and the false positives, and it
12 is very much true that for the screening situation it is the
13 false negatives that we are more interested in -- I don't
14 know how many times more interested, ten times, a hundred
15 times more interested, whatever --
16 DR. SICKLES: A thousand.
17 DR. TOLEDANO: A thousand times more interested?
18 Then build that into your study but you still do need to
19 show that the technology itself, after the learning curve,
20 is not going to increase the false positives. It is part of
21 the definition of effectiveness.
22 DR. SICKLES: What is unreasonable? I am not
23 talking about an outcome study where you can actually
24 measure it, but features analysis. What is unreasonable?
25 If you see features as well or better with digital, to make

1 the assumption, as you would with biopsy-proved cancers,
2 that sensitivity is the same, that if you see the features
3 of benign lesions as well or better that the specificity
4 would be the same.
5 DR. TOLEDANO: Nothing unreasonable about that,
6 DR. SICKLES: Okay, then you just build it in.
7 DR. TOLEDANO: Yes, build it in. That is fine.
8 DR. SICKLES: That is fine.
9 DR. BEAM: My thought would be that the proof of
10 the pudding is in the eating ultimately. The proof of the
11 pudding is in the eating. What I am saying is you would
12 have to do a study to look at it.
13 DR. SICKLES: I think Craig is less comfortable
14 with a features analysis study then the clinicians on the
15 panel are.
16 DR. TOLEDANO: We are both less comfortable with
17 that.
18 DR. BEAM: My point is just to keep in mind
19 exactly what it can and what it cannot do. It's simple.
20 Number two consideration is that these are found cancers.
21 So, they have been found already. So, we have to be a
22 little bit concerned about the unfound cancers.
23 DR. SICKLES: In a noninferiority study you don't
24 have to worry about non-found cancers, I would think.
25 DR. TOLEDANO: In a screening study you have to

1 worry about non-found cancers.

2 DR. SICKLES: In a noninferiority to film study,

3 you know, your baseline is what is found on film. Right?

4 DR. TOLEDANO: No way.

5 DR. SMATHERS: To keep the readers honest there

6 ought to be a small fraction of what you would call BIRADS 1

7 and 2 films in there or they are going to find something on

8 every film they see.

9 DR. ALAZRAKI: Right, and you will have lots of

10 those types in your sampling.

11 The third question is, I think, in doing this sort

12 of a study because someone looking at the study might come

13 up with a subset of patients that was not included. So it

14 might be beneficial to, a priori, convene a panel, expert

15 panel, to identify the patient groups, perhaps feature

16 types, feature categories.

17 For example, is it important to stratify by age?

18 Is it important to stratify by breast density, and so on and

19 so forth, in order to make a well-defined set of standards

20 for this--

21 DR. SICKLES: I understand where you are coming

22 from. Again, I would suggest if you narrow the time window

23 in which cases could be accrued and match that to the

24 expected accrual rate of all patients by that manufacturer

25 and the various sites that they have out there, you can get

1 a representative sample of what is out there without
2 substantial selection bias. I think you can do that.
3 DR. DESTOUET: I have a problem with that in the
4 sense that if you are looking at a screening population, the
5 number of cancers you are going to see in those women is
6 really small. And if you basically want to look at the
7 subset of cancers, you are asking the manufacturers to come
8 up with a large number of cancers.
9 It seems as though you are not talking about a
10 short interval of time. You are talking about a couple of
11 years.
12 DR. SICKLES: But they have that already.
13 DR. DESTOUET: Give me a retrospective dataset as
14 opposed to a prospective dataset but one in which you can
15 still eliminate bias from a retrospective dataset; is that
16 possible? Dr. Beam?
17 DR. SICKLES: Dr. Beam defers to Dr. Toledano.
18 That is a good question. You have to be concerned about it.
19 DR. ALPERT: We have acknowledged in all of our
20 discussions with manufacturers that the problem--it is the
21 one that we talked about very early this morning when we
22 began--that to do a prospective study in capturing of
23 cancers is the huge study that all of us are not talking
24 about.
25 So there is a component of every one of these

1 studies that is enriched and retrospective in order to get
2 the number of cancers with the variation amongst them that
3 we all believe is necessary in evaluating the technology.
4 It is one of the things we are dealing with in trying to
5 come up with study designs that are not true prospective
6 screening trials because we have acknowledged that those
7 will be, by force of nature, very large.
8 We are hearing from our manufacturers, from the
9 clinicians advising them, that those studies are too large
10 to be responsive to the need to develop information about
11 these technologies.
12 We have also heard, and I think we have all
13 stated, that we all believe that that is going to be
14 necessary at some point in time, whether it is by one
15 manufacturer, by the entire consortium of manufacturers,
16 with the help of the government in the form of discussions
17 that have come up with NCI.
18 We will have to have some understanding of that
19 prospective information and impact, but what we were dealing
20 with here is an expressed need by both the clinical
21 community and the manufacturer community for alternatives to
22 that study which is why these are all, by force, somewhat
23 retrospective, somewhat selected, and why we are so
24 concerned about selection bias and controls on selection
25 bias because it is the only way to get the dataset.

1 DR. SICKLES: I have to leave to catch a plane, so
2 I am going to give you about a one-minute conclusion. But
3 you are going to keep going. I think Dr. Hendrick had a
4 comment. I don't know if you can or can't recognize him.
5 From my point of view, and maybe the other
6 clinicians in this group will address this, I think you have
7 two types of study you can consider. I feel more
8 comfortable with the feature analysis study because it is
9 easier for me to handle the endpoints.
10 But I think either one of them will work. I think
11 you can address all of your concerns with either kind. What
12 I am uncomfortable with is giving you specific numbers for
13 the outcomes study because I just don't know where they
14 should be.
15 With that, I will leave you and I thank you for
16 asking me to come.
17 DR. ALAZRAKI: Thank you, Dr. Sickles.
18 DR. MONSEES: I would like to reiterate that, as a
19 radiologist, I feel more comfortable with the feature-
20 analysis study because it is something we can get our hands
21 around. It is something that we know and that we know will
22 resolve the issue one way or the other.
23 It may be equivocal. It may answer positive or
24 negative. But at least it is something that is not going to
25 be subject to variable interpretations, a variety of

1 different readers, et cetera, that we may not be able to get
2 our hands around. I feel much comfortable with that. I
3 think that, given a panel of expert readers who are doing
4 this, not just one individual, that we can really look at
5 this and tell whether it is equivalent or not.
6 DR. ALAZRAKI: I think FDA is hearing that the
7 clinicians on the panel are reasonably comfortable with the
8 feature analysis type of study as has been described, that
9 the statisticians are not quite as comfortable with that.
10 DR. BEAM: It is not a matter of comfort. It is a
11 matter of a level of proof, definition of proof. The
12 concern that I have is that we do not confuse feature
13 analysis as actually giving a sensitivity and specificity,
14 level-2 type considerations. To focus in on the feature
15 analysis, the FDA should be saying, then, that they are
16 happy with, basically, the level-1 type analysis.
17 DR. ALPERT: I was going to put on the table the
18 following heresy, if you will. The feature analysis,
19 depending on how the actual study is designed, but as an
20 approach, as I said earlier, seems to be answering the
21 question that would be answered by a phantom, if we had one,
22 which we don't have.
23 The question, then, or the heresy I am going to
24 put on the table, and I don't want the industry to panic,
25 please, is are we talking about essentially two pieces that

1 fit together to give the answer that we are looking for and
2 that is that the feature-analysis-type evaluation of the
3 technology tells us something about the crispness with which
4 the technology presents the biological abnormality.
5 The noninferiority approach tells us something
6 about sensitivity and specificity. It seems to me that what
7 we heard, actually, from Trex is that they had a little bit
8 of both in their analysis. They did the noninferiority
9 study and can tell us a lot about feature analysis as a
10 result of having a lot of films.
11 I think where I am going, and this is not
12 requirements--again, this is just understanding of what I am
13 hearing from the panel--that we are getting different
14 things. It would seem to me that if a non-inferiority study
15 were, in fact, conducted, and had appropriate representation
16 of the types of patients in which we are all interested,
17 then doing some type of feature analysis on a subset of
18 those films is a "doable," because you already have the
19 films.
20 You already have, in the same time, of the same
21 patient, same issue, and you know a lot about the patient.
22 So I am just suggesting--and I guess I am asking the
23 industry if, in fact, that construct is crazy or whether, in
24 the films that they are seeing, in the populations that they
25 are seeing and the kinds of studies they are developing, if

1 one couldn't, then, get some of this feature analysis that
2 the clinicians are looking for to make them more comfortable
3 with understanding how digital performs.
4 DR. GARRA: I would like to make a comment. I
5 think us clinicians would be comfortable with an outcomes
6 study. We just think it is an unnecessary expenditure of
7 resources. For what we would need to make our judgment, I
8 wouldn't need that. I think the other clinicians feel the
9 same way that the feature analysis would probably be less
10 expensive and fulfill the need.
11 DR. ALPERT: Our issue is how do we translate that
12 into an understanding of sensitivity and specificity and,
13 therefore, safety and effectiveness.
14 DR. GARRA: That is a policy decision on the part
15 of the FDA if you want specificity and sensitivity, then you
16 can--somebody is going to have to pay for it.
17 DR. ALPERT: Well, we are not funded to do the
18 research.
19 DR. GARRA: Right. So you make other people pay
20 for it.
21 DR. ALPERT: Very honestly, Dr. Garra, I think it
22 is worth commenting on. The agency is not funded to support
23 the research by the manufacturing community or whoever is
24 appropriate to do the research. It is a fact, not a choice.
25 The fact that this is an area that will cost in terms of

1 developing the data--it is not the only one that we regulate
2 where the cost of doing trials is extremely important and
3 high impact on companies choosing whether to come into the
4 environment or not choosing.
5 I am not belittling it. It is a very important
6 one. It is just one over which I just need to say that the
7 FDA has no control over how much money is available in that
8 research environment.
9 DR. GARRA: I understand that. I am just saying
10 that if the FDA makes the policy decision to go with
11 sensitivity and specificity, it is policy decision because
12 you haven't done it for other parts of things that could
13 change the imaging chain as well. That's okay. That can be
14 done.
15 I am saying, as a clinician, I feel more
16 comfortable with the feature analysis from a cost
17 standpoint. I think that the other is fine, but, in my
18 opinion, it is a little bit of overkill.
19 DR. DESTOUET: Dr. Toledano has told us that
20 actually we don't need 10,000 women or a huge cohort of
21 women to do an outcomes study. Am I wrong? Is there a
22 smaller number of women from whom we could get sensitivity
23 and specificity data? Have I misunderstood what you have
24 said?
25 DR. TOLEDANO: You haven't misunderstood. I want

1 to go back for a second to what Dr. Garra is saying. It is
2 two different levels of proof. The feature analysis gives
3 you the proof that the clinicians feel that they are seeing
4 the same information.
5 The level of being comfortable with that is that
6 the clinicians feel that if they know they are seeing the
7 same information, they can go and then perform at least as
8 well and that should be enough.
9 The other level of proof, the sensitivity and
10 specificity study, is on top of that. To the extent that it
11 is overkill, this is what we try to deal with, to not be
12 doing a 10,000 women or 100,000 women study. The people
13 from Trex claim that they have done such a study. There is
14 no problem with doing such a study. It is eminently doable.
15 You take the women who come to be screened for
16 analog. You take all of them who are scheduled for biopsy.
17 They get digital. You have some other rule for the women
18 who are not scheduled for biopsy. Some of them get digital.
19 Others of them are put on hold. 1,000 women, follow up
20 1,000. You don't need all 10,000.
21 If you know where they came from, you don't need
22 all 10,000. But that may be diagnostic more than screening.
23 We have to make sure that we are talking about the right
24 thing. You need to make sure that you have the appropriate
25 gold standard for your study. But screening study or

1 diagnostic study, you don't need 10,000 or 100,000 women to
2 have digital mammograms.
3 If the companies have several years' worth of
4 data, as long as they are honest and can tell you, "These
5 are all the people that we image," the issue of the
6 retrospective bias gets into who is going to pick and who is
7 going to disclose how many cases they have done.
8 If the companies are trying to get approved, which
9 I think they are, and if we are all supposed to be working
10 together, which I thought was the goal, that we all work
11 together, then you sit down and you say, "These are how many
12 we have imaged, and let's work with it from that standpoint
13 and not be trying to hide things."
14 I think digital has a lot of potential. I don't
15 think anyone is trying to hide failures of digital. So I
16 think we can work together and not have to worry so much
17 about the bias of retrospective selection. I think we can
18 do studies that don't involve 10,000 or 100,000 women.
19 DR. ALAZRAKI: Thank you. I would just like to
20 recognize Dr. Hendrick to make a very brief remark since he
21 was one of our speakers earlier.
22 DR. HENDRICK: One of the questions was could you
23 do an unbiased feature-analysis study. What we are actually
24 doing in the GE equivalence study is taking every woman who
25 comes in for diagnostic mammogram over the age of 40 who

1 consents to digital and doing digital mammogram at a number
2 of different sites.
3 That, I think, eliminates this bias of having to
4 have an ad-mixture of completely normal cases. It makes a
5 tougher population. In fact, it makes it ideal for doing a
6 feature-analysis study. But it takes every woman without
7 selection bias. So it does eliminate the potential for case
8 selection that could change the results.
9 If I could just respond to your point about having
10 to do an outcomes study not taking 10,000 women. You are
11 right; it doesn't take 10,000 women but it does require
12 follow up of all the women in your study. My understanding
13 of what was described by Trex is that they didn't follow up
14 the negative outcomes cases, so it is not a complete outcome
15 study. That why they called it some different name that I
16 hadn't heard before.
17 DR. TOLEDANO: Did Greg want to do that one?
18 DR. CAMPBELL: You said this twice, now. I don't
19 think you have to follow up everybody. You do have to
20 follow up a bunch of folks but you may not have to follow up
21 everybody which was one of the points that I think Alicia
22 had alluded to as well. In particular, you probably want to
23 follow up all the discrepancies between the digital and
24 analog and some fraction which may not be too large of the
25 double positives and the double negatives.

1 DR. HENDRICK: It is just that, in the real world,
2 I think radiologists understand that it is hard enough to
3 follow up the positive findings that you have if you are
4 going to require follow up of every positive case on digital
5 or film screening.
6 DR. ALPERT: This is not the real world. This is
7 a clinical trial. We are not in the real world yet. I am
8 putting the comment on the table very importantly while you
9 are at the mike because I think that is very important. The
10 issue that we are dealing with is developing a set of data
11 to support a marketing authorization.
12 We are not talking about using this as the
13 mechanism for monitoring once this is already introduced
14 unless there is a question we have to answer postmarket.
15 Again, this is something we face in FDA all of the time,
16 people telling us that they can't develop the data even in a
17 study setting.
18 I have to answer that studies are, in fact, for
19 the purpose of getting information that you don't normally
20 get in the real world. That is why they are constructs and
21 not just going in and doing regular practice. They are
22 really special and patients sign consent to be in them and
23 understand that things are going to be done to them that are
24 not being done in the real world.
25 DR. HENDRICK: I don't disagree with that.

1 DR. DESTOUET: I have a question. I am not sure
2 which subset of patients Dr. Hendrick is referring to that
3 we don't follow because MQSA guidelines require that we
4 follow up on all positive mammograms.
5 The only people we don't follow up are on the
6 negative mammograms. Is that what we are talking about?
7 DR. HENDRICK: Yes.
8 DR. ALPERT: Negative film screen, positive
9 digital, which are not patients that would otherwise be
10 followed.
11 DR. HENDRICK: Mostly, I am talking about negative
12 film screen, negative digital. I think you will get some of
13 the negative film screen, positive digital. But even in a
14 study situation, even in the best of situations, you won't
15 get all of them.
16 DR. ALPERT: We are getting off the track, but our
17 assumption is that, in any clinical trial, if 80 percent of
18 the patients actually finish the trial, we have got a
19 reasonably valid trial. We worry when we lose a lot more
20 than that. We would like to get 90 percent of the patients
21 entering the trial actually being followed for their
22 complete trial, but we know that there are losses. We are
23 not expecting perfection.
24 DR. ALAZRAKI: We are going to have to wind up now
25 because there are a number of panel members that have to

1 leave at 4 o'clock.

2 I would like to just give all of the panel members

3 an opportunity, if they want, to make any final statement

4 here. We will start, Brian, with you, if you have anything

5 to say.

6 DR. GARRA: I have probably said enough. I like

7 the idea of doing the outcome study. I think that would be

8 a really nice study. I am just not sure that we need this

9 to move ahead with this specific modality. The feature-

10 analysis study, I can see a lot of little problems that were

11 alluded to here that we need some careful, careful selection

12 of what features, whether you blind the viewers and all

13 these things. But I think the feature analysis alone is the

14 most cost-effective approach, but, certainly, the outcome

15 study would be a valuable way to go as well.

16 DR. TOLEDANO: I have certainly said quite a lot.

17 I guess the one thing I would say is can we please call it

18 something other than outcomes study because we are not

19 exactly doing outcomes analysis. Call it the better study.

20 No; call it the sensitivity and specificity study.

21 I think Dr. Garra wrapped things up very nicely

22 with his comments so I don't need to repeat them and I don't

23 need to repeat anything that I have said all day. I am done

24 talking and it is somebody else's turn.

25 DR. GRIEM: I would like to back up briefly. The

1 question of taking of a phantom and doing some preliminary
2 work on the phantom, to do a rough analysis before moving
3 in, and then I think the feature analysis would be quite
4 satisfactory.
5 DR. STERNICK: No comments.
6 MS. WHELAN: No comments.
7 DR. MONSEES: No additional comments.
8 DR. BERG: One comment. I think the public
9 probably would really like us to do the non-inferiority
10 study at least to have some sense of sensitivity of this
11 technology before we approve it for marketing.
12 I think, furthermore, in doing that, as has been
13 said and alluded to, that we will have a subset of cases
14 that could then be subjected to the feature analysis to make
15 radiologists happy. I think we have two different groups
16 that we are really trying to address.
17 I also think, finally, that feature analysis, we
18 need to clarify that it is really diagnostic not screening
19 when we are doing feature analysis. No one is going to try
20 to laboriously characterize mass borders on the basis of the
21 screening mammograms. One needs a magnification view to do
22 that.
23 Similarly, with calcifications. So I think we
24 need to be a little careful in our chronology there. I feel
25 there is a need for both.

1 DR. BEAM: The sensitivity and specificity study
2 could be called a diagnostic-accuracy study. That's one.
3 Number two, I still haven't heard an answer to my question
4 about whether or not level 1 evidence is sufficient. I
5 would hope that somebody at the FDA would make a decision
6 there.
7 DR. ROMILLY-HARPER: Not to much comment. I
8 second Dr. Garra's comments in that, to me, a feature-
9 analysis study is all I would like to see to get it on the
10 market and maybe look at a more intense study down the road
11 where we have large numbers of and we have, maybe, all the
12 manufacturers involved in a significantly large study that
13 would give us much more information.
14 DR. DESTOUET: I think it is very difficult to ask
15 the manufacturers to put this off much longer. I know that
16 the FDA really would like a tremendous amount of data. I
17 think those of us in practice recognize that there are clear
18 advantages to digital breast imaging even in the setting
19 that we are using now and I would anticipate that whole-
20 breast digital imaging would provide us much more
21 information in the proper setting.
22 I also agree with that panel members that feature
23 analysis would provide at least some sense of equivalency.
24 I know that is not very good, but at least some sense of
25 equivalency would probably--not to the level of the FDA's

1 full liking but at least to the level at which we could say
2 that this technology can be introduced to the market with
3 imaging capability that is as good as or, perhaps, better
4 than analog images.
5 Then, with postmarket surveillance, as Dr. Harper
6 has pointed out, we may be able to, indeed, get a better
7 handle on sensitivity and specificity in a much larger
8 population study, one that we probably could not require the
9 industry to produce now.
10 DR. ALAZRAKI: I would like to say that I also
11 think the feature-analysis study is the way to go at this
12 point weighing all of the issues of cost and time. I think
13 the clinical radiology community would be satisfied with a
14 nicely designed feature-analysis study.
15 Also, I would like to point out that some of the
16 other benefits of the digital imaging, the archiving, the
17 filmless--saving expense of film, and file-room costs and
18 all of this, and losing films and all of that, that all of
19 those benefits are substantial, and some acknowledgment of
20 those benefits should be, in some way, factored in to the
21 FDA's deliberations.
22 With that, unless the FDA wants to, in five
23 minutes, tell us something else.
24 DR. ALPERT: How about in one. I would just like
25 to thank everyone who attended, both in the audience who

1 came to the microphone, who just listened and will take
2 their messages home as well, and particularly to all of you
3 who sat at the panel and helped us.
4 I appreciate that you now understand what
5 difficulties we have in breaking ground with new
6 technologies. I heard that from a number of you during the
7 breaks. I really do want to thank you for helping us
8 because the more discussion we have on these issues and the
9 more we challenge each other on this understanding and these
10 issues, the better the result will be in all of us
11 understanding what we get and what we don't get in certain
12 assessments of new technologies and what we get and what we
13 don't get as products enter the marketplace.
14 That is extremely important for all of us, so I do
15 sincerely thank you for participating with us. I warn you
16 as well that, as we firm this up, you will be hearing from
17 us for more comment.
18 DR. ALAZRAKI: Mr. Doyle, you have some
19 announcements?
20 MR. DOYLE: Yes, just one. To help us put this
21 record together, and, as you know, there have been a lot of
22 comments and speakers, I would appreciate it if each one of
23 the speakers could send me a copy of his or her viewgraphs
24 or if they have any text of their presentation. It will
25 just help us to put a record of this meeting together and

1 get everything down in an accurate fashion.

2 DR. ALAZRAKI: Also, I would just like to thank

3 the panel and the lead discussants and all of the people who

4 worked hard to review all of the material that was sent out

5 to all of us.

6 I hereby adjourn the meeting.

7 [Whereupon, at 3:55 p.m., the proceedings were

8 adjourned.]

CERTIFICATE

I, **ALICE TOIGO,** the Official Court Reporter for Miller Reporting Company, Inc., hereby certify that I recorded the foregoing proceedings; that the proceedings have been reduced to typewriting by me, or under my direction and that the foregoing transcript is a correct and accurate record of the proceedings to the best of my knowledge, ability and belief.

Alice Toigo

ALICE TOIGO