

UNITED STATES OF AMERICA
 DEPARTMENT OF HEALTH AND HUMAN SERVICES
 FOOD AND DRUG ADMINISTRATION

+ + +

CENTER FOR DEVICES AND RADIOLOGICAL HEALTH
 MEDICAL DEVICES ADVISORY COMMITTEE

+ + +

RADIOLOGICAL DEVICES PANEL

+ + +

November 18, 2009
 8:00 a.m.

Holiday Inn Gaithersburg
 2 Montgomery Village Avenue
 Gaithersburg, Maryland

PANEL MEMBERS:

CARL J. D'ORSI, M.D.	Chair
CRAIG K. ABBEY, Ph.D.	Voting Member
THOMAS J. PAYNE, Ph.D.	Voting Member
BHARAT MITTAL, M.D.	Voting Member
ROBERT D. ROSENBERG, M.D.	Voting Member
MARVIN C. ZISKIN, M.D.	Voting Member
LORI DODD, Ph.D.	Temporary Voting Member
LEONARD GLASSMAN, M.D.	Temporary Voting Member
J. ANTHONY SEIBERT, Ph.D.	Temporary Voting Member
GEORGIA TOURASSI, Ph.D.	Temporary Voting Member
OTTO LIN, M.D.	Temporary Voting Member
YULEI JIANG, Ph.D.	Temporary Voting Member
DANIEL SWERDLOW, M.D.	Temporary Voting Member
KENNETH STEIER, M.D.	Temporary Voting Member
A. MARILYN LEITCH, M.D.	Temporary Voting Member
DANIEL BOURLAND, Ph.D.	Temporary Voting Member
DAVID KIM, M.D.	Temporary Voting Member
JOHN A. CARRINO, M.D., MPH	Temporary Voting Member
XIAO-HUA ANDREW ZHOU, Ph.D.	Temporary Voting Member

Free State Reporting, Inc.
 1378 Cape Saint Claire Road
 Annapolis, MD 21409
 (410) 974-0947

PANEL MEMBERS (Cont.):

GARY L. DUEHRING, Ph.D.
ROBERT UZENOFF

Consumer Representative
Industry Representative

JAMES P. SWINK

Designated Federal
Officer

FDA REPRESENTATIVES:

JANINE M. MORRIS
Acting Division Director, Division of
Reproductive, Abdominal and Radiological Devices

ROBERT OCHS

PEPER LONG
Press Contact

FDA PRESENTERS:

JOYCE M. WHANG, Ph.D.
NICHOLAS A. PETRICK, Ph.D.
THOMAS E. GWISE, Ph.D.
ROBERT C. SMITH, M.D., J.D.
CARA J. KRULEWITCH, CNM, Ph.D., FACNM

PUBLIC SPEAKERS:

JOHN DELUCA, M.D.
MARYELLEN GIGER, Ph.D.
MARGARITA ZULEY, M.D.
ROBERT NISHIKAWA, M.D.
STEPHEN VASTAGH
STEPHEN SLAVENS
TERRENCE SWEENEY
JULIAN MARSHALL
MORGAN NIELDS

INDEX

	PAGE
CALL TO ORDER - Carl J. D'Orsi, M.D.	6
FDA Conflict of Interest Disclosure Statement - James P. Swink	7
General Announcements - James P. Swink	11
Panel Introductions	12
FDA PRESENTATIONS	
Introduction - Joyce M. Whang, Ph.D.	16
Considerations on the Assessment of CADe Devices - Nicholas A. Petrick, Ph.D.	36
Statistical Issues in CADe Evaluations - Thomas E. Gwise, Ph.D.	75
Clinical Viewpoint - Robert C. Smith, M.D., J.D.	99
Post-Approval Consideration - Cara J. Krulewitch, CNM, Ph.D., FACNM	135
QUESTIONS FOR FDA SPEAKERS	141
OPEN PUBLIC HEARING	
John Deluca, M.D., iCAD	192
Maryellen Giger, Ph.D.	197
Margarita Zuley, M.D., ACR SBI	203
Robert Nishikawa, M.D.	207
Stephen Vastagh, MITA	210
Stephen Slavens, GE Healthcare	214
Terrence Sweeney, Philips Healthcare	218
Julian Marshall, Hologic	222
Morgan Niels, Bioptics	225
QUESTIONS FOR OPEN PUBLIC HEARING SPEAKERS	227

INDEX (Cont.)

PAGE

PANEL DISCUSSION AND QUESTIONS TO THE PANEL

Question 1	241
Question 2	259
Question 2(b)	259
Question 5(a)	297
Question 2(b)(iv)	297
Question 3(c)	301
Question 3(d)	302
Question 4	305
Question 4(a)	306
Question 4(c)	311
Question 4(d)	318
Question 5(b)	319
Question 3	342
Question 3(a)	343
Question 3(e)	358
Question 2(a)	377
Question 3(b)	400
Question 6	404

ADJOURNMENT	407
-------------	-----

M E E T I N G

(8:00 a.m.)

DR. D'ORSI: I would like to call this meeting of the Radiological Devices Panel to order.

I'm Dr. Carl D'Orsi, the Chairperson of this Panel. My area of expertise is in breast imaging and technology assessment. I'm Professor of Radiology and Hematology and Oncology at Emory University and director of the breast imaging section.

Mr. Swink, the Designated Federal Officer for today's Radiological Devices Panel, will make some introductory remarks.

MR. SWINK: Good morning, everyone. During the second day of this two-day Radiological Devices Panel meeting, we are seeking input from the Panel on two product areas and their associated guidance document. Specifically, we are looking for the Panel's clinical and scientific viewpoint on how best the Agency can determine a reasonable assurance of safety and effectiveness for this product with respect to the type and amount of data needed to make that assessment.

As part of the Agency's effort to encourage open discussion and increased transparency, we are

Free State Reporting, Inc.
1378 Cape Saint Claire Road
Annapolis, MD 21409
(410) 974-0947

1 providing time to allow the presentation of more than
2 one scientific viewpoint from within the Agency. The
3 scientific reviewers presenting today will provide
4 their views on the topic, and you may find that they
5 may differ in some aspects. It should be noted that
6 the viewpoints expressed today do not necessarily
7 represent a consensus opinion within the FDA or align
8 with existing Agency policy. With the Panel's
9 additional input, the Agency will apply the
10 scientific and clinical recommendations to the
11 regulatory framework set out by the statute and
12 regulations.

13 I'll now read the Conflict of Interest
14 Statement for today.

15 The Food and Drug Administration is
16 convening today's meeting of the Radiological Devices
17 Panel of the Medical Devices Advisory Committee under
18 the authority of the Federal Advisory Committee Act
19 of 1972. With the exception of the industry
20 representative, all members and consultants of the
21 Panel are special Government employees or regular
22 Federal employees from other agencies and are subject
23 to Federal conflict of interest laws and regulations.

24 The following information on the status of
25 this Panel's compliance with Federal ethics and

Free State Reporting, Inc.
1378 Cape Saint Claire Road
Annapolis, MD 21409
(410) 974-0947

1 conflict of interest laws covered by, but not limited
2 to, those found at 18 U.S.C. Section 208 and Section
3 712 of the Federal Food, Drug and Cosmetic Act are
4 being provided to participants in today's meeting and
5 to the public.

6 FDA has determined that members and
7 consultants of this Panel are in compliance with the
8 Federal ethics and conflict of interest laws. Under
9 18 U.S.C. Section 208, Congress has authorized FDA to
10 grant waivers to special Government employees who
11 have financial conflicts when it is determined that
12 the Agency's need for a particular individual's
13 services outweighs his or her potential financial
14 conflict of interest. Under Section 712 of the FD&C
15 Act, Congress has authorized FDA to grant waivers to
16 special Government employees and regular Government
17 employees with potential financial conflicts when
18 necessary to afford the Committee essential
19 expertise.

20 Related to the discussions of today's
21 meeting, members and consultants of this Panel who
22 are special Government employees have been screened
23 for potential financial conflicts of interest of
24 their own as well as those imputed to them, including
25 those of their spouses or minor children and, for

1 purposes of 18 U.S.C. Section 208, their employers.
2 These interests may include investments; consulting;
3 expert witness testimony; contracts/grants/CRADAs;
4 teaching/speaking/writing; patents and royalties; and
5 primary employment.

6 For today's agenda, the Panel will discuss
7 and make recommendations regarding the Agency's
8 regulatory strategy for computer-assisted detection
9 devices. CAD devices are devices intended to
10 identify, mark, highlight, or in any other manner
11 direct attention to potential abnormalities revealed
12 in the radiological data of the human body or imaging
13 device data during interpretation of patient images
14 or patient imaging data by a physician or other
15 healthcare professionals.

16 The Panel will discuss two draft guidance
17 documents entitled "Computer-Assisted Detection
18 Devices Applied to Radiology Images and Radiology
19 Device Data--Pre-market Notification [510(k)]
20 Submissions" and "Clinical Performance Assessment:
21 Considerations for Computer Assisted Detection
22 Devices Applied to Radiology Images and Radiology
23 Device Data--Pre-market Approval (PMA) and Pre-market
24 Notification [510(k)] Submissions." This is a
25 particular matter of general applicability.

Free State Reporting, Inc.
1378 Cape Saint Claire Road
Annapolis, MD 21409
(410) 974-0947

1 Based on the agenda for today's meeting and
2 all the financial interests reported by the Panel
3 members and consultants, a conflict of interest
4 waiver has been issued in accordance with 18 U.S.C.
5 Section 208 and Section 712 of the FD&C Act to
6 Dr. John A. Carrino. Dr. Carrino's waiver addresses
7 an anticipated research grant with an affected firm
8 at issue. For his services, he anticipates between
9 \$5,001 to \$10,000, but the exact terms and amount
10 have not been finalized. This waiver allows the
11 individual to participate fully in today's
12 deliberations.

13 FDA's reasons for issuing the waiver are
14 described in the waiver document, which is posted on
15 FDA's website at www.fda.gov. Copies of this waiver
16 may also be obtained by submitting a written request
17 to the Agency's Freedom of Information Office,
18 Room 6-30 of the Parklawn Building. A copy of this
19 statement will be available for review at the
20 registration table during this meeting and will be
21 included as part of the official transcript.

22 Robert Uzenoff is serving as the Industry
23 Representative, acting on behalf of all related
24 industry, and is employed by Fujifilm Medical
25 Systems, Incorporated.

Free State Reporting, Inc.
1378 Cape Saint Claire Road
Annapolis, MD 21409
(410) 974-0947

1 We would like to remind members and
2 consultants that if the discussions involve any other
3 products or firms not already on the agenda for which
4 the FDA participant has a personal or imputed
5 financial interest, the participants need to exclude
6 themselves from such involvement and their exclusion
7 will be noted for the record. FDA encourages all
8 other participants to advise the Panel of any
9 financial relationships they may have with any firms
10 at issue.

11 Now, I have a few general announcements.
12 If you have not already done so, please sign the
13 attendance sheets that are at the registration
14 outside the door. Transcripts of today's meeting
15 will be available from Free State Court Reporting,
16 Incorporated. They may be reached at (410) 974-0947.
17 Information on purchasing videos of today's meeting
18 can be found on the table outside of the meeting
19 room.

20 I would like to remind everyone that
21 members of the public and press are not permitted
22 around the Panel area, which is the area beyond the
23 speaker's podium. The press contact for today's
24 meeting is Peper Long. I would request that
25 reporters please wait to speak to FDA officials until

Free State Reporting, Inc.
1378 Cape Saint Claire Road
Annapolis, MD 21409
(410) 974-0947

1 after the Panel meeting has concluded.

2 If you are presenting in the Open Public
3 Hearing session today and have not previously
4 provided an electronic copy of your slide
5 presentation to the FDA, please arrange to do so with
6 Ms. AnnMarie Williams at the registration table.

7 And I would ask you to please silence your
8 cell phones and other electronic devices at this
9 time.

10 Thank you.

11 DR. D'ORSI: Thank you. At this meeting
12 the Panel will discuss and make recommendations
13 concerning the clinical performance assessment of
14 computer-assisted detection devices applied to
15 radiology images and radiology device data and CADE-
16 related pre-market notification 510(k) submissions.

17 Before we begin, I would like to ask our
18 Panel members and FDA staff seated at the table to
19 introduce themselves. State your name, your area of
20 expertise, your position and affiliation, and we'll
21 start at the left with Mr. Uzenoff.

22 MR. UZENOFF: My name is Bob Uzenoff. I
23 work for Fujifilm Medical Systems U.S.A. My area of
24 expertise is image quality, quality control. I work
25 in regulatory affairs, and I am the Industry

Free State Reporting, Inc.
1378 Cape Saint Claire Road
Annapolis, MD 21409
(410) 974-0947

1 Representative for this meeting.

2 DR. DUEHRING: I'm Gary Duehring, and I'm
3 the Consumer Rep for this meeting.

4 DR. CARRINO: Hi, I'm John Carrino, a
5 musculoskeletal radiologist at Johns Hopkins
6 University, and my area of expertise is healthcare
7 informatics.

8 DR. KIM: My name is David Kim. I'm at the
9 University of Wisconsin. My area of research is in
10 CT colonography.

11 DR. BOURLAND: I'm Dan Bourland, Wake
12 Forest University, Winston-Salem, North Carolina, and
13 I'm a radiation therapy physicist with interest in
14 imaging and radiation oncology.

15 DR. SEIBERT: Tony Seibert, Professor of
16 Radiology at the University of California-Davis in
17 Sacramento, California. I'm a medical imaging
18 physicist with an interest in detector evaluation and
19 imaging research.

20 DR. LEITCH: I'm Marilyn Leitch. I'm a
21 surgical oncologist and professor of surgery at UT
22 Southwestern Medical Center in Dallas. My primary
23 area of practice is in breast cancer.

24 DR. LIN: My name is Otto Lin. I'm a
25 gastroenterologist at Virginia Mason Medical Center

1 in Seattle; also a clinical associate professor of
2 medicine at the University of Washington. My area of
3 interest in colon cancer screening.

4 DR. ABBEY: I'm Craig Abbey. I'm a
5 researcher in the Department of Psychology at UC-
6 Santa Barbara. I also have an appointment in
7 biomedical engineering at UC-Davis, and my area of
8 interest is observer performance studies.

9 DR. ROSENBERG: I'm Robert Rosenberg,
10 Professor of Radiology, University of New Mexico. My
11 area of expertise is in mammography and outcomes in
12 the community.

13 DR. ZISKIN: I'm Marvin Ziskin. I'm
14 Professor of Radiology and Medical Physics at Temple
15 University in Philadelphia, and I'm the Director of
16 the Center for Biomedical Physics. My interests are
17 rather broad, including ultrasound, electromagnetic
18 effects, and image processing.

19 DR. MITTAL: My name is Bharat Mittal. I'm
20 Professor and Chairman of Radiation Oncology at
21 Northwestern University. My area of interest is user
22 radiation, radiation oncology, and treatment of head
23 and neck cancer.

24 DR. PAYNE: I'm Tom Payne. I'm a medical
25 physicist. I'm from Minneapolis, Minnesota. I

1 participate in all areas of medical physics. My
2 primary interest at the moment is in mammography and
3 CT.

4 DR. ZHOU: My name is Andrew Zhou. I'm a
5 Professor in the Department of Biostatistics for the
6 University of Washington and Director of the
7 biostatistics unit at the VA hospital over there. My
8 expertise is in statistical masses in diagnostic
9 medicine and health services research.

10 DR. DODD: My name is Lori Dodd. I'm a
11 biostatistician at the National Institute of Allergy
12 and Infectious Diseases. I was formerly at NCI for
13 seven and a half years, where I worked with the
14 cancer imaging program, and my area of expertise is
15 really in medical diagnostic testing.

16 DR. GLASSMAN: I'm Len Glassman, and I'm
17 the head of breast imaging at the Armed Forces
18 Institute of Pathology and in the private practice of
19 radiology in Washington, D.C. My research interest
20 is radiologic-pathologic correlation in breast
21 disease.

22 DR. TOURASSI: My name is Georgia Tourassi.
23 I'm Associate Professor of Radiology and Medical
24 Physics at Duke University Medical Center, and my
25 area of expertise is CAD technology, mainly in breast

1 imaging.

2 DR. JIANG: I'm Yulei Jiang. I'm Associate
3 Professor of Radiology and Committee on Medical
4 Physics at the University of Chicago. My area of
5 interest is in computer-related diagnosis of breast
6 cancer and prostate cancer in both radiologic images
7 and pathology images. I'm mostly interested in ROC
8 analysis and observer studies.

9 DR. SWERDLOW: I'm Dan Swerdlow. I'm
10 Assistant Professor of Radiology at Georgetown
11 University. My areas of interest are CT colonography
12 and imaging-guided interventional procedures.

13 DR. STEIER: Good morning. I'm Ken Steier.
14 I'm a pulmonary and critical care physician in
15 private practice in Long Island, New York.

16 MS. MORRIS: Janine Morris, the Acting
17 Division Director for the Division of Reproductive,
18 Abdominal and Radiological Devices. I've been with
19 the Agency for over 19 years.

20 DR. D'ORSI: Thank you. We'll now begin
21 with the FDA presentations. The first FDA presenter
22 is Dr. Joyce Whang.

23 DR. WHANG: Good morning. Thank you for
24 joining us today for the second day of our two-day
25 meeting of the Radiology Advisory Panel. Today we

Free State Reporting, Inc.
1378 Cape Saint Claire Road
Annapolis, MD 21409
(410) 974-0947

1 will be discussing two draft guidance documents
2 pertaining to computer-assisted detection devices for
3 radiological images. These devices will be referred
4 to as CAD devices or CADe devices. These guidance
5 documents grew from the two-day meeting that we had
6 in March 2008 with this Panel. Many of you
7 participated in that meeting, and we thank you for
8 your continued involvement, and we hope that you find
9 these guidance documents to be a good reflection of
10 the discussion we had in March.

11 The two draft guidance documents are
12 intended to be somewhat general in their
13 applicability. The first provides the general
14 expectations for 510(k) submissions for CAD devices
15 used for radiology images. The second focuses on the
16 clinical studies that may be used to support CAD
17 devices, both 510(k)'s and PMAs. These guidance
18 documents are not focused on specific clinical
19 applications as was much of our discussion in the
20 March 2008 meeting.

21 We hope that they capture the expectations
22 that will be applicable to various clinical
23 applications that exist now and in the future. In
24 your discussion today, we will ask you for your
25 scientific input regarding these guidance documents

1 and how they should be applied to specific possible
2 applications. These draft guidance documents were
3 issued on October 21st, and they are open for public
4 comment for 90 days. That takes us into mid-January.
5 They do not include sections of frequently asked
6 questions, F-A-Q's, or FAQs, but we are considering
7 adding FAQs in response to today's discussion and
8 other input that we received during the public
9 comment period.

10 Here's what I will be speaking about today.
11 First, I will review some of the key regulatory
12 concepts. For those of you who were here yesterday,
13 there's going to be some review, but I will try to
14 package them in a way that's focused on today's
15 subject matter. Then I will describe the types of
16 CAD devices that are covered by the guidance
17 documents, and I will provide summary of the
18 recommendations provided regarding the contents of
19 510(k) submissions, and I will summarize the
20 recommendations provided regarding clinical studies
21 that may support 510(k)'s or PMAs.

22 Before a CAD device can be marketed in the
23 U.S., it must go through either the 510(k) process or
24 the PMA process. Designation of 510(k) or PMA may be
25 related to the risks associated with the use of the

1 device or the presence of a predicate 510(k) device
2 with which the new device can be shown to be
3 substantially equivalent. 510(k) devices are usually
4 Class II devices, which means that the devices have
5 been considered to be moderate-risk devices for which
6 safety and efficacy can be ensured via some
7 combination of what we call general controls and
8 special controls. To be cleared for marketing, a
9 510(k) device must be shown to be at least as safe
10 and effective, that is, substantially equivalent, to
11 a legally marketed 510(k) device. For the CAD
12 devices that are 510(k), the predicates are usually
13 other 510(k) devices, although they could potentially
14 be devices that have been reclassified from Class III
15 to Class II.

16 At this time we have colon CAD devices and
17 lung CAD devices that are 510(k). This does not mean
18 that all lung CAD and colon CAD devices will be
19 subject to 510(k). It will depend on a comparison of
20 the intended uses and the technological
21 characteristics.

22 So how do we make that comparison? As we
23 discussed yesterday, for a new device to be cleared
24 through the 510(k) process, it has to be found
25 substantially equivalent to a predicate device. This

1 may occur if the two devices have the same intended
2 use and the same technological characteristics.

3 Now, a new CAD device usually represents
4 the implementation of some new software, and FDA
5 expects that new CAD devices will have different
6 technological characteristics from the legally
7 marketed predicate devices. So how do we consider
8 those different technological characteristics? If we
9 consider an example that came up yesterday, if we
10 talk about a new car, a new car with power windows,
11 it has the same intended use as the existing car.
12 It's going to get us from point A to point B. But
13 it's got these new technological characteristics,
14 these power windows.

15 So what we ask is do the new technological
16 characteristics raise new types of questions of
17 safety and effectiveness? So we think to ourselves,
18 well, what questions are raised by these power
19 windows? If all we care about is do the windows go
20 up and down, and we've asked that about the other
21 cars, that is not a new type of question. So the
22 device can still be eligible for 510(k).

23 Now, let's say along with these power
24 windows, we have some kind of virtual window that you
25 can't see, but it's there. It's supposed to keep the

1 wind out when you're driving. But we're wondering
2 what happens if you have an accident. What sort of
3 safety does this provide? If this is not a question
4 that we've considered with previous cars and not a
5 type of question we've considered before, then it's
6 not going to be eligible for 510(k).

7 So when a device has new types of
8 technological characteristics, the new technological
9 characteristics that do not raise new types of
10 questions of safety and effectiveness, the company
11 may need to provide data, scientific and/or clinical.
12 We're just going to demonstrate that the new or the
13 changed device is as safe and effective as a legally
14 marketed predicate device and that there are no new
15 types of questions of safety and effectiveness
16 related to the predicate device, or in comparison to
17 the predicate device.

18 Now, if the new technological
19 characteristics do raise new types of questions, the
20 company gets what we call an NSE letter: not
21 substantially equivalent. The default when something
22 is considered not substantially equivalent for this
23 reason is that the device was to be considered a PMA.
24 It would require a PMA as a Class III device. Now,
25 the company has the option of requesting that this

1 device be identified as de novo. De novo has a
2 specific regulatory meaning, a de novo 510(k). It
3 means that the risks can be well enough defined to be
4 controlled -- to be managed with the special controls
5 guidance document.

6 So the maker of the car with the virtual
7 window, who got our NSE letter, they could go ahead
8 and just work through the PMA process or they could
9 come back to us and say, yeah, but the risks involved
10 in this virtual window, look, they're X, Y, and Z,
11 and they can be handled with A, B, and C. So you
12 could do a special controls guidance document. It
13 could still be a Class II device.

14 Okay, there are also CAD devices that are
15 Class III devices. Oh, actually, I should highlight
16 that CAD -- among the CAD devices for Class II
17 devices, we have the colon and the lung CAD devices.
18 I think I said that. Okay, there are also CAD
19 devices that are Class III devices. Class III
20 devices that support or sustain human life, that are
21 of substantial importance in preventing impairment of
22 human health, and that present a potential,
23 unreasonable risk of injury or illness.

24 CAD III devices require a PMA to
25 demonstrate a reasonable assurance of safety and

1 effectiveness. The approved mammography CAD devices
2 are all Class III, as is one of the lung CAD devices.
3 In discussion Question 8, we'll be asking the Panel
4 to discuss the relative risks of different types of
5 CAD devices that may affect the appropriate
6 classification.

7 Note that neither 510(k)'s nor PMAs require
8 an absolute certainty about the level of safety and
9 effectiveness of the device. For 510(k)'s, the level
10 of safety and effectiveness must be substantially
11 equivalent to the predicate device. For PMA, the
12 level of safety and effectiveness must stand on its
13 own to provide a reasonable assurance of safety and
14 effectiveness.

15 However, while the requirements for the
16 level of safety and effectiveness may vary, the type
17 of testing -- for the type of testing that's needed,
18 there are similar options for testing 510(k) and PMA
19 devices. If clinical data are needed, there are
20 similar choices for control arms. In general, there
21 are some devices that are tested without a control
22 arm with studies that use objective performance
23 criteria.

24 More often, a clinical study will require a
25 control arm. The control arm might be a similar

1 device, for example, a predicate for a 510(k) device
2 or the control arm could be some other standard of
3 care. For example, unassisted reading might be an
4 appropriate control arm for CAD devices. Whether a
5 device is 510(k) or PMA, there are no inherent
6 requirements as to what the control has to be. There
7 will be further discussion of this in the
8 presentations to come and in discussion Question
9 2(b). 2(a).

10 Now, I'll provide the regulatory
11 definitions of some of the terms I've been using.

12 Safety is described in the Code of Federal
13 Regulations as follows: "There is a reasonable
14 assurance that a device is safe when it can be
15 determined, based upon valid scientific evidence,
16 that the probable benefits to health from use of the
17 device for its intended use and conditions of use,
18 when accompanied by adequate directions and warnings
19 against unsafe use, outweigh any probable risks."

20 For effectiveness: "There is a reasonable
21 assurance that a device is effective when it can be
22 determined, based upon valid scientific evidence,
23 that in a significant portion of a target population,
24 the use of the device for its intended use and
25 conditions of use, when accompanied by adequate

1 directions for use and warnings against unsafe use,
2 will provide clinically significant results."

3 You've just heard that safety and
4 effectiveness must be determined utilizing valid
5 scientific evidence. The definition of valid
6 scientific evidence starts with examples of where
7 such evidence could come from. "Valid scientific
8 evidence is evidence from well-controlled
9 investigations, partially controlled studies, studies
10 and objective trials without matched controls, well-
11 documented case histories conducted by qualified
12 experts, and reports of significant human experience
13 with a marketed device from which it can fairly and
14 responsibly be concluded by qualified experts that
15 there is a reasonable assurance of the safety and
16 effectiveness of a device under its conditions of
17 use.

18 "Isolated case reports, random experience,
19 reports lacking sufficient details to permit
20 scientific evaluation, and unsubstantiated opinions
21 are not regarded as valid scientific evidence to show
22 safety or effectiveness."

23 When considering what level of evidence is
24 needed to support a PMA or a 510(k), FDA applies the
25 least burdensome approach, as provided in the Food,

1 Drug and Cosmetic Act as amended by FDAMA in 1997, by
2 two provisions that are commonly referred to as the
3 least burdensome provisions.

4 Here's the provision that focuses on
5 clinical data for PMAs. It states that FDA can only
6 request the clinical data needed to establish device
7 effectiveness and that FDA will work with companies
8 to find the least burdensome methods for evaluating
9 device effectiveness in a way that is sufficient for
10 approving the device. Similarly for 510(k)'s, FDA
11 can only request information that is necessary, and
12 FDA must consider the least burdensome means of
13 demonstrating substantial equivalence.

14 The term least burdensome has been
15 interpreted by the Agency as a means of addressing a
16 pre-market issue through what amounts to the most
17 appropriate investment of time, effort, and resources
18 on the part of industry and the FDA. When the least
19 burdensome concept is conscientiously applied, it
20 should help to expedite the availability of new
21 device technologies without compromising scientific
22 integrity and the decision-making process or FDA's
23 ability to protect health. The least burdensome
24 concept should be applied to all pre-market
25 activities as well as post-market activities as they

1 relate to the pre-market arena.

2 Now, I'm going to switch gears and focus on
3 the draft guidance documents. First I'm going to
4 discuss the types of CAD devices covered by the
5 proposed guidance documents. Both guidance documents
6 pertain to computer-assisted detection devices
7 applied to radiology images and radiology device
8 data. Radiological data include those that are
9 produced during patient examination with X-ray, CT,
10 ultrasound, MRI, PET, et cetera.

11 By computer-assisted detection, we mean
12 computerized systems that incorporate pattern
13 recognition and data analysis capabilities to
14 identify, mark, highlight, or in any other manner
15 direct attention to portions of an image or aspects
16 of radiology device data that may reveal
17 abnormalities when the patient radiology images or
18 radiology device data are being interpreted by the
19 physician or other healthcare professional.

20 Here are examples of things that CAD
21 devices may be designed to identify and prompt. On a
22 mammogram, the CAD device might be designed to
23 identify possible microcalcification clusters and
24 masses. A CAD might be intended to highlight colonic
25 polyps on CTs or filling defects on thoracic CT or

1 brain lesions on MRI.

2 This does not include devices that are
3 intended for use during intraoperative procedures,
4 computer-assisted diagnostic devices, we call them
5 CADx, or computer triage devices. For example, a
6 CADx might be one that is intended to assess the
7 likelihood of the presence or absence of disease, or
8 a device intended to specify disease type, severity,
9 stage, or intervention recommended. A device
10 intended to diagnose Alzheimer's from MRI would not
11 be covered by this guidance.

12 A computer triage device is one that
13 reduces or eliminates some aspect of the clinical
14 care that is currently provided by a clinician, for
15 example, a device that indicates that some patients
16 are normal and therefore their radiological data do
17 not need to be interpreted by a clinician. These
18 devices are not covered by the guidance document.

19 The first guidance document is focused on
20 the information that should be provided for 510(k)
21 applications for Class II CAD devices. I will
22 provide some introduction to each section of this
23 guidance.

24 First, a submission should provide general
25 information about the device, including the target

Free State Reporting, Inc.
1378 Cape Saint Claire Road
Annapolis, MD 21409
(410) 974-0947

1 population. That would include the patient
2 population, organs of interest, diseases, conditions,
3 and abnormalities of interest; also, the type of
4 radiological data that are to be used with the
5 device, which would include imaging modalities, for
6 example, CT or MR, and if there are specific systems
7 or image acquisition parameters; also how the device
8 will be used in the clinical workflow. A 510(k)
9 application would also be expected to include
10 information about the algorithm, any processing it
11 does of images, how features are identified and
12 selected, types of models and classifiers used, and
13 how the algorithm is trained.

14 Then there is additional information
15 pertinent to the evaluation of the device. First,
16 there are the databases that are used to train and
17 test the devices. These will be discussed further in
18 the upcoming presentations. Then there is ground
19 truth, how it was determined, and finally, how was
20 the CAD scored? By scoring, we mean when you're
21 testing a device, how close to the disease does the
22 mark have to be to be considered to have identified
23 the disease? These last three items listed under
24 evaluation are key elements of any standalone or
25 clinical testing for a device.

Free State Reporting, Inc.
1378 Cape Saint Claire Road
Annapolis, MD 21409
(410) 974-0947

1 Standalone performance is one of the ways
2 of demonstrating CAD performance. Standalone
3 performance refers to how the device performs in the
4 absence of a reader. Does the CAD mark the correct
5 locations? There are several aspects of this testing
6 that the draft 510(k) guidance addresses. First is
7 the accuracy with which the CAD marks the locations.
8 This is done with a database, the ground truth, and
9 the scoring methods described in the previous slide.
10 A 510(k) submission should also provide definitions,
11 such as true positive, true negative, false positive,
12 false negative, and make sure that these definitions
13 are consistent with the intended use of the device.
14 Similarly, the basis on which results are reported,
15 for example, per patient or per lesion, must be
16 justified.

17 The guidance also recommends that
18 reproducibility testing be reported. For example,
19 for digitized image data, the placement of the film
20 in the scanner or the time when the scanning occurs
21 could produce data differences that may affect how
22 the algorithm performs. And we recommend algorithm
23 stability testing. For example, if the training set
24 is changed, how does the performance change?

25 In addition to standalone testing, a

Free State Reporting, Inc.
1378 Cape Saint Claire Road
Annapolis, MD 21409
(410) 974-0947

1 clinical study will usually be needed to show a new
2 CAD device is as safe and effective as a legally
3 marketed predicate. When we say clinical study here,
4 we are referring to reader studies. The guidance
5 states that because the reader is an integral part of
6 the diagnostic process for CAD devices, you can
7 assume that a clinical assessment will be necessary.
8 This clinical assessment should provide an estimate
9 of the clinical effect of the CAD on clinician
10 performance. You will be asked to discuss the need
11 for clinical data in Question 2(b) today.

12 For clinical assessment, various control
13 arms can be used, for example, a reading aided by a
14 predicate device for a 510(k) or unassisted reading.
15 And you're going to hear more discussion of the
16 selection of control arms today, and you will be
17 asked to discuss control arms in discussion Question
18 2(a). Additional recommendations for clinical
19 testing are provided in the second guidance document.

20 Continuing with the 510(k) guidance
21 document, there are also recommendations that
22 submissions provide summaries of the procedures that
23 will be used to train intended users of the device
24 when the device is marketed. And there are
25 recommendations of what should be included in device

1 labeling.

2 The second guidance document focuses on how
3 to design and conduct clinical performance assessment
4 studies so that they are well-controlled clinical
5 investigations. Again, I will speak to the various
6 sections of this guidance. The guidance discusses
7 some specific aspects of clinical study design.
8 Endpoints should be selected to demonstrate that the
9 CAD device is effective in a significant portion of
10 the target population. Primary and secondary
11 endpoints will depend on the intended use of the
12 device and should be fixed prior to initiating any
13 evaluation. Likely candidates for endpoints include
14 receiver operating characteristic curves, or ROC
15 curves, or some variant thereof, or sensitivity and
16 specificity. You're going to hear more about these
17 possible endpoints in the upcoming presentations.

18 For control arms, the guidance recommends
19 that for PMA submission, the most relevant comparator
20 is generally reading of images without the CAD. For
21 CAD devices intended as second readers, another
22 possible control is double reading by two clinicians.
23 For 510(k) submissions, direct comparison to the
24 predicate CAD device may be useful for establishing
25 substantial equivalence.

Free State Reporting, Inc.
1378 Cape Saint Claire Road
Annapolis, MD 21409
(410) 974-0947

1 Reading scenarios need to be defined. The
2 readers, the cases, and the reading scenarios need to
3 be randomized to reduce bias. If there are multiple
4 reading sessions where some cases are read multiple
5 times, the guidance document recommends that the
6 reading sessions be separated in time by at least
7 four weeks to avoid memory bias.

8 Regarding rating scales, investigators
9 should use conventional medical interpretation and
10 reporting where possible. ROC-based endpoints may
11 support collecting data with a finer rating scale
12 when supporters rate the lesion and/or the disease
13 status in a patient. Readers may need training on
14 using the rating scale.

15 As discussed in the 510(k) guidance, a
16 scoring procedure must be defined for determining
17 when the reader's interpretation matches the ground
18 truth. And submission should also include specific
19 instructions in training provided to study
20 participants on the use of the CAD device and details
21 on how to participate in the clinical study.

22 Patient data may be collected prospectively
23 or retrospectively, based on well-defined inclusion
24 and exclusion criteria. The guidance recommends that
25 submissions provide the protocol used for the case

1 collections. Note that cases collected for clinical
2 trial should be independent of the cases that are
3 used during device development, and they should be
4 new to the readers who are participating in the
5 clinical assessment of the device. In the
6 statistical presentation, Dr. Gwise will discuss the
7 pros and cons of prospective and retrospective
8 studies.

9 The dataset may be enriched with diseased
10 or abnormal cases for an efficient and less
11 burdensome representative case dataset. There may
12 also be what we call a stress test when the
13 population is enriched with patient cases that
14 contain imaging findings or other imaging data that
15 are challenging to clinicians but that still fall
16 within the device's intended use population. You
17 will hear more about enrichment of databases in the
18 talks to come.

19 If a study is based on non-U.S. data, the
20 submission should justify why non-U.S. data reflect
21 what is expected for a U.S. population with respect
22 to disease occurrence, characteristics, the practice
23 of medicine, and clinician competency. There should
24 be statistical and clinical justification of the
25 poolability of data from multiple sites. Submissions

1 should also describe the truthing process, that is,
2 how it is determined whether disease is present, and
3 if so, the extent or location of the disease or other
4 abnormal condition. And the guidance provides some
5 recommendations on how performance results are
6 reported. The guidance also indicates that there's a
7 possibility of a post-approval study as a condition
8 of approval for PMAs. Post-approval studies will be
9 discussed later this morning by Dr. Krulewitch, and
10 your comments on post-market studies will be asked in
11 discussion Question 9 this afternoon.

12 Now, I'm going to turn things over to the
13 other speakers. Dr. Nicholas Petrick will discuss
14 imaging science aspects of CAD devices.
15 Dr. Thomas Gwise will discuss statistical issues.
16 Dr. Robert Smith will present the clinical issues.
17 Dr. Cara Krulewitch will discuss issues regarding
18 post-approval studies.

19 The FDA speakers are going to provide some
20 different viewpoints on the types of evidence needed
21 to support CAD submissions. You're going to hear
22 that there's not full internal agreement on some of
23 these issues, and I presume we'll hear still more
24 viewpoints in the Open Public Hearing after lunch.

25 For your discussion today, we've provided

Free State Reporting, Inc.
1378 Cape Saint Claire Road
Annapolis, MD 21409
(410) 974-0947

1 some specific examples of difficult questions that
2 arise in the review of CAD devices. We look forward
3 to hearing your scientific and clinical perspectives
4 on the assessment of CAD devices. We will use your
5 input to provide pre-market guidance for industry to
6 know how a new CAD device should demonstrate a
7 reasonable assurance of safety and effectiveness, if
8 it's a PMA device, or substantial equivalence to a
9 predicate, if it's a 510(k) device.

10 Thank you for your assistance.

11 DR. PETRICK: Good morning. So I'm going
12 to talk about an imaging science viewpoint on
13 evaluation of CAD devices, and I'll just show you an
14 outline. I'll talk a little bit about controlled
15 reader studies, and in particular study endpoints and
16 control arms for those studies. I'll talk about
17 reuse of data and then talk about assessing algorithm
18 changes.

19 So, again, this is going to be imaging
20 science viewpoint, and I'm going to start off talking
21 about endpoints and controlled reader studies. And
22 there's again been some disagreement on how to power
23 those particular studies. This is related to
24 Question 3(e) the Panel will be evaluating or
25 discussing, and the Agency is seeking feedback on

1 which summary statistic can be used to power
2 controlled reader studies.

3 So the question, again, we're talking about
4 reader studies. In this case we're talking about the
5 reader being a loop. So the question is does CAD-
6 aided reading -- does CAD aid clinician or reader
7 performance? And, in particular, I'm going to talk
8 about a binary task. I'm going to talk about trying
9 to differentiate either normal from abnormal patients
10 or something like disease from non-disease patients.
11 In particular, I'm going to try to talk about making
12 this an objective assessment.

13 The clinical guidance permits controlled
14 reader studies, again, using enrichment of stress
15 populations. So for the vast majority of the CAD
16 submissions, it's likely that there'll at least be
17 enrichment where we'll have additional cases above
18 the prevalence level, and in particular, there'll
19 likely be stress testing with the inclusion of
20 challenging cases.

21 And what I'm just showing here, we can talk
22 about this binary task and make a binary decision,
23 and I just show a two-by-two table, which is going to
24 show the clinical truth versus the reader decision,
25 and very common endpoints of this are sensitivity and

1 specificity, and I just give those definitions over
2 on the left-hand side. This is often thought to be
3 very appropriate because clinicians often think in
4 these go/no go terms.

5 What I've done here is I've just plotted
6 sensitivity versus a false positive fraction or one
7 minus specificity on a curve. And, in particular,
8 when we're talking about binary endpoints, we
9 typically talk about pairs of endpoints. We talk
10 about sensitivity and false positive fraction or
11 sensitivity and specificity, like sometimes we're
12 talking about positive predicative value and negative
13 predicative value. But they typically go together in
14 pairs.

15 I just showed the operating point. We also
16 have to think about these with error bars associated
17 with them. They're not just fixed points in space.
18 And, in particular, there are three main sources of
19 error, and I'll get into the discussion of those
20 error terms later.

21 When we're comparing two devices, these are
22 two devices. In this case we're talking about CAD-
23 aided to CAD-aided reading, two different CAD devices
24 aiding the reader. The question becomes has Device 2
25 improved reader performance? And here I just plotted

1 two different points on the curve. And in this case
2 there's been an increase in sensitivity but also an
3 increase in the number of false positives. We might
4 have that same question if we increase sensitivity
5 and decrease false positives, or we decrease
6 sensitivity and decrease false positives, or in the
7 final category, it looked at decreasing sensitivity
8 and -- but also decreasing the number of false
9 positives.

10 Typically, another approach we might be
11 comparing is unaided reading to aided reading with a
12 single device. And what we've typically seen in CAD
13 submissions and what's seen in the literature is a
14 shift to the upper right, where you're trading an
15 increase -- you have an increase in sensitivity but
16 also an increase in the number of false positives.
17 So the question again is with these two particular
18 endpoints, is this is improved performance?

19 In order to make this a quantitative
20 assessment in a formal manner, we have to talk about
21 utility or risk benefit analysis. And this means
22 that you need to assign a numerical benefit for
23 finding disease, a numerical risk associated with
24 recalling non-disease patients, a numerical benefit
25 value for finding non-disease patients, and a

1 numerical risk associated with missing disease. If
2 you do this, there's a formal equation -- I just give
3 it at the bottom -- that allows you to assess the
4 overall utility for that approach. Now, I just want
5 to have you keep in mind that this formal utility
6 analysis is not a standard approach that's typically
7 used within CDRH, but it is a way to make this type
8 of assessment quantitative.

9 I just want to give you an idea of what
10 utility lines look like. This is using some of the
11 simplifications from the Wagner et al. paper in
12 evaluating these. And what they look like are
13 straight lines in the sensitivity versus false
14 positive fraction space. And then utility typically
15 increases towards the upper left-hand corner. So as
16 we move to the 0.1 point in the space we have
17 increased utility. Also keep in mind that the
18 utility is going to be a linear function of
19 prevalence of the disease and also the individual
20 risks and benefits associated with true positive and
21 false positive, et cetera.

22 So going back to our comparison, if we look
23 at this comparison of two different CADe devices
24 aiding the reader, the question is does this improve
25 performance? And if we can show in a statistical way

1 that the utility of the green Device 2 is higher than
2 Device 1, that would improve performance. So in this
3 case, where we have increased number of -- increased
4 sensitivity and also increased false positives, we
5 still have increased utility. So yes, it does
6 improve performance. Also keep in mind that there
7 are other areas of the space. We have now decreased
8 sensitivity, but decreased false positives. This,
9 again, has increased utility in this particular
10 example. And likewise in this, it's another region
11 of space of where we've increased utility.

12 Now, I've given a formal analysis for the
13 utility. We can also look at regions within the
14 space, and we know something about utility without
15 formally defining the risks and benefits. In
16 particular, I've defined a region with constant false
17 positive fraction and constant sensitivity. And what
18 we know about utility is that if we increase
19 sensitivity and decrease the number of false
20 positives, we'll always increase the utility
21 associated with that original Device 1 point.

22 So this is a region of space where we don't
23 necessarily have to formally define it. There are
24 other ways of defining space. But keep in mind,
25 there are other regions within that space that have

1 increased utility that do not fall within that range.
2 So it's a restrictive range.

3 So what are some of the complications with
4 utility analysis? Well, first of all, utilities need
5 to be considered and defined prior to conducting the
6 study. Likewise, utilities are those particular
7 risk/benefits for true positives and false positives
8 that can be subjective and controversial. Likewise,
9 utilities can change over time. As practice in
10 medicine changes, those utilities may change as well.

11 Another complication is that utility is
12 prevalence dependent, and in particular, the
13 prevalence in the CADE controlled reader study likely
14 won't match that found in clinical practice. So the
15 question is what is a particular prevalence that
16 should be utilized in the study? Likewise,
17 utilities will vary for different imaging tests. So
18 if you're talking about a screening test, the
19 utilities may be very different from a diagnostic
20 type of test.

21 Now, I'll just change back and go talk
22 about sources of variability with individual
23 endpoints, binary endpoints. There's case
24 variability. So based on a number of cases
25 associated with a study, there'll be some variability

1 associated with that. There's also variability
2 coming from reader skill. We're talking about reader
3 studies. As they have different reader skill,
4 there'll be more variability added. There's also
5 variability which I'm calling reader mindset
6 variability. This is the different operating
7 thresholds for different readers, and in particular,
8 reader thresholds change with experience and
9 training, patient risk factors, and evolution in
10 clinical practice. And, in particular, there's
11 always going to be a range of reader thresholds. We
12 see this in a number of different studies. And I'll
13 just give you a classic example. This is a paper by
14 Elmore, from the *New England Journal of Medicine* in
15 the mid-'90s, where they looked at -- investigated
16 inter-reader -- inter-radiologist variability in
17 mammographic interpretation.

18 And what I've just done here is plot on our
19 sensitivity versus false positive fraction curve each
20 of the 10 individual radiologists and their operating
21 points, and this is for the immediate workup rate
22 threshold. And you can see that the mean sensitivity
23 was 87 percent but had the range of 74 to 96 percent.
24 And the mean specificity was 44 percent, with a range
25 from 11 to 65 percent. So what this plot shows is

Free State Reporting, Inc.
1378 Cape Saint Claire Road
Annapolis, MD 21409
(410) 974-0947

1 that there's a large variability among readers, and
2 in particular that each reader had their own
3 operating threshold. And I'll revisit this later to
4 talk about how the reader skill interacts with this
5 plot.

6 I also wanted to talk about the difference
7 between controlled reader studies in clinical
8 practice. We're talking about in CAD evaluation,
9 talking about controlled reader studies typically, as
10 opposed to prospective clinical trials, and in these
11 studies there's typically going to be enrichment or
12 stress testing, so the population might be different.
13 Patient care is typically not going to be impacted in
14 controlled reader studies. Likewise, the clinicians
15 may have little experience with the new device, which
16 may influence how they operate. And, likewise, they
17 may be blind to other patient information.

18 So the conclusion of this is that the
19 sensitivity and false positive fraction or
20 specificity from a controlled reader study is not
21 going to equal the sensitivity and specificity in
22 clinical practice. These numbers will not be the
23 same. And, in particular, we have two different
24 variables that are going to be played against each
25 other.

1 And just to give you an example of this
2 from the literature, there's a study from Gur, et al.
3 that compared radiologists' performance for mammo
4 interpretation in controlled reader studies to
5 clinical practice. And this may be a little bit hard
6 to see, but in your plot you can see that there's
7 different sensitivity and different sensitivity for
8 those same readers in their clinical practice use of
9 those cases as opposed to what they did a laboratory
10 study. And, in particular, reader performance was
11 significantly better in the clinic, in this
12 manuscript, and inter-reader dispersion was
13 significantly lower within the clinic. So the
14 conversion from a controlled reader study to clinical
15 practice had influence both on the variability of the
16 readings as well as the actual operating points.

17 So what are the implications of this in
18 clinical practice? Well, first of all, we're clear
19 that the absolute sensitivity or specificity change
20 is unlikely to be predictive, and also that the
21 sensitivity and specificity may change at different
22 rates. So this is a complicating factor when you're
23 talking about looking at two endpoints. So there may
24 be a performance from -- I'm just showing a toy
25 example where the clinical reader study is on the

1 left-hand side, and this may be the implications in
2 clinical practice.

3 It may also be possible that the
4 sensitivity drops much more quickly than the false
5 positive rate. So those two points come together.
6 This has implications about the utility of that
7 particular operating point. Has that now increased
8 utility over the other point? It may also be
9 possible that we reduce false positive facet and
10 sensitivity decreased, and we see another change in
11 the utility associated with this. This makes the
12 analysis of these particular endpoints especially
13 more complicated.

14 I also wanted to talk about how the
15 clinician may be influenced or, in clinical practice,
16 how a device may change over time. And this is an
17 example from Dean et al., where they're actually
18 looking at prospective comparison of a CAD
19 implemented in clinical practice, and I'm just taking
20 a piece of their data out. I'm not talking about the
21 sensitivity part, but just showing you how the
22 implementation of a CAD may change over time.

23 They looked at the recall rate for
24 screening mammographic patients in their practice,
25 and in particular, six months before the CAD

1 installation, it was 6.2 percent. In the initial two
2 months of utilization, it jumped to 13.4 percent.
3 From months 3 to 24, it reduced down to 7.8. And
4 then that four-month follow-up period again reduced
5 down further to 6.75. So looking at this study, it's
6 clear that the recall rate with the CAD changes over
7 time, and in particular, that in the initial studies,
8 the device did not define the operating point. But
9 in fact, as the clinicians learn with the device,
10 their operating points will change over time.

11 So just so I summarize what I was talking
12 about in these binary decisions for sensitivity and
13 specificity endpoint, which is making this sort of
14 go/no go decision, it matches much of clinical
15 training and practice. That's a big advantage of
16 using that endpoint.

17 For objective comparison, it requires
18 considering utilities, and in particular, the
19 utilities of the particular risk/benefits can be
20 subjective and controversial. Utilities will be
21 specific to the testing dataset. Likewise, utilities
22 can change over time, and the utility analysis will
23 be prevalence dependent.

24 In addition, this type binary decision has
25 three sources of variability: cases, reader

1 threshold, and reader skill. So, again, more data is
2 typically going to be necessary to overcome these
3 additional sources of variability.

4 Now, I'm going to shift focus and talk to
5 you about a different way of evaluating the same
6 data, talking about multilevel decision, and this is
7 associated with what's called receiver operating
8 characteristics, or ROC analysis, where the goal is
9 to evaluate a range of tradeoffs in sensitivity for a
10 given task. And compared to sensitivity and
11 specificity, ROC provides more bang for the buck. In
12 particular, it provides more information, it can
13 clear ambiguous comparisons without resorting to
14 utility analysis, and likewise, it provides more
15 statistical power through averaging.

16 Some of the criticisms of ROC has been
17 clinicians don't rate patients, they really just
18 decide and act, and that all operating points are not
19 necessarily relevant. And I'll get back to those
20 criticisms in a little bit.

21 I also wanted to sort of generalize a
22 little bit more. I've talked about, particularly,
23 sensitivity and specificity endpoints. That's really
24 a patient-based type of analysis. And I'm also going
25 to sort of concentrate on ROC, which again is a

1 patient-based analysis. But we can also talk about
2 location-based analysis, things like LROC, location-
3 specific ROC, or free-response ROC. There are other
4 processes associated with region-based ROC to try to
5 incorporate location in the analysis. So for the
6 following discussion, I'm going to concentrate on
7 ROC, but keep in mind that this generalizes to
8 location-based analysis as well. And the specific
9 analysis would really depend on the CAD task at hand
10 and the particular selection.

11 So I'm going to go through, and I'm sure
12 most of you already know this, but I'm going to go
13 through and introduce ROC analysis, again, with the
14 goal to evaluate a range of tradeoffs in sensitivity
15 and specificity for a task. What I'm going to do is
16 try to show this in a way that keeps in mind that
17 binary decision, that go/no go decision by the
18 clinician.

19 So I just show a table of decision space on
20 the lower left, and we can go back to the original
21 decision, that original threshold from the clinician,
22 and ask them to make a binary decision of whether the
23 patient should be worked up or not. That puts a
24 point into ROC space. We can then go back and
25 revisit.

1 Let's take those cases that were called
2 negative by the clinician and say, look at these
3 cases and which are the cases you would recall first
4 or work up first relative to this subgroup? So they
5 reanalyze those cases. Once they do that, that will
6 provide a new threshold. They'll have actually
7 recalled more cases than the original. They'll then
8 be acting more aggressively. That puts another point
9 into the space.

10 We can do that same thing, look at the
11 subgroup of cases that they called positive and say
12 which one of these is the most suspicious that you
13 would actually work up first as opposed to others?
14 This would cut off some of the cases. This would be
15 considered a lower aggressiveness threshold, and that
16 puts another point in space. You can continue this
17 process of reevaluating the cases or splitting the
18 levels, with a goal to order the cases from least to
19 most suspicious. And if you do this, you can fill
20 out the curve and the number of points.

21 But this can be achieved in a number of
22 different ways. First, I just showed you a case
23 where you revisit the cases and do this type of
24 assessment. You could also score individual cases.
25 This has been the typical approach in CAD reader

1 studies in the literature. You can also do pairwise
2 comparisons where you look at pairs of cases over
3 time, or some sort of hybrid approach. It's
4 important to realize that when you get the continuum
5 of thresholds, this plots out what we call the ROC
6 curve and we get a whole curve. And I just wanted to
7 point out that ROC is really the recommended approach
8 from the image science community. In particular,
9 there's an International Commission on Radiation
10 Units and Measurements, ICRU Report 79, which deals
11 with the ROC analysis in medical imaging and the
12 potential advantages of this approach.

13 What I showed you before is the actual ROC
14 curve. We can talk about summary statistics. One of
15 the most common is area under the ROC curve, and I
16 just show that plot on the right-hand side. ROC
17 really can be defined -- can be looked at in a number
18 of different ways. It is the probability that a
19 disease ranking is greater than a non-disease
20 ranking. It's also the sensitivity average over all
21 specificities. And, in particular, it's measuring a
22 reader's ability to separate populations.

23 Again, I'll concentrate on this sort of
24 total area in the rest of my analysis, but we can
25 also be talking about partial areas or areas within a

1 region ROC space, and that may be a very valuable
2 type of analysis as well. And the particular
3 endpoint or summary statistic would depend on the
4 particular application.

5 So back to this issue of measuring a
6 reader's ability to separate disease from non-disease
7 populations. Here I just show two distributions of
8 cases. The goal is to try to differentiate those.
9 And for this particular overlap of the cases, you see
10 a particular ROC curve associated with that, and AUC
11 is .85. If this CAD device comes along and improves
12 a reader's ability to separate those distributions,
13 the disease from the non-disease, that will lead to
14 an increased ROC curve and increased area under the
15 ROC curve.

16 If, on the other hand, this device comes
17 along and for whatever reason makes it harder for the
18 reader to differentiate those distributions, those
19 distributions will come together, we'll see a lower
20 ROC curve, and we'll see a lower area under the ROC
21 curve. So, again, the ability to separate those
22 distributions, or as associated with increasing the
23 reader skill, is associated with increasing ROC
24 curves and increasing areas under the ROC curves.

25 Again, going back to this question of

Free State Reporting, Inc.
1378 Cape Saint Claire Road
Annapolis, MD 21409
(410) 974-0947

1 evaluating a CAD device, in this case we're looking
2 at evaluating a single device between unaided to
3 aided reading. Comparing unaided to aided reading,
4 we would look at comparing the ROC curves, and then
5 when the ROC curves are non-overlapping, to compare
6 the area under those ROC curves as a summary
7 statistic of the differentiability. And in this
8 particular example, we had two points which were
9 difficult to evaluate by themselves, two different
10 operating points. If we look at the ROC curves,
11 again, the comparison can often be clearer. There is
12 no need to assign the utilities to these, and it
13 accommodates the various reader thresholds, and this
14 is a very important advantage of ROC analysis.

15 ROC analysis has, certainly, sources of
16 variability as well. In particular, it has case
17 variability, just as the individual endpoints. It
18 also has reader skill variability. Again, with
19 different reader skill, there'll be more variability
20 associated with it. However, ROC curves include the
21 operating -- include all operating thresholds and
22 include differences in reading mindset. This means
23 that the reader threshold variability is absorbed
24 within the ROC curve, and so that component does not
25 come through in the process.

1 So, again, ROC mitigates significant
2 sources of variability. In particular, the reader
3 mindset variability is mitigated. And this by
4 design. ROC eliminates operating threshold
5 variability. Likewise, case and reader variability
6 is reduced through averaging. Area under the curve
7 is an average over an area. So that's averaging over
8 a range instead of a point estimate to reduce these
9 sources of variability as well. So there are a
10 number of statistical advantages to using ROC.

11 So let's just go back and revisit the
12 Elmore data. And one of the questions may be can the
13 data be fit by an ROC curve? And the answer to that
14 is, in this particular case, fit an ROC curve to the
15 immediate workup data. And what we see is that ROC
16 suggests that actually the readers have fairly
17 similar skill. There are a few that might be
18 slightly better and a few may be slightly worse. But
19 really what we're seeing is differences in threshold
20 and operating points, and this will be always
21 consistent in imaging tests with a reader
22 interpreting the data.

23 And I don't want to claim any credit for
24 looking at this originally. D'Orsi and Swets looked
25 at this in the mid-'90s. They plotted an ROC curve

1 for a different biopsy/no biopsy threshold. But,
2 again, these would be -- these data would fall on the
3 same curve. That workup rate threshold is really
4 just a different operating point along the same ROC
5 curve.

6 We can go back to that question of
7 controlled reader study versus clinical practice, all
8 those same difficulties comparing controlled reader
9 studies to clinical practice, and at least as many,
10 and there may be other factors as well for both the
11 study controls. So the implication of this is the
12 area under the ROC curve or the ROC curve itself is
13 not going to -- in a controlled reader study, is not
14 going to be the same AUC or ROC curve as found in
15 clinical practice.

16 What are the implications of this? Again,
17 the absolute AUC may not match. However, now we've
18 gone to one endpoint. Instead of thinking of two
19 endpoints trading off against each other, we have one
20 typical endpoint, and that means that the direction
21 of the change should be predictive. So we're not
22 sure exactly the magnitude of the change, but the
23 direction of that change should be predictive.

24 So just to summarize this multilevel
25 decision, I talked about ROC curves providing added

1 information over single endpoints, and in particular
2 AUC measures the separation between disease in
3 populations for the reader performing that particular
4 task. The ROC is efficient. It reduces variability
5 by design. It is limited to -- it has two main
6 sources of variability. That comes from cases and
7 readers. And these two sources are actually reduced
8 through averaging. And, again, it eliminates that
9 threshold variability component that are seen in
10 binary endpoints.

11 Back to the critiques of ROC. While
12 clinicians don't rate patients, they make go/no go
13 decisions. ROC really is not meant to exactly match
14 clinical practice. And, in fact, all controlled
15 reader studies may compromise this, and this may be,
16 I think, a legitimate compromise to make. ROC is
17 meant to efficiently evaluate a technology across
18 many different operating thresholds. This is the
19 task at hand in evaluating medical images.

20 The criticism also that not all operating
21 points are equally relevant. Again, readers operate
22 over a range. Multiple readers operate over a range
23 of endpoints. Individual readers operate over a
24 range. So area is a relevant measure. The question
25 is how to select that particular range, and using

1 something like the partial area may be an appropriate
2 tool for that.

3 So let's get back to the questions that the
4 Panel will be discussing. Which summary statistics
5 should be used to power a study? The viewpoint that
6 I'm advocating is that sponsors can choose to power
7 the study for sensitivity and specificity. That may
8 be the way they need to approach that for their
9 particular CAD device, or use some other type of
10 binary statistic. However, utility should be
11 considered, and that may be a region within the
12 utility space or it may be directly assigning
13 utilities. However, the sponsor should keep in mind,
14 this suffers from threshold variability, so likely
15 more data will be required.

16 The other option I would advocate is that
17 the sponsor should have the option to power the study
18 to show a statistically superior AUC or partial area
19 or whatever might be the relevant location measure.
20 This would still be required to collect or be
21 recommended to collect and report clinical
22 sensitivity and specificity output data in the
23 labeling. The advantage of this is that it increases
24 statistical power, and it really increases the
25 information on the technology that we would receive.

1 All right. So that's a long section on
2 endpoints. I'm going to go now talking about control
3 arms in controlled reader studies. Again, this is an
4 area of disagreement within the Agency. The guidance
5 suggests a couple of different or a number of
6 different potential study control arms. The two most
7 likely, I think, are going to be unaided reading and
8 the direct comparison with a predicate device. So
9 I'm going to lay a couple of scenarios for using
10 these two different approaches. Again, this is
11 related to questions. The Agency is seeking feedback
12 on which control arms are viable within each pathway,
13 510(k) and PMA, or Class II or Class III CAD device
14 submissions.

15 And so in this particular arm, I'm going to
16 lay out a pathway where you use a predicate control,
17 and in particular they use the same cases and
18 readers. So the goal is to compare the aid of the
19 CAD-2 with the aid provided to the reader of CAD-1.
20 In particular, there's a single test dataset and
21 there's a single set of readers that evaluate both
22 devices. So the goal would be to look at the aid
23 provided by CAD-1, the aid provided by CAD-2, and
24 statistically compare those together.

25 So the CAD-2 aided reading should be

Free State Reporting, Inc.
1378 Cape Saint Claire Road
Annapolis, MD 21409
(410) 974-0947

1 substantially equivalent to the CAD-1 aided reading
2 if it's a 510(k) submission, or if it was a Class III
3 submission, to show safety and effectiveness for PMA
4 supplements. This would establish the benefit of the
5 CAD-aided compared with the predicate because
6 especially when the same cases and readers are used,
7 this allows for a direct comparison. However, it may
8 not lead to showing superiority for the CAD aid
9 over -- so the CAD device-aided reading over unaided
10 reading. And it depends on how the study is powered
11 and what particular endpoints are used in the study.
12 So it may show that they're equivalent, but it may
13 not be powered enough to show superiority over
14 unaided reading.

15 Likewise, it requires access to both the
16 sponsor's CAD device as well as the predicate CAD
17 device, and this can be a complication because
18 previous CAD algorithms may not be available. If
19 they're coming from a different company, that could
20 be complicated. And, likewise, original devices that
21 may have been approved a number of years ago may not
22 be currently on the market anymore, so again, they
23 would not be available for a comparison. However,
24 this may be very valuable for within a single company
25 comparing an updated algorithm to an older algorithm.

1 I'll talk about a second pathway using
2 unaided reader control, but with different cases and
3 readers. So, again, the goal is to compare Device 2
4 with Device 1 in aided reading. However, the study
5 would be having two different sets of readers and
6 datasets. So the device would be compared
7 individually between unaided reading and aided
8 reading, and Device 2 would then do the same thing,
9 the different dataset, different readers, compare
10 unaided reading with aided reading. Both devices
11 would need to establish that CAD-aided reading is
12 statistically superior to unaided reading. Again,
13 the comparison is that both CADEs improve reader
14 performance and each device would stand on its own
15 with respect to improving unaided reading.

16 This is the paradigm used for Class III
17 submissions without a predicate. It only requires
18 access to the sponsor's CAD algorithm and test data.
19 And, in fact, we can use multi-reader and multi-case
20 analysis to actually allow for a direct comparison of
21 the performance when the cases and the readers are
22 sampled from the same populations.

23 Back to the question. Again, we're going
24 to talk about what are valid control arms, and we'll
25 talk about that in terms of Class II and Class III

1 devices, or PMA and 510(k) submissions. The
2 viewpoint, again, that I'm advocating is that for
3 original PMAs, to allow unaided reading control. So
4 that's a single CAD device showing superiority of
5 that CAD-aided reading over unaided reading and as
6 the current recommendation for Class III CADe devices
7 coming into the Agency.

8 For 510(k) and PMA supplements, the options
9 would be that I'm advocating is to allow for the
10 predicate device control. This allows for a
11 comparison of the aided readings between the
12 different devices, and in particular, if the same
13 cases are used and the same readers, you could do a
14 direct comparison. If they're different cases and
15 readers, as long as they're sampled from the same
16 population, that would still offer a comparison. And
17 this would typically be a non-inferiority study,
18 especially for 510(k)'s.

19 I would also advocate allowing unaided
20 reading as a control. So the comparison is that both
21 devices improve unaided reading as compared -- that
22 both devices improve unaided reading without
23 necessarily determining which device is better. If
24 different cases and different readers are used, as
25 long as they're sampled from the same population, in

1 fact, we could actually be able to compare those
2 devices as well. Superiority of the CAD-aided
3 reading over unaided reading for both devices would
4 be necessary. And, again, this is the approach for
5 Class III PMA submissions currently.

6 I'm going to shift focus again, only a
7 couple more times. I'm going to talk about readers
8 of data. At the last Panel meeting, the Panel -- I'm
9 just going to give a quote out of that. "The Panel
10 had severe concerns about the reuse of test data, and
11 that optimally a new test set should be obtained.
12 However, we," the Panel, "realize" -- or that's you
13 -- "that there will be certain circumstances where
14 that will either be unnecessary or so burdensome that
15 a lesser solution would be acceptable." So starting
16 from that perspective about reuse of data, the Agency
17 again is seeking additional feedback on how to
18 approach data reuse.

19 And I'll just lay out the general idea of
20 how CADs are developed. Typically, there's a
21 training dataset that's used for algorithm
22 development. Typically, there's some sort of
23 internal validation dataset that says the algorithm
24 is good enough that we are willing to look at a
25 regulatory dataset to evaluate it. And when it

1 passes that, then there's this independent regulatory
2 dataset that's used for algorithm testing. If that's
3 successful, then that device would come in to the
4 Agency for approval.

5 What I'm going to concentrate on is not
6 those first two blocks about validation or training
7 but talk specifically about that regulatory or what
8 I'm going to call the test database, the test
9 dataset, and how that might able to be reused.

10 So there are a couple of different
11 approaches the FDA could take. First of all, the FDA
12 could allow reuse without replacement. This
13 particular approach, however, would mean that
14 learning will occur, and even if it's inadvertently,
15 there's information in that data when you get results
16 back from the test dataset, and over time you could
17 take that -- use that to advantage in how your device
18 is developed.

19 We could also talk about complete
20 replacement. This would certainly be a conservative
21 approach. The downside of that is datasets would be
22 unlikely to grow in size over time, and increasing
23 the size of a database is very useful for
24 understanding how well that device performs. And, of
25 course, the other approach is partial replacement

1 with each reuse. The question would then be how many
2 cases need to be replaced on each reuse.

3 I just wanted to point the Panel to what
4 we're talking about when we're talking about error.
5 In this case we're talking about mean squared error.
6 There's actually two components. There's a component
7 associated with the bias and there's a component
8 associated with the variance associated with error.
9 The bias would be the sum of all the study biases,
10 and if reuse was allowed, the reuse bias would be one
11 of those components. The variance is actually a
12 function of the sample size, and if we're talking
13 about data reuse in a controlled reader study, it
14 actually comes from the readers as well. So there
15 are two components, potentially, to the variance.

16 So I'm going to look at a particular
17 scenario where there's limited feedback provided to
18 the developers, and in particular, the developers
19 only see the summary performance. They don't see
20 subgroup, or they don't have the results reported
21 back on subgroups or individual case performance. So
22 it's just the total performance for that particular
23 algorithm.

24 In this scenario, if the test samples are
25 used repeatedly, the developer could optimize the

1 algorithm to improve performance, and in particular
2 you could look at a scenario where you could increase
3 the bias associated with reuse, the reuse bias based
4 on each reuse of the dataset.

5 What are some ways to potentially mitigate
6 this reuse bias? Well, here I just look at a dataset
7 that has N observations or N cases, and I've just
8 broken it up into a number of different subgroups of
9 size M . So on the first use of that data, the
10 developers haven't seen any of the results coming
11 back from that yet, so the whole dataset's available
12 to be used. So the first dataset, you use all N
13 observations.

14 On the second use, one potential way of
15 looking at that is to delete M of the cases and then
16 add some J new cases. In this case we're thinking
17 the J is bigger than M . And then you could randomly
18 sample from that larger population for N observations
19 and test again. You could repeat that same process
20 on a third reuse. In this case we just assumed
21 there's some different number of cases added, again,
22 re-sampling down to get N observations. This is a
23 way to control the bias of reuse.

24 It would require strict management of the
25 test dataset. What information is fed back to the

1 developers is obviously very important. By
2 supplementing data on each reuse, that would control
3 for reuse bias. By random sampling of the data on
4 each use, that would be an additional control on
5 bias. And potentially incrementally increase data on
6 each use, that actually would control for the
7 variance. So we can control for both the bias and
8 variance through this approach.

9 Again, back to the Question 3 for the
10 Panel. It's to look at the question of reuse. And
11 the approach that I'm advocating is to allow some
12 reuse of the test data, again, with tight controls,
13 accounting and management of the regulatory dataset,
14 to recommend incremental increase in the dataset on
15 each reuse, and then random sampling from that
16 dataset on each reuse. This would allow the dataset
17 to grow over time, which would control both the bias
18 and variance. It also provides a sponsor with some
19 flexibility to use data efficiently, and it also
20 manages the risk of training to the test dataset.

21 All right, the final area I'm going to talk
22 about, which is algorithm changes where clinical
23 assessment may not be recommended. And there are
24 number of questions that deal with this, and really,
25 the question is under what conditions would the

1 Agency potentially accept a surrogate endpoint, such
2 as standalone performance, in lieu of clinical
3 performance assessment? And this is certainly an
4 important question for CAD algorithm and CAD
5 companies with changes coming into the Agency.

6 So I just tried to identify three different
7 types of -- basic types of changes. There may be
8 changes to the algorithm, but no changes to the data
9 source. So this would be something like an algorithm
10 update, but the data is coming from the same basic
11 imaging hardware. There could be no changes to the
12 algorithm, but now the application of that CAD is to
13 a new piece of hardware. So that could be a CAD
14 applied to a data coming from a new CT platform. Or
15 there could be changes to both the algorithm and the
16 data source, and this may happen if the algorithm is
17 adjusted for application, say, to a new piece of
18 hardware.

19 I just wanted to mention that most
20 algorithm changes are likely to be incremental, at
21 least after the initial approval of the device. Not
22 all changes will be incremental, and multiple
23 incremental changes would maybe lead to a major
24 change. But, in general, these are going to be
25 incremental changes.

1 So one approach the Agency could take is a
2 conservative approach, which would be to have the
3 sponsor study the impact of every change on reader
4 performance. So that would be a new controlled
5 reader study or clinical study for any change in the
6 hardware, change in the acquisition protocol, or
7 change in the algorithm.

8 I just wanted to try to highlight some of
9 the information that's available to make a decision
10 on whether that was necessary or not. In particular,
11 we would have two pieces of information. Assuming
12 this was an initial submission, we would have the
13 standalone performance data as well as the
14 performance in the controlled reader study.

15 Keep in mind that standalone studies will
16 likely have less variability than reader studies.
17 This would come from two sources. One is that
18 there's not going to be reader variability in the
19 standalone data, and typically there's more
20 standalone data available for assessment as well. So
21 the standalone data will actually be more sensitive
22 to measuring change.

23 One type of improvement may be that the CAD
24 improves, and in this case I'm going to talk about
25 improvement in the sense that we're talking about

1 increasing sensitivity and potentially decreasing the
2 number of false positives, that type of change, which
3 would be typically what we see of devices coming in,
4 but not necessarily the only type of changes. In
5 this case, which is maybe a little hard to see on the
6 slides, the CAD update falls within the confidence
7 bounds of the actual original submission and
8 standalone performance. So the question for the
9 Panel is does CAD improvement within those original
10 confidence intervals require a new reader study?

11 We can have a second type of change, which
12 is that the device actually has improvement that
13 falls outside of those confidence intervals. And,
14 again, what about -- the Panel, I think, would need
15 to discuss what about improvements that fall outside
16 those confidence bounds? They're improvements in the
17 sense that they're improvements in standalone
18 performance. Why does that have meaning as far as
19 the clinical performance?

20 I just also wanted to give a toy example of
21 what the clinician may be seeing from a CAD device,
22 so I picked sort of a potential CAD sensitivity of 80
23 percent at two false positives per image. And what I
24 wanted to highlight is the number of false marks the
25 clinician would see relative to a true mark when

1 they're working with this device in clinical
2 practice.

3 At that particular rate, if we're talking
4 about sort of a screening prevalence, which I just
5 assumed was .005, they would see roughly 500 false
6 marks for every true mark they saw in clinical
7 practice. If the prevalence was something more like
8 a diagnostic prevalence of .1, they'd see roughly 25
9 false marks for every true positive mark. And now
10 I'm just assuming for this example, for that
11 particular initial submission, that there was some
12 clinical study done to establish that it had clinical
13 impact.

14 If there was a modification that came in
15 to, say, improve sensitivity to 85 percent but reduce
16 false -- and reduce false positives to, say, 1.75,
17 the clinician would still see roughly 400 false marks
18 for every true mark for a screening population, and
19 roughly about 21 false marks for every true mark in a
20 diagnostic type of situation. If that change was,
21 say, 85 percent, down to 1.5 false positives, the
22 clinician would still see roughly 350 false positives
23 for every true, and roughly 18 false positives for
24 every true in a diagnostic type of situation.

25 So what this means is that there's likely

Free State Reporting, Inc.
1378 Cape Saint Claire Road
Annapolis, MD 21409
(410) 974-0947

1 to still be a large ratio of false positive to true
2 marks, and the question, I think, for the Panel to
3 potentially consider is will the reader notice the
4 impact of this change, and in particular, what is the
5 clinical impact of that type of change?

6 I also wanted to give a historical
7 consideration for what the FDA has done with hardware
8 changes. So these are CAD algorithms supplied to
9 different hardware. For a CAD applied to CT or
10 radiographic data, the CAD has been limited to a
11 specific range of acquisition parameters. However,
12 there's been no restriction on including new CT
13 hardware.

14 For a CAD applied to screen-film
15 mammography, there's been no restriction on any new
16 hardware or particular acquisition parameters.
17 However, it has required standalone data to support
18 each new film digitizer that has come along with that
19 device.

20 For CAD applied to digital mammography, the
21 Agency has asked for standalone data to support each
22 new system, and with the new guidance, we expect to
23 recommend new controlled reader studies for each new
24 FFDM hardware.

25 There is also other information the Agency

Free State Reporting, Inc.
1378 Cape Saint Claire Road
Annapolis, MD 21409
(410) 974-0947

1 will potentially have about changes, certainly
2 information about the device, information on the
3 details of the change, and information on the
4 motivation for the change. The Agency will also have
5 hardware -- imaging hardware and protocol
6 information, imaging physics hardware as well as
7 non-clinical testing from that device, and details of
8 imaging acquisition protocols.

9 The Agency may also have information on
10 reproducibility testing about the algorithm, which
11 would be the variability of that algorithm with
12 respect to imaging hardware and protocol changes.
13 The Agency may also have stability information on the
14 algorithm. So that would be variability information
15 of the algorithm with respect to algorithmic
16 variations. And the Agency will also have standalone
17 testing data, which will be a direct assessment of
18 the change on clinical cases, but without the reader.
19 This gives the Agency at least some ability to assess
20 the impact of changes on subgroups.

21 Now, I gave you that toy example, and I
22 just said sensitivity increased 85 percent. It's not
23 clear if that was across the board for every subgroup
24 had an increase in sensitivity, or maybe one subgroup
25 had a large increase and another subgroup had a

1 decrease. By looking at standalone data, it at least
2 gives the Agency some ability to look at the impact
3 of that change on subgroups.

4 There are a number of different questions
5 dealing with this issue of change. Question 2(b)
6 deals with scenarios related to CAD algorithm
7 changes, in particular performance changes and prompt
8 and output changes. Question 4 deals with
9 application to new imaging hardware. Question 5
10 deals with the more general question of can the Panel
11 help identify what might be considered minor
12 modifications that would not need clinical
13 performance assessment?

14 And so I will now try to summarize what I
15 tried to say. I talked about endpoints, in
16 particular sensitivity and specificity operating
17 points. I talked about the additional variability
18 from different reader thresholds for these particular
19 selections of endpoints. And I also mentioned a
20 consideration that utility is required for
21 quantitative comparisons.

22 I also talked about ROC, which measures a
23 clinician's ability to separate disease from
24 non-disease population, and it's designed to account
25 for a range in thresholds. The question for the

1 Panel would be should ROC be an option to power a
2 controlled reader study, in particular in the
3 situation where we have enrichment and stress
4 testing?

5 I talked about control arms and that
6 unaided reading is a current control in Class III PMA
7 submissions, where the device stands on its own. I
8 also talked about a direct comparison with predicate
9 requires access to that predicate device. And the
10 question for the Panel is can unaided reading be a
11 control when comparing to a predicate device?

12 I talked about the reuse of data and the
13 fact that it adds bias. But by supplementing and
14 incrementing the data, it can control both the bias
15 and variance. And, again, the question for the Panel
16 is under what conditions can data be reused?

17 And, finally, I talked about algorithmic
18 changes, and I talked about different types of
19 changes, changes to the algorithm but no changes to
20 the data source, changes to the hardware or
21 acquisition parameters, and changes to both of those.
22 I also mentioned that most algorithms are likely to
23 be incremental, at least after initial submission.
24 And, again, the question for the Panel is when is a
25 new controlled reader study recommended for algorithm

1 changes or changes in imaging hardware?

2 Thank you.

3 DR. D'ORSI: Dr. Smith is going to present
4 a clinical view.

5 UNIDENTIFIED SPEAKER: No, Dr. Gwise.

6 DR. D'ORSI: Oh, excuse me, I'm told it's
7 Dr. Gwise who will present a statistical view.

8 DR. GWISE: Good morning. My name is
9 Thomas Gwise. I'm a mathematical statistician and
10 Acting Team Leader with the Division of Biostatistics
11 in the Office of Surveillance and Biometrics. And I
12 hope we can all stay awake long enough to get through
13 another statistical talk.

14 Here's a brief outline of what I'll talk
15 about. I'll cover some basic statistical concepts,
16 talk about study designs, prospective and
17 retrospective study designs, and I'll try to focus on
18 areas of interest with respect to the questions
19 you'll be asked, and these involve choice of
20 endpoints, controls, and data use issues.

21 First of all, two dimensions are always
22 considered when we're evaluating diagnostic test
23 performance. How well can the test detect disease
24 cases, and how well can the test correctly identify
25 the non-disease cases? These are sensitivity and

1 specificity, respectively. ROC curves are plots of
2 sensitivity and specificity considering all possible
3 cutoffs, as Dr. Petrick just described.

4 So one important thing we have to take into
5 consideration when we're evaluating a diagnostic test
6 is does the test add any value to the clinical
7 process? That is, for example, is a diagnostic test
8 for bone mineral density better than just using a
9 person's age in diagnosing osteoporosis? Another
10 example would be does the use of CAdE device improve
11 the diagnostic performance of readers? And this
12 could be an increase in sensitivity and specificity,
13 or perhaps an increase in the area under the ROC
14 curve, or maybe improved reading time for the same
15 performance in one of those metrics.

16 Devices are studied under the intended use.
17 The vast majority of submissions for CAdE devices to
18 date have been for those labeled as second readers,
19 aides to physicians, where the user is directed to
20 completely evaluate images, as practice dictates,
21 before initiating the CAdE. As such, it's expected
22 that using the device in accordance with the label
23 will improve performance of the physician.

24 Because the study conduct matches the
25 intended use, it's generally believed that a good way

1 to test for a change in performance is to do a multi-
2 center, prospective, randomized, controlled clinical
3 trial, where we would randomize patients to the
4 respective experimental conditions, unassisted image
5 reading and reading assisted with CAD, manage the
6 patients according to the evaluations as in routine
7 clinical practice, then follow up to determine their
8 disease state and analyze the data.

9 The benefits of doing this is we study the
10 device under its intended use, that is, routine
11 clinical practice, where reader decisions affect
12 patient management. Doing this will get good
13 estimates of performance under the intended use.

14 Some drawbacks to a randomized, controlled
15 clinical trial is that in populations where
16 prevalence is low, a prospective study would take a
17 long time to accrue enough patients. And there's
18 also a risk to participants if we use the device
19 under study to manage patients. And this would
20 require an investigational device exemption.

21 So we have a few popular surrogates or
22 proxies for determining diagnostic performance in the
23 population. These are retrospective reader studies
24 and standalone studies, which are studies of the
25 device performance without any reader involvement.

1 So retrospective reader studies. I'm
2 calling retrospective reader studies reader
3 evaluations that are made off line on a retrospective
4 dataset of images on which disease status of patients
5 has been established according to some ground-truth
6 rules. And usually we'll use a multi-reader, multi-
7 case design where multiple readers read some or all
8 of the images and the sample is enriched with
9 disease-positive cases.

10 Retrospective reader studies don't pose a
11 significant risk because readers -- a significant
12 risk to patients, that is, because the reader results
13 are not used to manage the patients. So, in
14 addition, an IDE would not be required. And they're
15 very efficient. Their relatively small sample size
16 can result in precise estimates of our performance
17 metrics.

18 Some drawbacks to these retrospective
19 reader studies is reader behavior may not be the same
20 as in routine clinical practice because the readers
21 know their readings do not matter to the patients, so
22 they may act differently in the laboratory
23 environment. Readers may detect enrichment, which
24 could affect their reading behavior. And enrichment
25 causes spectrum bias. And more about this in a

1 moment. For example, enriching with challenging
2 cases could cause a downward bias in reader
3 performance and upward bias in CAdE effect on the
4 reader.

5 So I'll talk in a little bit more detail
6 about some of these complications, the reader
7 variability issues, enrichment-related biases, choice
8 of controls for these studies, and some assumptions
9 that we make.

10 So I'll show you some data on reader
11 variability. This is from the study by Beam, et al.,
12 where 108 U.S. mammographers were asked to read a
13 common set of 79 mammograms and provide a rating
14 according to the BI-RADS scale, from 1 to 5, where 5
15 is the highest level of suspicion of cancer.

16 And looking at this image, we can see that
17 sensitivity ranges from about 45 percent to about 95
18 percent, and specificity ranges from about 40 percent
19 to again about 90 percent. So there's a wide
20 variability in reader performance.

21 So companies have submitted studies with
22 from 5 to 20 readers. We would want the reader
23 sample to be representative of the intended use
24 population. A small number of readers involved in
25 these studies may not be generalizable to the

1 population.

2 I'll talk a little bit about enrichment,
3 which is the process of supplementing the image
4 sample with disease-positive images. Performance
5 estimates obtained with enriched study samples would
6 likely be different than performance in the intended
7 use population. We infer that the differences in
8 performance between modalities may be qualitatively
9 applicable to the intended use population if the
10 spectrum of disease is properly represented.

11 Different case mixes of lesion types will
12 likely result in different performance estimates.
13 This is spectrum effect. For example, in
14 mammography, a CADE may have more difficulty
15 detecting some masses than microcalcifications. A
16 sample in which the proportion of microcalcifications
17 to masses is large will give a higher performance
18 estimate compared to a sample in which that
19 proportion is smaller.

20 I'm going to show you an example that
21 illustrates the spectrum effect. The box on the left
22 there represents reader scores on a scale from 0 to
23 100, and the disease-negative patients' images are
24 over here on the left and the positive images are on
25 the right and we see, from the ROC curve, that the

1 test does a good job of separating these two
2 distributions. And I simulated this to represent
3 cases that are easy to diagnosis.

4 Now, I have the same number of cases here
5 in this set of data. This is simulated data also.
6 And, in fact, the disease-negative cases are the same
7 exact ratings, the difference being the disease-
8 positive cases, I'm simulating them to be more
9 difficult to detect, and that's reflected in the
10 lower -- some of the lower scores. And we see that
11 this affects the ROC curve and the area under the
12 curve. Now, the thing to keep in mind is the ratio
13 of disease positive to disease negative has not
14 changed, only the case mix.

15 So we consider a sample of images enriched
16 with a large proportion of disease-positive cases
17 easily detected by readers and CADes. It may be
18 difficult to see a difference between the two
19 modalities. And that's depicted here.

20 A stress test is a study in which a sample
21 of images is enriched with a large proportion of
22 positive cases considered to be difficult to detect
23 by readers and CADs, the goal being to show that the
24 device can add value in cases that are difficult for
25 readers. And this would be -- the goal is to see a

1 difference between the two modalities.

2 Context bias. Readers in a study
3 environment will become aware of the environment and
4 could change their reading behavior in response. And
5 this is defined in a paper by Egglin et al. And
6 investigators attempt to mitigate this by estimating
7 the relative performance between the two devices.

8 Now, I'll talk a little bit -- give a
9 little bit of background for questions on endpoints.
10 This is a little more specific to ROC and sensitivity
11 and specificity. ROC curves show how well a test can
12 separate disease test scores from non-disease test
13 scores. Now, we assume that a decision variable can
14 model a reader's decision process, for example, the
15 probability of malignancy score, where readers are
16 instructed to rate an image with respect to the
17 probability that it represents disease. The ratings
18 I've simulated here are for 25 healthy and 25
19 disease-representing images. And this is a similar
20 setup to what we were looking at before the data on
21 the left and the ROC curve.

22 Now, the data I've simulated here, both of
23 these distributions are Gaussian. ROC curves are
24 invariant to monotone transformations, that is, the
25 relative ranking is the key to the ROC curve. So if

1 we take another look at our ROC curve with the
2 Gaussian data, we can see if we were to merely look
3 at the ratings, we would have the same curve. And
4 these are the ratings from 1 to 50.

5 So Gur wrote in this 2007 paper, a very
6 large fraction of responses for certain detection
7 tasks are in the extreme ranges of the scale. A
8 similar pattern is not uncommon in reader data that's
9 submitted to the FDA.

10 So if we look back at our ranked scores and
11 ROC curve, we can see that these ranked scores could
12 very well be in the ends, and we could get the same
13 ROC curve or possibly some questionable cases that
14 get scored somewhere in the middle. And the point is
15 the same ROC curve could arise from many different
16 data distributions.

17 Certain tasks that are binary in nature are
18 better represented by a binary endpoint, both
19 conceptually and statistically. In simulations, Gur
20 et al. show that a binary task is evaluated with less
21 bias and less variability if a binary scale, rather
22 than a continuous scale, is used. For a task that is
23 essentially binary, such as detecting
24 microcalcifications, how rigorous can we expect
25 relating rankings to be?

1 A little bit more on ROC endpoints.
2 They're good for comparing tests over all possible
3 cutoffs, and they use information efficiently. I'll
4 talk about a few drawbacks to their use. Now, the
5 difference between ROC curves or the area, the
6 difference in area between two ROC curves can be
7 interpreted as the average difference in sensitivity
8 over all specificities. And I've just drawn two ROC
9 curves here to represent two different modalities.

10 So the question comes up, with something
11 like this, would the difference in area be
12 comparable? Now, once a subtraction is done, these
13 two areas might be very similar. But in this plot on
14 the left, we see that most of the difference in
15 sensitivity occurs in a region of the plot where
16 specificity is close to zero.

17 And over here we see the change in
18 sensitivity is fairly large in a region where
19 specificity is also large. So the question is should
20 we be averaging over all the false positive fraction
21 region? And this would depend on clinical context or
22 the use of the test and the device.

23 So, again, do we want -- and this is
24 something we would have to consider. Do we have this
25 area over here to influence our statistical

1 inference? Perhaps we could use a partial AUC,
2 looking at the difference in partial AUC, with some
3 context-dependent bound or perhaps some other device-
4 specific criteria. And although I haven't seen it
5 used, it may be possible to weight regions of this
6 unit square according to their clinical relevance.

7 Now, using a threshold like sensitivity and
8 specificity as a metric for the threshold, they're
9 intuitive. They are directly applicable to what
10 happens in practice. The patient is sent for a
11 further workup, a biopsy or whatever comes next in
12 the diagnostic procedure or not. We don't need to
13 worry about adapting readers to scores they may not
14 be -- scoring systems they may not be familiar with
15 in a laboratory setting. We won't have those biases
16 to worry about. And they mimic reality. The data
17 we'll get from post-market studies will necessarily
18 be binary threshold because it will be actual data
19 from clinical practice.

20 Now, I'm going to describe an example here.
21 I'm calling this the "Keep All Positives from the
22 Unaided Read" rule, and this is because I couldn't
23 think of a better way to call it. Several second
24 CAD -- second reader CADe devices, or labels, rather,
25 require and imply that positive findings on the

1 initial unaided read should not be negated by the
2 CADe-aided read. And here I've just excerpted some
3 information from labels that contain this rule, these
4 two device labels.

5 Now, if this star represents the
6 sensitivity and false positive fraction for the
7 unaided reader, we can use this keep all positives
8 rule to help us define a hypothesis test. Using this
9 rule, we can see that we will not have any -- it's
10 not possible to decrease in sensitivity when going
11 from the unaided to the CAD-aided modality.
12 Similarly, it's impossible to have a decrease in the
13 false positive fraction if -- going from one modality
14 to the other, if this rule is followed.

15 In addition, we could further bound this
16 region and this is -- this bound is defined in a
17 method similar to a method described in Biggerstaff.
18 But there are other ways that we could use to define
19 bounds and to define a hypothesis test. For example,
20 say it's believed that if the false positive fraction
21 or the false positive rate of a screening test
22 exceeds some value, then we might believe that people
23 would not want to get the test. So this expert
24 knowledge could also be used to define the upper
25 level bound on the false positive fraction in

1 designing a test like this.

2 Such a process could be -- I'm sorry. Let
3 me see. Let me catch my thought here. So this is
4 one example of defining a hypothesis test. Often,
5 researchers will define hypotheses tests in a rather
6 ad hoc manner in choosing, say, a difference to
7 detect. And CDRH is currently pursuing a decision
8 analysis initiative in an effort to make decisions
9 more transparent and reproducible. Such a process
10 could be applied to defining hypotheses tests in a
11 way that quantitatively combine expert knowledge and
12 specific risks posed by false negatives or false
13 positives.

14 Decision analysis often includes defining
15 utilities. Defining hypotheses tests to examine
16 readers' clinical operating points, that is,
17 sensitivity and specificity, is more appropriate than
18 treating all levels of the specificity region as
19 equal and averaging over those.

20 So now I'll talk a little bit about sample
21 size, and this is a sample size needed to answer
22 questions about one specific dataset, and this is
23 going to involve two specific ROC curves. The
24 information that I'm about to provide is from Zhou,
25 Obuchowski and McClish, 2002. So I'm going to use

1 this magnified version of these two ROC curves to
2 explain the information that comes in the next slide.

3 This is a change in sensitivity at a given
4 false positive fraction or a false positive rate of
5 20 percent. This is a change in the partial area
6 under the curves over this interval of false positive
7 rates from 10 percent to 20 percent. And if we
8 divide that change in partial AUC by the interval,
9 we'll get an estimate of average change in
10 sensitivity over that interval.

11 So here are the answers to some of those
12 questions. The sample size necessary to find a
13 difference in AUC between these two specific curves
14 would be 278 images, where half of those images are
15 disease positive and half of those images are disease
16 negative. We see that the sample size necessary to
17 answer some of these other questions varies quite a
18 bit, and the largest being to answer the question how
19 many images would it take to see a difference in
20 about 10 percent sensitivity if we hold specificity
21 at 99 percent?

22 Now, given that FDA has seen sample sizes
23 for reader studies over 600 images, the numbers here
24 and the differences here do not seem prohibitive to
25 answer the questions that we really want to answer.

Free State Reporting, Inc.
1378 Cape Saint Claire Road
Annapolis, MD 21409
(410) 974-0947

1 Now, this is another example of ROC curves,
2 and this is not an uncommon problem where we have
3 crossing ROC curves. It's difficult to interpret
4 this change in AUC in a situation like this. And
5 perhaps a post hoc, partial AUC -- look at the
6 difference in partial AUC might be a rescue. But if
7 we were to do that, we have to worry about Type I
8 error implications of choosing that bound after
9 seeing the data.

10 Now, the point of this is if we were to
11 have sized this study for a change in AUC over the
12 whole interval, there may not be enough information
13 to look at differences in partial AUC or differences
14 in sensitivity and specificity to answer our study
15 questions. And inadequate information is, to me, the
16 definition of a failed study. You don't get anything
17 out of it.

18 So I talked a little bit about endpoints
19 and the fact that sensitivity and specificity are
20 more relevant than ROC AUC in the dichotomous
21 decisions made in some readings. I showed some
22 drawbacks to using ROC analysis, such that they're
23 not always easy to interpret. We could have crossing
24 curves. We have to adapt readers to the scoring
25 system for the laboratory environment.

Free State Reporting, Inc.
1378 Cape Saint Claire Road
Annapolis, MD 21409
(410) 974-0947

1 And, finally, I've just included a quote
2 from the 2008 Panel by Dr. Berry, where he's
3 suggesting that sensitivity be the metric to look at
4 in this type of analysis. And there's a few other
5 references that I've listed here that support using
6 sensitivity and specificity in screening-type exams
7 or the one by PEPI (ph.), specifically to
8 mammography.

9 I'll talk a little bit about control arms.
10 It's assumed that the effectiveness or the clinical
11 utility can be shown by comparing unaided image
12 reading to CADe-aided image reading. So as
13 Dr. Petrick said, we have several questions we'll ask
14 you about control arms.

15 Now, what I've done here is just a little
16 schematic to represent a non-inferiority test, and
17 this is a confidence interval for the difference in
18 reader performance using, say, a newer CADe device
19 compared to a predicate device and the success
20 criteria being that the confidence interval is
21 greater than some predetermined limit.

22 So I'll just briefly discuss two possible
23 study designs, the first one being the readers read a
24 common set of images under three modalities, unaided
25 reading, CADe-aided reading with the study device,

Free State Reporting, Inc.
1378 Cape Saint Claire Road
Annapolis, MD 21409
(410) 974-0947

1 and CADe-aided reading with the predicate. We'd
2 randomize image order and use a washout period of,
3 say, four weeks, for example, to minimize the
4 possibility of memory biases and then compare the
5 performance results. The unaided reading, including
6 unaided reading modality, ensures that we can find
7 the difference between the unaided reader and the
8 CADe-aided reader to show clinical utility. And in
9 this type of a study, we could define a
10 non-inferiority delta.

11 Now, here's another possible study design
12 where we would look at unaided reading versus CAD-
13 aided reading and use the same sort of procedures
14 where we randomize image order, use washout
15 procedures, and then compare our results in that
16 study to label information or prior study
17 information.

18 Again, considering this second design, if
19 we have the two studies, say, and in the predicate
20 study, the study contained relatively difficult-to-
21 detect diagnosed images and used mostly experienced
22 readers, while the new device study, say, had mostly
23 easier-to-detect diagnosed images with less
24 experienced readers, but in the end, the differences
25 in performances looks similar.

1 So as we talked about earlier, we have a
2 wide variety of performance levels for readers, and
3 we also talked about spectrum bias. So in Design 2,
4 the comparison across studies is confounded by
5 spectrum bias and reader differences. Using such a
6 study design comparing changes across enriched
7 studies effectively reduces the question of
8 substantial equivalence to one of whether or not the
9 CAdE device offers any increase in performance over
10 the unaided read. With respect to performance
11 comparing across enriched studies, it invites
12 imprecise or erroneous substantial equivalence or
13 non-substantial equivalence conclusions due to
14 confounding. And this could be confounding from case
15 mix differences, reader differences, differences in
16 the laboratory environment, lighting, any number of
17 differences between the studies.

18 So I've just updated my little schematic of
19 the non-inferiority test to include the need to
20 compare the devices in the same study and also the
21 requirement that there is some clinical utility over
22 the reader alone shown in the study -- through the
23 study.

24 Now, I'll talk a little bit about
25 standalone studies. Standalone studies cannot show

1 clinical utility because there's no reader involved.
2 Standalone studies may be useful in comparing a CADE
3 device to a previous version or investigating the
4 performance of the device without the reader. An
5 example where this would be especially beneficial is
6 studying a sample large enough to characterize all
7 important strata. This would be disease, differences
8 in lesion types, and also differences in non-disease
9 cases that would potentially mimic lesions and
10 produce false positives. And this would be useful
11 information to have in the labeling of these devices.

12 Enriched studies suffer the same
13 complications as reader studies with respect to
14 sample enrichment. The results are not generalizable
15 across studies. Performance estimators apply only to
16 the sample. They're not simple random samples of the
17 population, and they do not represent standalone
18 performance in the population.

19 And now I'll talk a little bit about reuse
20 of test data. Some companies have proposed reusing
21 test data in evaluating updated versions of their
22 devices. And specifically, we're talking about
23 complete reuse of the data, the same datasets being
24 submitted as in previous studies. And I'll leave it
25 at that.

1 So if we're thinking about examining
2 multiple updated versions of a device, we're really
3 looking at different devices. And if we test
4 different devices on the same set of data, usually
5 sponsors must account for multiplicity and a
6 Bonferroni-type correction is often used. And if we
7 were to do that, how do we think about Type I error?
8 How do we think of alpha at .05 when it's not clear
9 that these sorts of reusing the data tests would be
10 preplanned and there'd be some set number of them?

11 Another worrisome problem is the teaching
12 to the test. Each upgrade iteration on the same data
13 could be considered training. The data is being seen
14 by the users, and something is being learned about
15 that dataset. So after a series of iterations, you
16 would expect that something would be -- some
17 knowledge would be gained about the dataset. And
18 this is going to add bias to the process of
19 evaluating these devices, and it's difficult to
20 quantify this bias.

21 Now, I'm going to present an example here.
22 This example is from Simon, et al. It illustrates
23 the problems of over-fitting in the context of
24 developing algorithms for class predication with gene
25 expression data. The large number of features within

1 relatively small samples make this a good parallel to
2 the situation faced by CADE developers. Simon,
3 et al. randomly generated datasets of 20 profiles
4 having 6,000 features each. Then they arbitrarily
5 assigned each to one of two classes. They developed
6 and evaluated classifiers using three processes. The
7 first they called a nearly unbiased class validation
8 method, and in this method they chose one member of
9 this 20-member group, and they left it out. They
10 built the classifier on the remaining data, and they
11 classified that last point. And then they went
12 through that process to look at all 20 of those
13 points.

14 Next is the re-substitution method, or what
15 Simon, et al. called the re-substitution method. And
16 this is similar to what would happen if we allowed
17 unlimited reuse of data, continually looking at the
18 same dataset teaching to the test. This example is
19 extreme, but it is the limit where we would end up if
20 we allowed this unconditionally. So the way this
21 would work is you'd build the predictor or the
22 classifier on the full dataset, and then you reapply
23 the predicator to each specimen. And the third
24 method I'm not really going to talk about, but it's
25 somewhere in between those two.

1 And this is repeated 2,000 times to get
2 those data. Now, I'll point out here that this Y
3 axis has been truncated. We go from .1 to .9. The
4 re-substitution method, which I'm calling the
5 teaching to the test and that's similar to looking at
6 the data over and over again, 98.2 percent of
7 datasets had zero misclassifications. They were
8 perfect in classifying every point in those 20-point
9 datasets 98 percent of the time. But we have to
10 remember that this is random data, so we would expect
11 about half of the items in these 20 member sets to be
12 misclassified. We'd expect to see a distribution
13 like this, and this is what they see with their
14 nearly unbiased method of class validation. So this
15 illustrates the problems we could potentially face if
16 we were to continually reuse datasets testing the
17 same device or upgraded versions.

18 Any variation of reusing data would raise
19 many difficult review issues, such as data integrity
20 and access controls. Who has access to the test data
21 and when? Do we have a theoretical basis for the
22 procedures? Are they published and are the
23 assumptions verifiable? And selection bias. How
24 were the images chosen to go into this dataset that
25 would be continually reused? And last but not least

1 is how is Type I error controlled in such a
2 situation?

3 Now, I'll move on to talk about using only
4 standalone data in clearance for approval of a CADe.
5 A change in marker style can affect reader behavior,
6 and I've listed two references here that discuss
7 this. Changes in prevalence, as we discussed earlier
8 and is discussed in the paper by Egglin et al.,
9 changes in prevalence affect reader behavior. Now,
10 we can deduce that changes in CADe-marked placement
11 or frequency could impact reader behavior. A change
12 to the algorithm is a change to the device. The
13 device is acting on reader diagnosis. It's difficult
14 to know a priori what change to an algorithm will
15 produce a change in diagnostic performance. Reader
16 studies investigate reader-device interaction.
17 Standalone studies investigate only device
18 performance.

19 So to summarize my discussion, I talked
20 about endpoints for reader studies and that binary
21 endpoints are more relative to the study question. I
22 talked a little bit about sample size and what sort
23 of an endpoint would we recommend sizing the studies
24 for. I talked about control arms and the fact that
25 if we compare across studies, we are essentially

1 reducing the question to is a device better than the
2 reader alone? We just talked about reusing test data
3 and the implications that has, the problems for bias
4 and incorrect -- problems with performance of a
5 device tested on reused data. And, finally, I talked
6 about evaluating CADes without readers and the fact
7 that such an evaluation would not represent clinical
8 practice or the reader/device interaction. Thank
9 you.

10 DR. D'ORSI: Thank you, Dr. Gwise. Let's
11 take 15-minute break. Come back here at 10 after
12 10:00.

13 (Off the record.)

14 (On the record.)

15 DR. D'ORSI: And the next presenter will be
16 Dr. Robert Smith on the clinical view. Why don't we
17 begin? And it's like herding cats. So just begin
18 and --

19 DR. SMITH: It seems like you're missing a
20 lot of the Panel members.

21 DR. D'ORSI: Okay, let's start. It looks
22 like a quorum, at least.

23 DR. SMITH: Let me just wait for Janine to
24 come back.

25 MS. MORRIS: I'm right here.

Free State Reporting, Inc.
1378 Cape Saint Claire Road
Annapolis, MD 21409
(410) 974-0947

1 DR. SMITH: Oh, thank you. Good morning.
2 My name is Robert Smith. I'm the radiologist in the
3 Radiological Devices Branch at CDRH. I was a
4 practicing radiologist for 15 years before I came to
5 the Agency. My areas of specialty are MRI, CT,
6 ultrasound, and medical imaging physics. And I would
7 like to acknowledge one of my FDA colleagues,
8 Dr. Sophie Paquerault, who provided assistance in
9 putting together this talk.

10 I'm here to talk about the clinical
11 viewpoint, although actually this is just another
12 statistical talk disguised as a clinical talk. I'm
13 just kidding. I'm going to talk about CADe in
14 clinical practice, the benefits, risks, and clinical
15 significance of CADe, very briefly, on the proposed
16 draft CADe guidance documents, since you've already
17 heard a lot about that. I'm going to come back to
18 the regulation of CADe devices as proposed in the
19 draft guidance, which I think is quite important.

20 I'm going to go back and look at what was
21 said at the March 2008 Panel meeting on this similar
22 topic. And then I'm going to talk briefly about the
23 state of the science. And then I'm going to spend a
24 lot of time talking about the issues and questions
25 for the Panel, and I'm going to go through it almost

1 in order, but it may go slightly out of order in the
2 numbering of the questions.

3 It's important to understand how
4 radiologists interpret images in order to identify
5 where and how CAdE devices might be helpful in a
6 clinical interpretation paradigm. So as
7 radiologists, we do detection, description, and the
8 description over the years has been of
9 characteristics that are helpful in distinguishing,
10 you know, abnormal from normal findings. Then we try
11 to come up with a differential diagnosis, that is, we
12 analyze the findings.

13 And finally we end up with a recommendation
14 for what I'll call the clinical action. And the
15 clinical action could be no further action. It could
16 be additional imaging the same day. It could be
17 additional follow-up imaging the following day or a
18 couple months later. We could recommend a
19 non-imaging diagnostic test. We could recommend
20 biopsy. It'd be pretty unusual, but I guess could
21 recommend surgery.

22 So what are the effects of using these CAdE
23 devices? Although they're intended to aid in the
24 detection task, each step in the image interpretation
25 cascades down the interpretation chain and ultimately

1 into the clinical action. Therefore, use of these
2 devices can affect diagnosis as well as the clinical
3 action.

4 So what are some of the benefits, risks,
5 and clinical significance? And these are important
6 when the Agency is weighing safety and effectiveness.
7 The factors that determine these benefits and risks
8 will be device specific. They can be organ specific.
9 They can be disease specific. They could even be
10 imaging finding specific, and that could vary on the
11 significance of false positive or false negative
12 findings, and it also is going to come down to the
13 clinical action.

14 I'm going to try to talk about specific
15 CADe devices. There are really too many of these to
16 talk about in a single talk, but I will talk about
17 some of the more common ones and some of the ones
18 we've seen at the Agency.

19 For screening mammography CADe devices, the
20 breast is the organ of interest, breast cancer is the
21 disease of interest, and the specific imaging
22 findings that these devices may try to detect are
23 masses, microcalcifications, architectural
24 distortion, perhaps focal asymmetry.

25 And a specific false positive imaging

Free State Reporting, Inc.
1378 Cape Saint Claire Road
Annapolis, MD 21409
(410) 974-0947

1 finding may actually be associated with a different
2 risk because of different clinical action. If you
3 have a false positive mass, well, you may end up just
4 doing an ultrasound. Very minimal, if any, risk. If
5 you have false positive microcalcifications, you may
6 need to get some additional X-ray imaging, which
7 includes some additional radiation. It could lead to
8 a biopsy. Architectural distortion could lead to a
9 biopsy or other kinds of imaging. It could even lead
10 to MRI. These are just some examples. But you need
11 to consider these when you're considering the
12 benefits and risks of these devices.

13 A false negative screening mammogram could
14 result in a delay of diagnosis. I have here for one
15 year, given current screening practice in the United
16 States. I suppose that could change, based on the
17 recommendations that came out the other day, but I'm
18 really not going to comment on that.

19 CT colonography CAde devices. The colon is
20 the organ of interest. You're looking for colonic
21 polyps. I guess a manufacturer could try to say
22 they're looking for adenomatous versus
23 non-adenomatous. But you're really just looking for
24 the imaging finding of a polyp. That's the disease
25 or condition.

1 And they can be falsely positive, and that
2 can be based on the number and/or size of detected
3 polyps because that will usually determine the
4 clinical action. And this again may be associated
5 with different risks, depending on what the clinical
6 action is. It may just be you do an optical
7 colonoscopy, which probably was going to be done
8 anyway. There's minimal risk of that, you know, up
9 to a point. Or it may be, if it's a surveillance CT
10 in somebody who has known adenomatous polyps, you may
11 do a closer interval follow-up.

12 A false negative result on a CT could have
13 a lot more implications. If that's going to be used,
14 say, if it's being used for screening to stop a
15 patient who would otherwise have a colonoscopy, if
16 the patient is in a low-risk category, there could be
17 a delay of a diagnosis of an adenomatous polyp or
18 even cancer by three to five years.

19 So a false negative has a much greater
20 significance than a false positive because current
21 clinical practice is to obtain the optical
22 colonoscopy anyway. Again, that's up to a point.
23 Obviously, if you're calling everybody positive, it
24 wouldn't make much sense.

25 So when you're assessing the safety and

1 effectiveness of CT colonography, you probably should
2 give significantly greater weight to the false
3 negative exam because that's going to have a lot more
4 implications for the patient.

5 I just briefly want to talk about the
6 proposed guidance document. As Dr. Whang described
7 this morning, one is specific to 510(k) submissions.
8 The other that relates to clinical performance
9 applies to a 510(k) and PMA. It's important to
10 remember that these guidance documents cover a lot of
11 different possible devices, diseases, et cetera, so
12 they really provide general guidelines about CAdE
13 devices, and they were not written to be specific to
14 a particular organ or disease.

15 And the clinical one is obviously
16 applicable to Class III and Class II devices. It's
17 safety and effectiveness for PMA, and as I'll get
18 into in more detail again a little later, substantial
19 equivalence for a 510(k), which, remember, if there
20 are differences in technological characteristic, it
21 means at least as safe and effective.

22 And these guidance documents request basic
23 elements like device description, device standalone
24 testing, describe the clinical testing. And
25 virtually all the guidance documents, these in

1 particular, say that these do not -- these guidance
2 documents are guidance. It's the current thinking of
3 the Agency. They don't create or confer any rights
4 for any person. It doesn't operate to bind FDA or
5 the public, and the Agency will always consider
6 alternative approaches, if those approaches satisfy
7 the statutory requirements.

8 I just want to talk a little bit about the
9 regulation of CADE devices as proposed in the draft
10 guidance. It's important to remember that FDA does
11 not clear or approve general CADE technology. We
12 clear or approve individual specific devices based on
13 safety and effectiveness for the intended use.

14 Most Class III devices, as has been
15 described, are cleared -- excuse me -- are approved
16 for marketing through a PMA, and unlike a 510(k), a
17 PMA, it can be but it's not typically or doesn't have
18 to be a comparison to other legally marketed devices,
19 but the data would generally stand on its own to
20 demonstrate safety and effectiveness for the intended
21 use.

22 The 510(k) applies to most Class II
23 devices. The term there used is clearance as opposed
24 to approval. And here the manufacturer needs to
25 demonstrate substantial equivalence to another

1 legally marketed device. That is referred to as the
2 predicate device.

3 And the reason intended use is important,
4 and I'm just going to touch on it today, is because
5 this guidance document does cover both PMA and 510(k)
6 devices. Intended use is defined in the regulations,
7 21 C.F.R. 801.4, and these are excerpted directly
8 from the regulation. Intended use -- and this
9 applies to a PMA device -- it's the objective intent
10 of the persons legally responsible for the labeling
11 of a device and the intent is determined by such
12 persons' expressions or may be shown by the
13 circumstances surrounding the distribution of the
14 device.

15 The objective intent may, for example, be
16 shown by labeling claims, advertising matter, or oral
17 or written statements. It may be shown by the
18 circumstances that the article is, with the knowledge
19 of such persons or their representatives, offered and
20 used for a purpose for which it is neither labeled
21 nor advertised.

22 And the intended use of a 510(k) device.
23 This is defined in Section 513 of the Food, Drug and
24 Cosmetic Act, which states -- and these are,
25 directly, excerpts from the Act -- "Any determination

1 by the Secretary of the intended use of a device
2 shall be based upon the proposed labeling submitted
3 in a report for the device under Section 510(k)."

4 "'Labeling' is defined in Section 201(m) of
5 the Act as 'all labels and other written, printed, or
6 graphic matter upon any article or any of its
7 containers or wrappers, or accompanying such
8 article.'"

9 The important thing to remember is that the
10 "Proposed labels, labeling, and advertisements
11 sufficient to describe the devices, its intended use,
12 and the directions for use are required to be
13 submitted in a 510(k)." So that information is
14 required to be in the 510(k) "for review during the
15 Substantial Equivalence determination."

16 Just to go back to the definition of
17 safety. This is again directly from the statutes.
18 Obviously something we use all the time.

19 "There is a reasonable assurance that a
20 device is safe when it can be determined, based upon
21 valid scientific evidence, that the probable benefits
22 to health from use of the device for its intended use
23 and conditions of use, when accompanied by adequate
24 directions and warnings against unsafe use, outweigh
25 any probable risks."

1 Effectiveness is defined as follows:

2 "There is a reasonable assurance that a
3 device is effective when it can be determined, based
4 upon valid scientific evidence, that in a significant
5 portion of a target population, the use of the device
6 for its intended use and conditions of use, when
7 accompanied by adequate directions for use and
8 warnings against unsafe use, will provide clinically
9 significant results."

10 510(k) clearance of CAdE devices. I know
11 we've talked about this a number of times, but we
12 talk about it all the time at the Agency, and it
13 never hurts to go over it again. There are two
14 distinct paths based on technological
15 characteristics.

16 Path 1, a device is substantially
17 equivalent if, compared to a predicate, it has the
18 same intended use and the same technological
19 characteristics. If that's the case, we can stop
20 right here. Nothing else needs to be done. It will
21 be found substantially equivalent.

22 If, however, there's a difference in
23 technological characteristics, then we go down to the
24 second path, assuming it has the same intended use
25 and has different technological characteristics,

1 which again can be a change in material, design,
2 energy source or, what we're most interested in here
3 today, software, and the information submitted to FDA
4 does not raise different, that is, new types of
5 questions of safety and effectiveness, and
6 demonstrates that the device is as safe and effective
7 as the legally marketed predicate device.

8 I just want to briefly go back and just
9 provide some excerpts to you from the March 2008
10 Panel meeting because I think that will be helpful.
11 Here are some excerpts relevant to mammography CAD
12 devices. This was Question M2(a), to discuss the
13 role of reader performance testing in the clinical
14 evaluation of CAD devices.

15 Part (a) was, if you believe reader
16 performance testing should be considered, please
17 provide your comments on the following: the
18 appropriate primary endpoints; the merits of per
19 lesion, per view, per breast, per patient endpoints;
20 whether the effectiveness analysis should be
21 conducted separately, or not, for cancers manifesting
22 as masses and microcalcifications; and if reading
23 time should be assessed.

24 This is an excerpt following the
25 discussion. "That reader studies should be the

Free State Reporting, Inc.
1378 Cape Saint Claire Road
Annapolis, MD 21409
(410) 974-0947

1 primary analysis is unchallenged and the opinion of
2 the Committee." It describes Question Subpart (i):
3 "ROC analyses were thought to be a good thing, but we
4 really couldn't get anybody to commit much as to how
5 good."

6 An additional comment was made after all
7 the questions were answered that also pertained to
8 ROC. "The shape of the ROC curve may be very unusual
9 in certain cases. Just looking at the overall area
10 under the curve is probably not going to be
11 meaningful in itself, that you are going to have to
12 look at the shape of the curve and you may have to go
13 to a partial area under the curve in the areas of
14 greatest interest to look for small differences
15 between tests; otherwise, they are going to be
16 swamped by the overall distortions in the curves."

17 With regard to Subpart (ii): "That per
18 patient endpoints with the reader study is very
19 important, although per lesion and per view should
20 not be completely ignored."

21 Subpart (iii): "Whether effectiveness
22 should be conducted for cancers with different
23 findings (i.e., microcalcifications versus masses),
24 the answer was clearly yes."

25 For Subpart (iv): "It was the sense of the

1 Committee that reading time is an important factor
2 for labeling."

3 Question M2(b): "Are there specific
4 situations where reader performance testing may not
5 be necessary?" One comment went unchallenged. That
6 for minor modifications, that standalone testing
7 would probably be sufficient. And I underlined the
8 word minor because we never really got to what minor
9 means, and I'm going to come back to that because I
10 think I might be able to provide you some help, since
11 you asked for help at the last Panel meeting.

12 Here are some excerpts that are relevant to
13 CT colonography CADe. Question C3 was, Please
14 discuss the role of standalone performance testing in
15 clinical evaluation of CADe devices, colon CAD
16 devices.

17 Subpart (a) was if you believe standalone
18 testing should be requested, provide your
19 recommendations and comments on whether certain
20 substrata, pathology, et cetera, are important.

21 Subpart (b): If you believe that there are
22 specific situations where standalone performance
23 testing may not be important, please comment.

24 This is an excerpt from the Panel summary
25 following the discussion. "Standalone testing is

Free State Reporting, Inc.
1378 Cape Saint Claire Road
Annapolis, MD 21409
(410) 974-0947

1 important. Polyp size, minimum six millimeters and
2 larger. The CT dose and imaging protocol needs to be
3 known to make sure that it's clinically relevant. We
4 want to enhance the set with polyps of varying
5 locations, including the flexures which may be more
6 difficult to find."

7 "We want to enhance with flat polyps that
8 we know are more difficult to find. And we want to
9 know the demographics of the test set to make sure
10 that it is clinically appropriate with the usual
11 patient population. And that we believe that
12 standalone testing is important in all instances."

13 Question C4: Please discuss the role of
14 reader performance testing in the clinical evaluation
15 of colon CAD devices. Do you think it should be
16 considered in the evaluation and, therefore, what are
17 the appropriate primary endpoints? Comment on ROC
18 analyses. Comment on the merits of per lesion, per
19 segment, or per patient endpoints and whether reading
20 time should be assessed, and if so, how?

21 These are excerpts again from the summary
22 discussion. Reader performance testing should be
23 done. Clinically effective sizes are greater or
24 equal to six millimeters. ROC analysis is
25 appropriate, as is FROC analysis, but either one.

1 The general consensus was that the endpoint should be
2 per lesion, rather than per segment and per patient,
3 knowing, however, that if we choose ROC analysis, it
4 would be converted to a per patient analysis.

5 And Subpart (b), Question C4: If you
6 believe that there are specific situations where
7 reader performance testing may not be necessary,
8 please comment on what these might be.

9 Again, we're back to minor changes in the
10 algorithm, that standalone testing would be adequate,
11 but we still have to decide what minor would be. And
12 then a comment was made: "That's what the FDA is
13 going to do for us, I hope." Well, partly we're
14 here -- you guys are here back today because we'd
15 like you to help us with that, although I will
16 provide some advice on what I think about that. But
17 we really do want to hear from you.

18 What's the current state of the science
19 with these devices? What does the scientific
20 literature reveal about mammography CAdE? These
21 devices have been well studied for at least 10 years.
22 There's a lot of literature out of there. And
23 standalone testing has shown a very high sensitivity
24 to mark calcifications; much lower sensitivity to
25 mark masses, architectural distortion, or focal

1 asymmetry; and a false positive mark rate generally
2 between two and four marks per patient. The more
3 recent devices are certainly at the much lower end of
4 this.

5 Reader performance testing has shown
6 conflicting results for the testing of invasive
7 cancers, still a bit of a controversy. There is
8 certainly a trend toward CAD improving radiologic
9 detection of calcifications, especially DCIS, by
10 readers. There's no improvement for the detection of
11 masses in the larger studies, no definite statistical
12 improvement, and there is an increased recall --
13 there can be an increased recall rate when using
14 these devices. In some studies, the increases are
15 statistically significant, but not always.

16 What does the scientific literature reveal
17 about CT colonography CADe? Well, here I think it's
18 very important to look at how can you do -- how well
19 can you do without the CADe devices because that
20 tells you, well, how much room is there for
21 improvement and which areas might improvement be best
22 utilized for?

23 Per patient sensitivity is approximately
24 0.9 in the largest recent studies for polyps that are
25 greater than or equal to 10 millimeters. Per polyp

1 negative predictive value, extremely high for polyps
2 greater than or equal to 10 millimeters. The per
3 patient negative predictor value, which is very
4 important for these devices, is also extremely high,
5 even going down to six millimeters.

6 What about CT colonography with CADe?

7 Well, there's much more limited data in the published
8 literature. There certainly is a lot of evidence
9 that CADe tends to detect the 6- to 10-millimeter
10 polyps that are more visually conspicuous, that is,
11 less flat polyps. It seems to have more difficulty
12 with the flat polyps. And there's a quote here from
13 a recent publication by Dr. Summers. "CADe
14 developers may need to specifically target flatter
15 and less conspicuous polyps for CAD to better assist
16 the radiologist to find polyps in this clinically
17 important size category."

18 Now, I want to come to and I'm going to
19 spend a lot of time on, which are really the
20 questions for the Panel. I'm going to start with
21 Question 2(a). And I know you've already heard some
22 materials about this in the statistical talks and
23 that relates to valid control arms.

24 Going back to 510(k)'s, the control arms
25 for a 510(k) to demonstrate substantial equivalence,

1 remember, a new device needs to be compared in some
2 manner to a legally marketed predicate device because
3 it has to be shown to be at least as safe and
4 effective as the legally marketed device. And a
5 comparison can be direct or indirect. In a direct
6 comparison, you would use the predicate device as the
7 control arm. In the indirect comparison, you could
8 use an unaided read as the control arm to estimate
9 the clinical impact of the new CAD on one set of
10 cases and readers and try to somehow compare that to
11 the estimated clinical impact of the predicate
12 device, which would usually have been obtained on a
13 different set of cases and readers. It would usually
14 be reported in the labeling of the predicate device
15 or perhaps in the published literature.

16 When you use the predicate device, you're
17 using a direct comparison. There you can have device
18 standalone performance testing using the same cases,
19 the same scoring methodology, the same ground-truth
20 methodology, et cetera. And for clinical performance
21 testing, you could use the same cases and the same
22 readers. It's fairly straightforward.

23 Question 2(b) was under what conditions the
24 Agency should consider accepting a surrogate
25 endpoint, such as standalone performance, in lieu of

1 clinical performance data for CADe?

2 If the control is the predicate device, can
3 device standalone performance testing be sufficient
4 for a new device, which would include an updated
5 version of a previously cleared device, or perhaps
6 the addition of a new image input to a previously
7 cleared device? And it might be sufficient. It
8 would be sufficient, for example, if the new device
9 and the predicate device show the same number and
10 type of CAD marks. The marks appear at the same
11 location and the marks have the same prompting
12 formats because it's important to remember, even
13 changes in prompting formats may influence the
14 reader. Otherwise you may very well need clinical
15 performance testing.

16 When you use the unaided read, that's the
17 indirect comparison. And the device standalone
18 performance testing and the clinical performance
19 testing on the predicate was typically done by the
20 manufacturer of the legally marketed device. And it
21 would typically have been done using different cases
22 and different readers, again, the different scoring
23 methodology and different ground-truth methodology,
24 which can obviously create complications when in a
25 510(k) you're trying to make a comparison.

Free State Reporting, Inc.
1378 Cape Saint Claire Road
Annapolis, MD 21409
(410) 974-0947

1 Just an example, if you're using the
2 control as an unaided read, you're trying to estimate
3 the clinical impact of the new device compared to the
4 estimated clinical impact of the predicate device.
5 Again, that's as reported usually in the labeling of
6 the device. And here's an example. Assume the
7 estimated clinical impact of the new CAde device is
8 identical to the estimated clinical impact of the
9 predicate device. But suppose that the new device
10 testing was done on cases that were easy to detect
11 and used inexperienced readers whereas the predicate
12 device testing may have been done on difficult-to-
13 detect cases with experienced readers. So you end up
14 with the same outcome, but this is clearly not a
15 valid comparison for demonstration of substantial
16 equivalence because of the differences in the cases
17 and the readers, even though you got the same
18 outcome.

19 As far as control arms for a PMA
20 supplement, I'm really not going to get into that
21 very much. The statutory regulations are different.
22 In general, if you have a PMA-approved device and you
23 make a change that affects safety and effectiveness,
24 you need to submit a supplement. But the specific
25 conditions are described in 21 C.F.R. 814.39 for

1 those who are interested.

2 Question 3(e): Manufacturers typically
3 report ROC curves, the area under the curve,
4 sensitivity, specificity, for clinical performance
5 studies. Should the studies be powered for all
6 summary endpoints? If not, which endpoints should be
7 used to size or power the study? It's my opinion --
8 I'm the clinician here -- that endpoints must be
9 based on clinical action. It's important to
10 remember, then, the guidance document was drafted to
11 be generally applicable to all CAdE devices. But I
12 think it's implicit that in order to evaluate device
13 safety and effectiveness based on the definitions
14 that I've given you, the endpoints must be selected
15 based on the clinical action for the appropriate
16 organ, disease, or imaging finding that the device is
17 intended to be used for.

18 For CT colonography, the clinical action is
19 taken on the patient, so I believe the primary
20 endpoint should be at the patient level. The
21 clinical action is typically based on the number and
22 size of polyps, but there isn't a uniformly accepted
23 means of deciding on that clinical action based on
24 the number and size.

25 There is some controversy in the

Free State Reporting, Inc.
1378 Cape Saint Claire Road
Annapolis, MD 21409
(410) 974-0947

1 literature. The radiology literature has come up
2 with a system similar to the BI-RADS system to
3 interpret mammography. And this is a recent
4 publication. It came out not that long ago, by Rex
5 et al. It was an estimation of impact of -- he
6 called it ACR recommendations, but it was really the
7 C-RADS published previously on CT colonography
8 reporting for resection of high-risk adenomatous
9 findings. And they found that if CTC, rather than
10 colonoscopy, were used and they assumed 100 percent
11 sensitivity of CTC for polyps greater than or equal
12 to six millimeters, and they used what they called
13 the ACR interpretation, which is really the
14 C-RADS recommendations, they found that 29 percent of
15 all patients and 33 percent of screening patients
16 greater than or equal to 50 years, with high-risk
17 adenoma findings, would've been interpreted as
18 normal, and an additional 18 to 23 percent of these
19 groups with high-risk adenomatous findings could've
20 had a polypectomy delayed for at least three years.

21 So this study looked at over 10,000 cases.
22 It's from the gastroenterology literature. So I
23 think there is some controversy as to how best to use
24 the findings on CT colonography, whether you're using
25 CAD or not, in order to manage the patients. But

1 clearly it has very important implications when
2 you're assessing the safety and effectiveness of a
3 device.

4 The clinical benefit of CT colonography --
5 this is not necessarily just related to CAD devices
6 -- does remain uncertain. There is some
7 controversies in the literature. This is from the
8 Center for Medicare and Medicaid Services. They made
9 a coverage determination in May of this year, on
10 May 12th. This is an excerpt from that.

11 "We have determined that there is
12 insufficient evidence on the test characteristics and
13 performance of screening CT colonography in Medicare-
14 aged individuals." So this really looked at just
15 screening and Medicare-aged individuals. CT
16 colonography can also be used for surveillance, which
17 would be a diagnostic use, and it obviously can be
18 used in younger patients in the Medicare population.
19 So it's just important to remember the context of
20 this determination.

21 There is a difference in the relative risk.
22 I think I touched about it before on false positive
23 versus false negative. The false negative has a much
24 greater significance than a false positive because,
25 again, current clinical practice is to obtain optical

1 colonoscopy, and this needs to be taken into account
2 when assessing safety and effectiveness.

3 What about mammography CADe devices? There
4 the clinical action is clearly based on the type and
5 location of findings, and so the primary endpoints
6 should be based on the finding type and must include
7 location. So here you may want to use FROC, LROC,
8 some type of methodology that does include location
9 if you use an ROC methodology.

10 The clinical action example for positive
11 screening mammography is diagnostic mammography.
12 There the risk is related to the additional
13 radiation. It can lead to biopsy. The clinical
14 action when you're using diagnostic mammography,
15 again, for a positive mass, may be ultrasound, which
16 has minimal, if any, risk and it has near 100 percent
17 positive predictive value for a simple cyst. The
18 clinical action when you have a positive for a
19 microcalcification may very well be biopsy, and
20 generally the biopsy has a predicative value of about
21 30 percent when biopsy is performed. A false
22 negative mammographic finding may delay diagnosis.

23 So what is the relative risk of a false
24 positive versus a false negative? Again, it's
25 important. The relative risk of a false positive is

1 variable, and it will depend on the exam type,
2 diagnostic versus screening, the finding type, mass
3 versus microcalcification. And so, again, you need
4 to take that into account when assessing safety and
5 effectiveness or when deciding on a metric to measure
6 your performance.

7 CADE for chest CT. If you're detecting
8 lung nodules, the clinical action for a positive
9 chest CT is going to be a lung biopsy of one or more
10 the nodules at specific locations. CT-guided lung
11 biopsy does have a significant adverse event rate, at
12 least from the published literature, of about 10 to
13 20 percent of pneumothorax.

14 For CADe chest X-ray, again, if you're
15 looking for lung nodules, the clinical action for a
16 positive chest X-ray is going to be a chest CT. The
17 clinical action for a negative chest X-ray is likely
18 to be no further imaging follow-up.

19 There are a lot of different performance
20 metrics, and I'm not here advocating any particular
21 performance metric. What I am here to let you know
22 is that at the FDA, we'll keep an open mind and we
23 consider any relevant performance metric. If
24 sensitivity and specificity is the relevant metric
25 based on the clinical action, then that's what should

1 be included in the testing.

2 ROC may be appropriate. Remember, it's not
3 location specific. You may want to consider using a
4 partial area in the clinically relevant portion of
5 the curve, as was discussed previously. You have
6 location-specific LROC. You also have FROC. There's
7 lots of different methodologies, and in fact, there's
8 a nice publication I'll come to shortly.

9 For sensitivity and specificity, as a
10 clinician, these are the things that are most easy
11 for us to follow and understand. Sensitivity, the
12 probability of the test is positive for when disease
13 is present. Specificity, the probability of the test
14 is negative when disease is absent. These are well
15 known and understood by clinicians, and they're more
16 closely tied to clinical practice.

17 This is just an example just to make the
18 point that if you're using ROC analysis, there may be
19 a very small portion that's clinically relevant.
20 Here, at a false positive rate greater than 10
21 percent, you know, whatever the test was, you may
22 want to cut it off. For example, for screening
23 mammography, generally it's recommended that you have
24 a false positive rate of 5 to 10 percent. So that
25 may be the relevant portion of the curve when you're

1 looking at screening mammography. The point I'm just
2 trying to make is that you really need to pick your
3 metric based on the clinical action and the safety
4 and effectiveness.

5 This is a recent publication that just came
6 out in the *Journal of the American College of*
7 *Radiology*. It was entitled "ROC, LROC, FROC, AFROC:
8 An Alphabet Soup." And these are just two excerpts
9 from that article.

10 "A strong consensus is emerging in the
11 medical imaging community on the necessity of using
12 task-based image quality assessment, where image
13 quality is quantified based on the performance of an
14 observer on a clinically relevant task."

15 "Like all other techniques, evaluation
16 techniques themselves are techniques that must be
17 evaluated. Currently, research is underway in the
18 medical imaging community relating to evaluate the
19 appropriateness of the competing approaches for
20 analyzing both FROC data and multi-class ROC data."

21 Question 5(a): The following questions
22 seek input on the regulatory significance of minor
23 modifications to CADe devices. I'm going to try to
24 spend a fair amount of my remaining time on this
25 topic.

1 The modifications can be to a device
2 previously cleared through a 510(k); a new 510(k) for
3 the modified device; a device previously approved
4 through a PMA, if a PMA supplement comes in.

5 For 510(k) devices, what modifications
6 require submission of a new 510(k)? And we touched
7 on this briefly yesterday. There is an FDA guidance
8 document on this, which I think is very instructive
9 on this issue of what's a minor modification, and
10 that's the guidance document deciding when to submit
11 a 510(k) for a change to an existing device.

12 And this is a reproduction of what I think
13 is the key chart from this guidance document, and I'm
14 just going to walk through this chart. I excerpted
15 the part that I thought was relevant to these
16 devices, which is the software change. There's lots
17 of other things in this chart, but we're going to
18 start with the software change in the upper right
19 corner.

20 If there is a software change, the answer
21 is yes. Then you get down to series of questions.
22 Does the software change affect the indications for
23 use? It may be changing the organ of interest. It
24 may be changing the input to the device. But if it
25 affects the indications for use, then the answer is

1 yes, you do need to 510(k).

2 If the answer is no, the next question is,
3 Are clinical data necessary to establish safety and
4 effectiveness for the purposes of substantial
5 equivalence? And, again, this is a question that the
6 manufacturer needs to answer. You're not coming to
7 the Agency for us to answer this question. If the
8 answer is yes, if you believe the clinical data are
9 necessary, then you need to come in with a 510(k).
10 If the answer is no, you don't believe clinical data
11 are necessary, then the next question is do results
12 of design validation raise new issues of safety and
13 effectiveness? If the answer is yes, then you need
14 to come in with a new 510(k). If the answer is no to
15 all of these questions, then you just need to
16 document this in the master file for the device and
17 you do not need to come in to the Agency with a
18 510(k).

19 So just to run through these a little more
20 closely, does the change affect the indications for
21 use? As with an explicit labeling change, if the
22 change affects the indications, i.e., if it creates
23 an express or implied new indication, then a new
24 510(k) should probably submitted.

25 Are clinical data necessary for the

Free State Reporting, Inc.
1378 Cape Saint Claire Road
Annapolis, MD 21409
(410) 974-0947

1 purposes of determining substantial equivalence?
2 Whenever the manufacturer recognizes that clinical
3 data are needed because bench testing or simulations
4 are not sufficient to assess safety and effectiveness
5 and thus to establish the substantial equivalence of
6 a new design, a 510(k) should be submitted. And this
7 is excerpted from the guidance.

8 And the preceding discussion on deciding
9 when to submit a 510(k), it's applicable or can be
10 used, you know, a similar thing for a PMA. Just keep
11 in mind, it is a different regulation, and I'm really
12 not going to go into the details of the PMA
13 supplements. But, again, if anyone is interested,
14 it's in 21 C.F.R. 814.39.

15 Question 5: Input on the regulatory
16 significance of minor modifications to CADe. So I'm
17 going to come back to this. At the March 2008 Panel
18 meeting, FDA received the following comment: "We
19 (the Panel) still have to decide what minor would be.
20 That's what the FDA is going to do for us, I hope."
21 I think the FDA guidance clarifies that a minor
22 modification is a modification that does not require
23 submission of a 510(k).

24 Here's an example of a minor modification.
25 This is based on standalone testing that the

Free State Reporting, Inc.
1378 Cape Saint Claire Road
Annapolis, MD 21409
(410) 974-0947

1 manufacturer does. You know, I show the ROC curve
2 for the original CADe device. I just show one point,
3 the operating point for the updated device. It's got
4 a higher sensitivity. It's got fewer false
5 positives. I'm presuming it's outside of the
6 confidence range of that exiting curve. This is a
7 minor modification. It does not affect device
8 standalone performance from the original device, and
9 for which no labeling change is needed. If the
10 manufacturer is making a labeling change, that may be
11 a different story. If they're just making small,
12 incremental changes, they test it, they don't believe
13 it affects safety and effectiveness, they justify it
14 and they're not changing the labeling, there's no
15 need to come in to the Agency.

16 And, you know, it's the best I can do for a
17 quick example, but I'm sure there are many other
18 examples. And the manufacturers know a lot more
19 about this than I do because they do this testing all
20 the time. It just requires that you use the analysis
21 in the guidance document, you document what you're
22 doing, and you don't necessarily have to come in to
23 the Agency with 510(k). I hope that helped clarify.

24 Question 4(d). Another issue that's a
25 difficult area for us that we could use your help

Free State Reporting, Inc.
1378 Cape Saint Claire Road
Annapolis, MD 21409
(410) 974-0947

1 with, a new imaging input to a breast CADE device.
2 Is a clinical performance assessment, i.e., a reader
3 study, necessary to assess the CADE for use with the
4 new FFDM device, or is a surrogate endpoint, such as
5 standalone performance data, sufficient to
6 demonstrate comparable performance based on the
7 specifications of the device?

8 Here's an example. Suppose a CADE device
9 is approved for use with image input from FFDM 1 and
10 the CADE manufacturer seeks approval for use with a
11 new imaging input from a different FFDM device,
12 FFDM 2. This is an example actually of a PMA
13 supplement for the addition of a new imaging input.
14 And these numbers are just made up.

15 Suppose for the approved CADE input, there
16 were 600 cancer cases, the mean size, 1.5. You can
17 see the distribution of masses and calcifications,
18 distribution of density. That's a typo on the
19 normals. I didn't mean 2,000; I meant to say 200.
20 And say you have the new imaging input. In the case
21 characteristics, your database has 100 cases and the
22 mean size is different. Now, it's 2. A little
23 different error.

24 And now you have 75 percent masses and 25
25 percent calcifications, different than it was before.

1 And suppose the density distribution is different and
2 the number of normals may be different. The
3 databases here clearly are different in many
4 characteristics. And unless the database
5 characteristics are essentially the same, if you do
6 provide performance testing, say you have device
7 standalone performance on the approved device, it may
8 have shown a sensitivity of 91, it shows you the
9 confidence intervals, the false positive rates, you
10 come in with the new device input, the sensitivity
11 may be fairly close. Here I just made up the number
12 85. Give the confidence intervals. How do we
13 compare these? Unless the databases are essentially
14 the same, it's not valid to simply just compare these
15 performance estimates.

16 Here's another example. Suppose you have a
17 new imaging input, and in this case, assume that the
18 database characteristics are essentially the same.
19 And suppose the approved CADE input actually had
20 clinical performance and it showed it used ROC
21 analysis and it showed that a change in area under
22 the curve with the device minus the device was .02;
23 it was a statistically significant result.

24 Well, even if the databases are essentially
25 the same, if the device standalone performance

1 estimates differ, then the clinical performance
2 estimate of the approved CADE device would not
3 necessarily be an estimate of the clinical
4 performance of the CADE device with the new imaging
5 input. So it may be difficult for us to make that
6 assessment.

7 But there is at least one set of conditions
8 such that clinical performance testing is not
9 necessary. If, for example, you have the same reader
10 performance on FFDM 1 without the CADE device versus
11 FFDM 2 without the CADE device, meaning there's no
12 significant difference per relevant mammographic and
13 patient characteristics using the same cases and the
14 readers.

15 And if you have the same CADE standalone
16 performance testing using the same cases, the same
17 scoring methodology, the same ground-truth
18 methodology, then you wouldn't need clinical
19 performance testing. Otherwise, I believe clinical
20 performance testing is necessary, as well as device
21 labeling, to reflect actual performance on the new
22 imaging input for FFDM 2.

23 Question 5(b): Mammographic CADE devices
24 contain separate and distinct algorithms that detect
25 masses versus microcalcifications. These devices

1 really then bundle two separate and distinct
2 functionalities, and the following questions seek
3 input on whether this distinction should have
4 regulatory significance.

5 As far as bundling of CADe devices, even if
6 you had just a single algorithm that was utilized to
7 detect more than one finding when the findings are
8 clearly distinct, device safety and effectiveness may
9 be significantly different for each of these specific
10 findings. For example, for mammography CADe devices,
11 they should be separately tested and labeled for
12 detection of masses versus microcalcifications.

13 We certainly know the performance is
14 different based on the published literature. Unless
15 there's sufficient testing for architectural
16 distortion and/or focal asymmetry, the device
17 labeling should include a warning that the device is
18 not intended for and/or has not be tested for
19 detection of these mammographic findings.

20 Question 5(c): Mammography CADe devices
21 are currently labeled as second readers. Do you
22 believe that these devices are used in a second-read
23 mode by the majority of radiologists who use the
24 devices in clinical practice? If not, is this is an
25 important issue the Agency should address through a

1 regulatory means?

2 And this gets back to what Dr. Gwise had
3 labeled as the "Keep All Positives from the Unaided
4 Read" rule, or it's also been called the never change
5 your mind rule from the unaided read. Basically it
6 means you do an unaided read and the device is
7 labeled never change your mind, never call something
8 that you call positive, never change it to negative.
9 Well, effectiveness is defined in terms of clinically
10 significant results. So CADe devices must be tested,
11 labeled, and intended for use in a manner consistent
12 with clinical practice. Therefore, unless the "Keep
13 All Positives from the Unaided Read" rule, described
14 by Dr. Gwise, is consistent with clinical practice,
15 it should not be used for testing or labeling of CADe
16 devices.

17 Just in summary, the proposed CADe guidance
18 documents, remember, contain general recommendations
19 to device manufacturers on the device description,
20 standalone performance testing, and clinical
21 performance testing. Specific CADe devices
22 regulation or type of data to be provided for
23 clearance of approval requires accounting for the
24 clinical action. Device standalone testing and
25 device clinical testing are necessary to establish

1 safety and effectiveness and to properly label CAdE
2 devices.

3 Thank you.

4 DR. D'ORSI: Thank you, Dr. Smith. The
5 next speaker is Dr. Krulewitch, who's going to speak
6 about post-approval considerations.

7 DR. KRULEWITCH: Good morning.
8 I'm Cara Krulewitch. I'm an epidemiologist and Team
9 Leader with the Division of Epidemiology in the
10 Office of Surveillance and Biometrics. I would also
11 like to acknowledge Dr. Ronald Kaczmarek, who is a
12 radiologist and epidemiologist and participated in
13 the development of these slides.

14 Today I will talk about post-market
15 surveillance and the total product lifecycle; general
16 principles that apply to post-approval studies; the
17 unique case of CAdE in this process; some findings
18 from the literature; a discussion of the actual
19 conditions of use and the potential effects of the
20 actual conditions of use on the findings; clinical
21 study findings and their implications if clinical
22 studies are required; and post-approval study
23 challenges and CAdE evaluation issues.

24 After devices have been reviewed through
25 either the 510(k) or PMA process, and possibly an

Free State Reporting, Inc.
1378 Cape Saint Claire Road
Annapolis, MD 21409
(410) 974-0947

1 Advisory Panel review, they are approved or cleared
2 and then are monitored through a number of methods,
3 including post-approval studies as a condition of
4 approval for PMAs, and medical device adverse event
5 reporting. Post-market monitoring is guided by
6 several post-market authorities, medical device
7 reporting, which I'm not going to go into detail
8 about, but does allow for adverse event reporting,
9 both mandatory and voluntary, conditions of approval,
10 and post-market surveillance.

11 Post-approval studies on Class III PMA
12 devices can be ordered at either the time of approval
13 or after approval. They are under the authority of
14 C.F.R. Title 21 Section 814.82, where FDA may impose
15 post-approval requirements at the time or after
16 approval of the PMA and continuing evaluation and
17 reporting on the safety, effectiveness, and
18 reliability of the device for its intended use.

19 Additionally, in Section 522 of the Food,
20 Drug and Cosmetic Act, post-market surveillance can
21 order a Section 522 study, and this can apply to
22 either Class II or Class III devices when there is
23 failure or a reasonable likelihood of serious adverse
24 events, or it's expected that there will be
25 significant use in pediatric populations, or the

1 device is implanted for greater than one year, or if
2 it's a life-supporting, life-sustaining device used
3 outside of a user facility.

4 We have a number of post-market tools, as I
5 talked about, where we can evaluate the safety and
6 effectiveness of devices once they've gone to market.
7 In addition to adverse event reporting and the
8 mandated studies I discussed, we also conduct applied
9 epidemiologic research, and the Sentinel Initiative
10 may offer a new way for active surveillance of these
11 devices.

12 The general principles for post-approval
13 studies include that the objective is to evaluate
14 device performance and potential device-related
15 issues in a broader population of people than was
16 done in the clinical study, over extended periods of
17 time after pre-market establishment of reasonable
18 assurance of safety and effectiveness of the device.
19 Post-approval studies should not be used to evaluate
20 unresolved issues from the pre-market phase that are
21 important to the initial establishment of device
22 safety and effectiveness.

23 Post-approval studies are needed to gather
24 essential information on the longer-term performance
25 of devices, retreatments and product changes, the

1 real-world performance where the effect of the
2 patients and the clinicians is considered, the
3 effectiveness of training programs when they're
4 necessary, the performance in subgroups that were not
5 evaluated in the clinical study, and outcomes of
6 concern that may have evolved during the evaluation
7 of safety and effectiveness.

8 As noted by Dr. Gwise, the data to support
9 safety and effectiveness or determine substantial
10 equivalence in the CAdE arena are most often
11 retrospective, enriched reader studies. These
12 studies are not generalizable to the population, as
13 discussed, due to the potential biases that are
14 associated with reader studies. Additionally, these
15 studies do not assess reader variation or the impact
16 on reader actions in the condition of use, and these
17 may change performance across enriched studies. They
18 also do not account for variations due to radiologic
19 findings, pathology, or subgroups. To date, there
20 are no post-approval studies as a condition of
21 approval for CAdE devices.

22 Similar to Dr. Smith's presentation, we
23 conducted several literature reviews, which were in
24 your Executive Summary, that evaluated CAdE devices,
25 both for mammography CAD and CT colonography and

1 lung. And we did find that in many of these studies,
2 there was an increase in false positives. That may
3 lead to increases in recall and biopsy rate,
4 potential for psychological stress, cost, morbidity
5 and complications from additional diagnostic
6 procedures, potential for missed cancers, variation
7 in subgroups among both readers, persons, and the
8 cancers, reported study designs that may not always
9 be used as indicated in the labeling, and potential
10 for additional radiation exposure.

11 When considering the actual conditions of
12 use, the case mix is different than retrospective
13 reader studies. There are changes in the ratio of
14 false positive marks to the true positive marks that
15 may occur because there is a difference when
16 evaluating enriched data, since in screening
17 populations, the prevalence of the condition being
18 screened for is lower and the sensitivity and
19 specificity is not pre-specified.

20 Under actual conditions of use, there may
21 be further intervention for the patient, such as
22 recall and biopsy or the potential for missed
23 cancers. Additionally, the physician may have
24 pressure to constrain the recall rate, and there are
25 potential medico-legal concerns, and these

1 implications may alter reader performance.

2 When clinical studies are included in the
3 application packet, there may be questions that arise
4 out of these clinical studies, in particular, device
5 performance in subgroups, both the patient subgroups
6 and the user subgroups, less experienced readers
7 versus more experienced readers, pathology subgroups,
8 and the radiologic findings that are noted, such as
9 architectural distortions, microcalcifications, et
10 cetera. The initial pre-market study may not be
11 powered for device performance in these subgroups,
12 and this may lead to equivocal subgroup results.

13 Additionally, there are challenges in the
14 post-approval study to conduct studies. These
15 include recruitment, clinical sites, physicians, and
16 patients. Because we cannot compel subjects or
17 physicians to participate, and because large
18 studies -- large sample sizes may be required, these
19 studies may also take significant amounts of time.

20 Additionally, because there is a low
21 prevalence of a disease of interest in the real-world
22 setting, the sample size may be very large, and this
23 may increase the burden for the sponsor and may
24 require a larger effort for the conduct of the post-
25 approval study compared to the clinical study in the

1 pre-market setting.

2 There is a potential for a wider variety of
3 endpoints, and these include recall rate, biopsy
4 rate, cancer detection stage, and mortality, and we
5 will be asking the Panel to provide input on the key
6 endpoints to be studied in post-approval studies.

7 As Dr. Smith just discussed, when
8 evaluating CADe devices, there are differences in
9 risks based on the findings presented because of the
10 clinical action that may follow. This risk asymmetry
11 results in a potential harm from one additional false
12 positive not being equal to one missed cancer among
13 all devices. And we will be requesting Panel input
14 regarding the balancing of the risks of increased
15 false positives against the benefit of increased
16 sensitivity and risk for all CADe devices.

17 In summary, CADe performance under actual
18 conditions of use may be different than what is found
19 in retrospective reader studies. Post-approval
20 studies for CADe devices can address unanswered
21 questions for subgroups, device performance under
22 actual use, and provide a greater range of study
23 endpoints.

24 Thank you.

25 DR. D'ORSI: Thank you very much. Since we

Free State Reporting, Inc.
1378 Cape Saint Claire Road
Annapolis, MD 21409
(410) 974-0947

1 have some time and we have a pretty full agenda, I'd
2 like to make a slight variation in the order of the
3 agenda. What we'll do now is open -- excuse me.
4 Yes. So as I was saying, what I'd like to do now is
5 open our Panel for questions to this morning's
6 speakers, then we'll adjourn for lunch and come back,
7 hear the afternoon speakers and reopen the Panel
8 questions to this morning's speakers, if any other
9 questions are needed, and this afternoon's speakers.
10 So let me open the discussion with the Panel. And,
11 again, if there is a specific response to a question,
12 please come up to the podium and state your name.

13 So I'm going to open that up now to the
14 Panel. Any questions that we have from this
15 morning's speakers? Yes.

16 DR. DODD: I have a question about the
17 definition of CADe versus CADx. So in Dr. Whang's
18 presentation, it was said that CADe is defined to
19 identify, mark, or highlight portions of an image.
20 Is that just a binary task? Say there's a
21 probability of malignancy, does that become CADx?
22 Could somebody clarify that?

23 DR. PETRICK: Could you go and ask the end
24 of your question one more time?

25 DR. D'ORSI: Please state your name.

1 DR. PETRICK: This is Nick Petrick from the
2 FDA.

3 DR. DODD: Okay. So Nick, if I -- so I go
4 in and mark an area of interest on an image, I
5 understand that to be part of CADe. But if I then
6 associate some probability of malignancy or some
7 probability of this being an interesting finding,
8 does that fall over into CADx?

9 DR. PETRICK: That would be CADx. It would
10 be some sort of combination, probably, of both
11 devices, if that probability score is put as an
12 output to the reader. If that's an internal -- if
13 you have a classifier that differentiates regions and
14 that's an internal function and all that's put out is
15 that prompt or that marking to the device, that would
16 be the CADe functionality. If you then add to that
17 this probability of malignancy score, then that would
18 be a combination of CADe and CADx. That would be
19 outside the scope of the particular guidances.

20 DR. DODD: Okay. So if I'm doing a
21 standalone performance and I'm just looking at the
22 presence or absence, I can't do an ROC analysis
23 unless I have some probability of malignancy or some
24 continuous output?

25 DR. PETRICK: We would look at that

Free State Reporting, Inc.
1378 Cape Saint Claire Road
Annapolis, MD 21409
(410) 974-0947

1 standalone performance output. So it could be an
2 FROC or an ROC type of analysis. So as far as
3 standalone performance, you could look at the overall
4 curve and as long as -- it's really this definition
5 of what information is going to be provided to the
6 actual reader. And we would certainly like to see
7 the FROC curves for devices and not just one single
8 operating point in the standalone data. And it's
9 really this question, that the clinician will only
10 see either prompt or if they're going to see some
11 additional information about the diagnostic
12 information, that would move it to a different
13 category, and it'd be outside the scope.

14 DR. DODD: Okay, but the standalone
15 performance would --

16 DR. PETRICK: Sure, you can still do that
17 part.

18 DR. DODD: -- not divide that threshold?

19 DR. PETRICK: That's right.

20 DR. DODD: Okay, thank you.

21 DR. D'ORSI: Dr. Zhou.

22 DR. ZHOU: So that's actually my follow-up
23 questions. If I understand it correctly, there's
24 three different types of the design, the standalone,
25 reader performance, the clinical performance, and

1 also we have to show an ROC curve. So could somebody
2 clarify what's the differences among all three of
3 them, in terms of the data you have to generate the
4 ROC curve? I mean, in order to generate an ROC
5 curve, I think you have to have some interpretations
6 on the results. So that's how you can distinguish
7 those three types.

8 DR. PETRICK: So I guess I'll try to
9 clarify it a little bit. Within the company, when
10 they have their own device, they will have outputs
11 from a classifier. That allows them to do either ROC
12 or FROC or some type of performance measurement on
13 standalone performance. When you go to evaluating
14 the reader performance, this is a reader using the
15 CAD device, the output is going to be a prompt for
16 that reader, and now the reader would either provide
17 some sort of binary decision if it's a binary
18 endpoint in that reader study, or they would provide
19 some sort of ROC score or they would do multiple
20 comparisons to provide ROC data to allow a comparison
21 on that controlled reader study.

22 So I think what Dr. Dodd was talking about
23 is in the standalone data, and there are sort of two
24 different pieces of information. The standalone
25 data, we can -- the company has the ability to look

1 at ROC or FROC types of data there, in most cases.
2 And then when they actually do the reader study,
3 that's a different study. The output from the device
4 is not just a prompt, and the reader would either
5 provide the scores or the binary outputs.

6 DR. ZHOU: But could you stay there?

7 DR. PETRICK: Sure.

8 DR. ZHOU: So if you have standalone
9 studies, then how can you give the ROC type of data,
10 which the data would have to say what is the presence
11 and what's your confidence level on the presence or
12 absence of lesions?

13 DR. PETRICK: So this is Nick Petrick again
14 from FDA. The classifier would do that. Typically,
15 there's an output from a classifier that tries to
16 differentiate regions, and the output of that
17 classifier typically, it doesn't have to be, but
18 typically it's going to have some sort of numerical
19 value associated with it.

20 DR. ZHOU: The region here referred to the
21 tumor type, or is it presence or absence of tumor or
22 the lesion?

23 DR. PETRICK: So we're talking about two, I
24 think, two different parts. One is the mechanics of
25 the CAD algorithm, which is trying to put marks on

1 the image. It's trying to identify regions. It may
2 be segmenting regions. It may just be looking at
3 boxes or regions of interest within the image. That
4 algorithm in itself will do some sort of
5 classification. It will try to differentiate normal
6 from benign. In the definition of a CAD device, it's
7 that marking of a region on the film for the
8 clinician to use that defines it as a CADe device.
9 That's just the prompt output.

10 DR. ZHOU: Yeah. So maybe that's actually
11 the other issue. When you talk about film, are those
12 films from the diseased patients? Where do those
13 films come from?

14 DR. PETRICK: Well, they're from the
15 population of patients that are going to be evaluated
16 with the device. So they're clinical examinations of
17 the patient. And some will be diseased, obviously,
18 and others will be normal.

19 DR. ZHOU: Yeah. So I still don't get it.
20 So basically you're using the same kind of the
21 patient population that's clinical assessment or the
22 reader assessment for standing alone assessment,
23 right?

24 DR. PETRICK: Right. I mean, typically
25 they'll come from a population of real clinical

1 cases, where someone went back and truthed them and
2 marked the locations and then the device is now
3 tested. There needs to be a scoring algorithm
4 associated with it. There needs to be truthing
5 associated with it. And it's really a clinical --
6 the images used for standalone are clinical images.
7 It just doesn't have the next step where the reader
8 actually interprets the images with the CAD.

9 DR. ZHOU: So let's say you have -- the
10 image comes from the colon cancer patients. The
11 standalone machine will tell me, based on the
12 score, to say what's the presence and absences of the
13 colon cancer on that image?

14 DR. PETRICK: Well, in colon CADs, they're
15 typically looking for polyps.

16 DR. ZHOU: Yeah.

17 DR. PETRICK: So on that particular image,
18 it'll mark regions. Typically they'll have -- you
19 know, hopefully it marks regions that has true
20 polyps, and it's also likely to mark regions which
21 are normal tissue or some other type of tissue that
22 are not actual polyps. So there'll be true marks,
23 there'll be prompts on true polyps, and there'll be
24 some other marks potentially on normal tissues or
25 other regions that are false positives.

1 DR. ZHOU: Okay.

2 DR. PETRICK: That'll be the output of the
3 CAD.

4 DR. ZHOU: So you're making some sort of
5 diagnosis already?

6 DR. PETRICK: Well, the CAD algorithm is
7 making a determination of what it should mark and
8 what it shouldn't mark. When it prompts on the
9 output, it just puts a mark on the film. It says,
10 look at this region, and have the clinician say --
11 the clinician is supposed to look at that region and
12 then reevaluate and determine whether there's
13 actually a polyp there or not. It doesn't provide
14 the diagnostic information directly to the clinician,
15 but the fact that it's prompting particular
16 locations, you know, it has some ability to
17 differentiate regions. I mean, that's the basic
18 premise of how a CAD would work.

19 DR. D'ORSI: Yeah. Can I get a little
20 clarification on the -- I think it'll help all of us
21 on ROC and CAD. If the device is throwing up a set
22 of binary decisions for you to make, let's say four,
23 and you make a decision on one that you recall, where
24 are you getting the gradations from the reader to put
25 into an ROC curve? I think that's what you're

1 getting at. Unless you ask the reader what is your
2 confidence that this is real, from absolutely not
3 real to absolutely real, which you don't ask. So
4 that's what I'm -- I think that's where we're having
5 trouble.

6 DR. PETRICK: Okay. So you could do the
7 study where, if you're doing the controlled -- if
8 you're doing the reader study part to evaluate the
9 CAD and all it's doing is exactly what you said, it
10 puts, say, four marks on the image, then you would
11 ask the reader potentially to rate that, give some
12 scaling from 1 to 10 or some other number, about your
13 confidence that this is a polyp or that this is a
14 cancer or whatever it may be that you're evaluating.
15 What I tried to lay out in my presentation is you can
16 actually do this in multiple ways. You can make a
17 whole bunch of binary decisions and then have the
18 reader actually reevaluate subsets of cases, and that
19 would also allow you to get an ROC curve, or you can
20 actually look at comparisons. You can look at case
21 one and compare it to case two and determine which
22 one you think is more suspicious than the other.

23 And, again, that would be another approach
24 to getting ROC type of data to use in the analysis.
25 So there are multiple ways to do it. Typically it is

1 this idea of scoring. That is the reader study part
2 of this. The CAD has its own algorithm that does
3 something similar, but it's really -- it's the
4 algorithm itself. It's not associated with the
5 reader at that point.

6 DR. D'ORSI: Thank you. That's very
7 helpful. Any other comments, questions? Dr. Kim.

8 DR. KIM: I mean, I might have a wrong
9 understanding, but in terms of the standalone, I
10 think the difference between that and the reader
11 performance is that the CAD actually does create a
12 likelihood ratio, and at some point there is a level
13 that you say this is real, this is not. So you can
14 take that ordering from the CAD to create your ROC
15 curve. And that's how you can get the curve from the
16 computer itself.

17 Also I had a quick question. I understand
18 that if you give a likelihood score to the reader
19 that would have CADx functions, would that be
20 considered the same if you just ordered the polyps,
21 the polyp markings without giving a likelihood score?
22 So going from most likely polyp to least likely but
23 not giving a score.

24 DR. PETRICK: That would again, I think,
25 fall outside the scope of this particular guidance.

1 We would call that at least having some CADx
2 functionality associated with it as well.

3 DR. D'ORSI: Lenny.

4 DR. GLASSMAN: Nick, stay there. I think
5 you're probably the right person. Let me ask a
6 question because it's got a long preamble. The
7 question is for 510(k) CAD submissions, a new unit or
8 a significant modification based on a predicate. Why
9 is clinical testing needed at all? Now, let me go
10 back to the beginning of the question, which is we've
11 got safety and efficacy.

12 Quite frankly, to me safety is a non-issue
13 here because it is the reader, the radiologist who is
14 the safety function, not the CAD device. The CAD
15 device is a marker. That's all it is. Now, on
16 efficacy, if we've already -- if we're already
17 satisfied that the predicate device is safe and
18 efficacious, if we could compare the new device or
19 the significant modification to either the original
20 predicate or a new version of the predicate that we
21 can reasonably assume is at least as safe and
22 efficient as the original predicate, and we got a
23 similar number and similar location of true positives
24 and false positives, can't we make the assumption
25 that that device is equivalent or at least equivalent

1 because we're using the same new test set? So the
2 issue of different sized lesions and everything goes
3 away. And under those circumstances, why would we
4 need a clinical reader study?

5 DR. PETRICK: I mean, I'll speak for
6 myself. I agree with you. If we can show that you
7 get basically the same location and either have
8 equivalent or better performance, that I don't -- I
9 agree, I think that that has implication for clinical
10 practice and what the CAD might be able to do. It
11 doesn't directly measure it, which I think is some
12 other people's problem with it, is we don't have a
13 direct measure for it. But I do agree that in that
14 case, the potential for having a significant impact
15 on clinical change seems to be fairly modest in my
16 opinion. And I think Dr. Gwise can say as well.

17 DR. GWISE: My name's Thomas Gwise. I'd
18 just like to add a little bit to what Dr. Petrick
19 said. If the device is being changed, we need to
20 look at the change, that does the change affect the
21 change in the reader's diagnosis? How does a change
22 in the number of prompts affect a reader's
23 interpretation of the image? You think about the
24 paper by Egglin, and if you change the frequency, the
25 number of CAD marks, you could get a difference in

1 the reader performance.

2 So if you went from, say, having two to
3 four false positives per image and that's changed now
4 to very few in comparison, how would that change the
5 reading, the interpretation? Now, where do you draw
6 the line? How do you differentiate a big change in
7 the frequency of marks to a small change? And what
8 effect does that have on reader performance overall?
9 And does that change the safety and effectiveness?

10 DR. D'ORSI: Yeah, Rob.

11 DR. ROSENBERG: Yeah.

12 DR. SMITH: This is Robert Smith. I just
13 wanted to add one comment to that, to Dr. Glassman.
14 Maybe I could help you with an example that was
15 actually in my slide, the example that I gave, where
16 you wouldn't need a clinical performance test, and if
17 you use the same dataset, the same scoring
18 methodology, the same ground truth, et cetera. And
19 the problem is if -- and the marks have to be in the
20 same locations, the same number. Even subtle things
21 like the size, the shape, using a dashed line for a
22 prompt, using a solid line, even those subtle things
23 can affect the impact on the user. So unless all of
24 those things are the same, there's no way to know how
25 it will affect the user.

1 And even just decreasing the number of
2 false positives, you don't know if those false
3 positives are in the same location. A new device may
4 mark fewer false positives, but the things that it
5 marks may be more difficult for the reader to dismiss
6 as a false positive. So it's a little more
7 complicated than what you're suggesting.

8 DR. GLASSMAN: Well, can I have one quick
9 follow-up?

10 DR. D'ORSI: Let's do it after Rob, okay?
11 Just keep it in your head. Rob.

12 DR. ROSENBERG: Yeah, I think you may have
13 answered the question, but do we know whether
14 decreasing the false positives has a known effect on
15 reader performance versus increasing the false
16 positives? In other words, if we know the direction
17 of the change in the algorithm, does that allow us to
18 require or not require reader performance testing?

19 DR. SMITH: Well, if all other things are
20 equal, and even the locations, and you're just
21 decreasing the number of false positives, but the
22 ones that are left are exactly the same as they
23 would've otherwise been, under that scenario and
24 everything else is exactly the same, you probably
25 wouldn't need to do any other testing.

1 DR. D'ORSI: Lenny.

2 DR. GLASSMAN: Just one more question.

3 Would it be reasonably likely, however, using the
4 reasonably word from the FDA regulation, though, that
5 if the differences were minor, that the clinical
6 impact would be minor and you could do away with the
7 reader study? Or is that just too unknown?

8 DR. SMITH: It's unknown to me. I think,
9 again, the minor comes down to -- well, the
10 definition I like of a minor modification is that's
11 when the manufacturer doesn't come in to us to take a
12 look at it. They're making small changes. They
13 should know why they're making the changes. They
14 should be testing what the effect is, and they
15 shouldn't be coming in to the Agency under that kind
16 of circumstance.

17 DR. JIANG: Can I just ask --

18 DR. D'ORSI: Yeah, Dr. Jiang.

19 DR. JIANG: Yeah, I'd like to follow up.

20 DR. PETRICK: Can I just comment on that?

21 DR. D'ORSI: Just state your name.

22 DR. PETRICK: This is Nick Petrick from the
23 FDA. I mean, I think I agree with you. I think that
24 there doesn't seem to be a large risk that there
25 would be major changes to clinical performance. And

1 the idea that the manufacturer should come in, I
2 mean, that's actually something new that we haven't
3 discussed internally at the FDA. It's news to me
4 from today. So that may be a viable option, but
5 that's something that we haven't really had time to
6 discuss within the FDA.

7 DR. D'ORSI: Dr. Jiang.

8 DR. JIANG: So this is going to be a
9 follow-up to Dr. Smith's example. So I think if you
10 have the exact same computer algorithm, the same
11 marks, everything the same, you repeat the same
12 reader study, the readers are not going to give you
13 the exact same answers because there's variability of
14 the readers. So given that, the question that I have
15 is, is there a particular example you can think of
16 that you would suspect the reader performance would
17 differ, given the same computer performance, given
18 the same standalone performance that you wouldn't --
19 you know, the standalone performance is okay.

20 DR. SMITH: You're saying again under the
21 conditions where you're using the same cases, the
22 same scoring methodology, the same ground truth, et
23 cetera, and the marks are the same locations, the
24 same size, the same prompt type, I think I've
25 answered that as no, I wouldn't -- you're right,

1 there'd be variability to different readers using it.
2 but if you tested it and you did a multi-reader,
3 multi-case study, you'd expect to have the same
4 outcome. Obviously, the question would be, well, why
5 would a manufacturer be coming in to the Agency, you
6 know, under that kind of scenario?

7 It could just be for a labeling change.
8 They may have made a minor modification that maybe
9 wouldn't require coming in to the Agency, but they
10 want to change the labeling to claim that they've
11 decreased the false positives or whatever the case
12 may be. Well, to get that labeling change, you
13 probably do need to come in to the Agency. It could
14 just be for that purpose. So it also depends on the
15 reason that the manufacturer is coming in to the
16 Agency.

17 DR. D'ORSI: Yeah, Dr. Dodd.

18 DR. DODD: So I just want to address this
19 to Dr. Petrick. So if in a standalone performance
20 study you have an increase in sensitivity but you
21 begin to have a pretty big increase in the false
22 positive marking rate, you know, at some point it
23 seems to me that you would want to do a clinical
24 study. So how would you recommend drawing that line?
25 How much of an increase?

1 DR. PETRICK: So this is Dr. Petrick again.
2 That's a tough question. I don't have a good answer
3 for that, and I guess I couched my slides in the
4 sense that we would expect to see increasing
5 sensitivity or at least stable sensitivity for
6 decreases in false positive, or stable false
7 positives for increases in sensitivity. And, again,
8 I think there is risk associated with changes to an
9 algorithm, where you have a subgroup that gets a
10 large improvement in performance, but that has a
11 negative impact on something else.

12 And I just don't have a clear
13 differentiation, but I do have, at least in my
14 opinion, again, that if the marks are fairly
15 consistent with each other, not necessarily exactly
16 the same but consistent with each other as far as
17 false positives, you've increased sensitivity,
18 especially in the scenario where we have a very large
19 number of false positive marks for the clinician to
20 read relative to true marks, which is typical in
21 screening CAD devices, the risk that there's going to
22 be a large impact on clinical practice is fairly
23 minimal.

24 DR. D'ORSI: Can we get a little more
25 guidance on the risk/benefit? And this is open to

1 anybody. I think we'll need some guidance on this
2 question related to clinical testing. The risks for
3 CAD-mam, CAD-colon are basically recall imaging.
4 However, the risk for CAD lung is intervention, if
5 I'm understanding this correct, and if I'm not,
6 whoever is knowledgeable on that can weigh in. Yes.

7 DR. STEIER: Yeah, I was going to even ask
8 about that because on Slide Number 260 on Page 130,
9 where it talks about CT scan of the lung, it says the
10 next step with a positive CAD would be -- or a CAT
11 scan would be biopsy, but it might really be PET
12 scan. So I don't know if that really matters, but it
13 does certainly to the patient, and certainly a PET
14 scan is going to have a much different risk profile
15 than a biopsy. So, in clinical practice, as a
16 pulmonologist, it would be unusual to proceed to a
17 biopsy without getting a PET scan, at least in the
18 area I practice in.

19 DR. D'ORSI: Okay, that's valid. That's
20 all I wanted to know. So the risks and benefits are
21 about the same in all these CAD devices. Good, thank
22 you. Yes, Dr. Zhou.

23 DR. ZHOU: Maybe this is for Dr. Petrick.
24 I want to go back to the standalone. So if you're
25 able to actually report the accuracy from standalone,

1 that's actually very important information because
2 that represents the accuracy due to machine himself.
3 If that one is 100 percent, who needs a radiologist?
4 So that's actually more important information, I
5 feel.

6 DR. PETRICK: I mean, I agree. If you have
7 100 percent accuracy, you have no false positives and
8 always get the lesion, we would be talking about
9 computer diagnosis and not computer-aided diagnosis
10 in that scenario. And so in this particular case,
11 we're really talking about these devices. It's not
12 devices that are going to work on their own. These
13 are devices that will always be interacting with the
14 clinician. That's the type of devices we're talking
15 about today.

16 DR. ZHOU: So that actually raises the
17 issue about what's the value of ROC curve estimated
18 for standalone if you don't believe that, if you
19 don't believe that would give you useful clinical --
20 useful information.

21 DR. PETRICK: I think it does give you
22 useful information. At some point, though, there has
23 to be cut put on that ROC curve and standalone. So
24 what comes out isn't an ROC output, but it is an
25 actual I've cut it, I'm going to get four marks per

1 image and that's going to lead to a sensitivity of
2 whatever it is, 80 percent, 90 percent, 70 percent.
3 So the output of the CAD will always have cut point.
4 The utility of looking at the FROC or the ROC curve
5 is you can understand what the tradeoffs are in cut
6 points that the company potentially made. It may be
7 that they picked a point that makes a lot of sense.
8 They may not have made sense. And, likewise,
9 companies sometimes want to adjust that cut point
10 after -- on a new revision of the software. And so
11 instead of saying now you wanted two false positives
12 per image, I'm going to move down back to 1.5. And
13 it again tells you the tradeoffs that you're looking
14 at in standalone performance only.

15 DR. ZHOU: Yeah, I feel like that for
16 standalone, you need to use some different
17 measurement in the ROC curve, even for the FROC or
18 the ROC, to evaluate the standalone performance. So
19 they're different from the clinical performance or
20 reader performance.

21 DR. PETRICK: Yeah, I guess would -- I
22 think we do look at cut points, and we do look at
23 what's typical as FROC, and I guess, in my opinion,
24 those are the appropriate measures. We're looking at
25 this numerical output coming out from a classifier.

1 ROC is certainly developed based on the fact that you
2 have this numerical output from a classifier
3 developed for radar applications, where you really
4 have this tradeoff between sensitivity and
5 specificity and you're interested in understanding
6 what that whole tradeoff is.

7 DR. D'ORSI: How about this side of the
8 table? You've been pretty quiet. Yes.

9 DR. BOURLAND: This may be for Dr. Whang.
10 I have a question about this is all for digital
11 images, but some of these images may come from analog
12 format, radiography and mammography. And can someone
13 just give a brief status on what number of these
14 input images are then analog converted to digital? I
15 assume that's changing over time. And then the
16 question is relevant for database, basically,
17 testing.

18 DR. PETRICK: Yeah, I don't know if someone
19 else -- this is Nick Petrick again -- if someone else
20 has better numbers. I'm not actually sure what the
21 relevant numbers are, but most of the CAD algorithms
22 that were approved for mammography screen-film
23 systems have been updated to include at least some
24 portion of digital mammography systems.

25 Obviously, for CT, those have been approved

1 for, you know, digital format. And X-ray, I actually
2 don't know. There's a chest X-ray device which was
3 approved for screen-film X-ray. Is that correct? Or
4 was that digital? And so I'm actually not sure of
5 the status of that. Someone else may have more
6 information.

7 DR. BOURLAND: Okay. And then the follow-
8 up question for this is, relative to Category II and
9 Category III devices, all mammo CADe goes -- are at
10 Class III, I think is what I understood, plus one
11 lung, and the others are all Class II. So what's the
12 distinction?

13 DR. SMITH: The distinction could just be
14 the indication that the manufacturer seeks. For
15 example, for the chest devices, there's a
16 manufacturer who wanted to have on their labeling
17 that it could detect cancers, specifically, as
18 opposed to detecting nodules. That would make it or
19 had made it a Class III device. Even though it could
20 be exactly the same algorithm, exactly the same
21 device, just what it wants to be indicated for could
22 make it a different device, a different
23 classification.

24 DR. BOURLAND: Okay. And then maybe this
25 is the question going back to a little bit of CADe

1 versus CADx. So in the case of some device that
2 would detect cancers, is that a CADx or is that a
3 CAdE?

4 DR. SMITH: That's a good question. I'm
5 not sure there's necessarily a uniform opinion inside
6 the Agency. I kind of view all of these CAD
7 detection devices as having some CADx functionality
8 because they're not just trying to detect imaging
9 findings, they're trying to detect something of
10 significance, which, you know, is typically going to
11 be cancer. I mean, even a device in the colon trying
12 to find polyps, you know, you're trying to find
13 polyps that may be early cancers. But the way we're
14 try to draw the line is if it's just trying to draw
15 your attention to a region of the image, you know,
16 that's what's seemingly making it a CAdE device.

17 But you're right, you could say they all
18 have some CADx functionality to them because they're
19 all trying to detect, you know, often, cancer. The
20 breast devices obviously are -- even though the
21 labeling may say it's detecting a region of interest,
22 what's a region of interest on a mammogram other than
23 a finding that may be a cancer? I don't know what
24 else is of interest.

25 DR. D'ORSI: Let me just ask this question

Free State Reporting, Inc.
1378 Cape Saint Claire Road
Annapolis, MD 21409
(410) 974-0947

1 as guidance for how to set up the clinical study, as
2 necessary. If the device, which I think is just a
3 detection device for findings and I really think, and
4 others can weigh in, that that's all it is. It is
5 not separating benign from malignant, as people have
6 said, because that's a CADx device. Would it be
7 easier to set the endpoint as a finding versus a
8 malignancy? What this may do is open up a much
9 larger number of women to be included; two, you would
10 obviously, hopefully, include cancers in those
11 finding groups. Is that something that is amenable
12 in a case study, or is the Agency strictly focused on
13 malignant versus non-malignant?

14 DR. PETRICK: This is Nick Petrick again.
15 And we are not completely focused on malignant and
16 non-malignant. And, in fact, the lung CT CADs were
17 approved for finding nodules and actually a certain
18 type of nodule, where the truth -- it becomes now a
19 question of how do you truth that. And in the case
20 of lung CT, there was a panel of experts that looked
21 at the cases and decided what was truth and what
22 wasn't. And, in fact, we actually looked at
23 variability of that truthing panel as well.

24 And likewise for polyp detection. We're
25 looking at typically detecting polyps and not

1 necessarily detecting colon cancer. So the unit is
2 really based to the manufacturer to decide what is
3 appropriate, and then based on that particular
4 decision, then the question of how do you truth it,
5 what is the appropriate method to determine what it
6 is and how to go forward with evaluating.

7 DR. D'ORSI: But the FDA would be open to
8 something like that if truthing was solid? Okay.

9 DR. PETRICK: Yes.

10 MS. MORRIS: Can I just follow up on that?

11 DR. D'ORSI: Yes, Janine.

12 MS. MORRIS: The review of any medical
13 devices and the level of evidence that's needed is
14 always going to be dominated by how the company comes
15 in and labels it for the intended use, indications
16 for use, and all of their claims that they may want
17 to make. So that's where we start. We start with
18 what the company is trying to either label the device
19 for and that's what will set the bar for us. Now,
20 certainly there are going to be arguments of, you
21 know, something that is too subtle, when it's very
22 obvious it's for something and we would work that
23 out. But that's the starting point.

24 DR. D'ORSI: Thank you. I think this is
25 important just to pursue a little bit further with

1 your guidance. If the manufacturers of these CAD
2 devices said we are marketing this for the detection
3 of actionable findings, would that be satisfactory
4 as -- this becomes, I think, I don't know if others
5 agree, this becomes relatively important both to the
6 intention of a CAD device and in clinical case
7 studies, to numbers.

8 MS. MORRIS: I think that you're going to
9 have various opinions, but again, that type of
10 phrasing is, you know, moving away from how much
11 you're relying on the device for the diagnosis. And
12 if there's clinical utility for that, and that we can
13 have established safety and effectiveness based on
14 that, then it's a reasonable approach. But you're
15 still always going to have varying points of view of,
16 well, what does that mean?

17 DR. D'ORSI: Thank you.

18 DR. SMITH: This is Robert Smith. If I
19 could just comment on that. I mean, the first
20 question we would ask in reviewing that is, well,
21 what do you mean by actionable? I would want to get
22 at exactly what that meant, and we might ask that the
23 labeling be changed to reflect what it meant. And
24 you'll also have to remember there is the truthful
25 and accuracy statement.

1 You know, the information provided to the
2 Agency has to be truthful and it has to be accurate.
3 The accuracy, you know, there might be required. It
4 might be required that it be more specific in the
5 labeling because I don't know what actionable means.
6 It depends on the device, the organ of interest. You
7 know, it would have to be defined.

8 DR. D'ORSI: Something you'd take an
9 additional action on. I'm going to call this woman
10 back. Yes. Yes, Dr. Leitch.

11 DR. LEITCH: This is Marilyn Leitch. So I
12 do think this is just very confusing because what do
13 the clinicians think they're using it for? Okay.
14 And the clinicians, I think, think they're using it
15 to avoid missing a cancer. That's what they think
16 they're using it for. So, in that sense, they don't
17 care about all of the markings that are, you know,
18 fibroadenoma or something like. They don't care
19 about that. You know, what they want to know is keep
20 me from missing a cancer. I think that's why
21 clinicians use it.

22 And so to say it doesn't have any -- you
23 know, what we're talking about doesn't have
24 diagnostic implications, it does to how the clinician
25 applies it, I think. And if they see a bunch

1 markings that are irrelevant to that, I think it
2 impacts how they view the value of the device for
3 their practice.

4 DR. D'ORSI: Any other comments, questions?
5 Yes.

6 DR. STEIER: Yeah, again, as a practicing
7 pulmonologist, I'd be petrified, probably, to see a
8 CAD report showing lots of findings and possibly
9 actionable items. We're frequently in a bind with
10 patients, you know, 85, 90-year-old patients with all
11 kinds of findings on MRI and CAT scan, as it is,
12 without the computer-aided component. And what do
13 you do? And what difference does it make to the
14 patient? And is the treatment going to be worse than
15 the cure, you know, than the disease and those kinds
16 of things? I mean, you're talking about, in CAT
17 scan, pneumonectomies and big procedures.

18 And, you know, then when you find all of
19 these -- take out of all these nodules which turn out
20 not to be cancer, that are scar tissue and granulomas
21 and all kinds of other things, it really -- and now
22 the patient's left missing a lobe of their lung and
23 six insignificant granulomas. So the whole thought
24 of a lot of actionable items, you know, we find
25 enough of those without the computer-aided diagnosis.

1 With the computer-aided, it's going to -- to me it
2 would create a whole scenario of medico-legal and
3 ethical problems.

4 DR. D'ORSI: Yeah, I understand. I won't
5 pursue that. Any other comments? Yes, Dr. Leitch.

6 DR. LEITCH: Maybe the FDA could say why
7 these were separated out into the categories because,
8 you know, if you think about, you know, the use for
9 some of the problems we're talking about, you know,
10 if you wanted to get to the point of, well, we're
11 going to screen studies to see when the physician
12 needs to look it, then having a lot of markings is
13 fine because you do want to pick up anything that
14 might require a physician to look at it, as opposed
15 to the many studies that really are okay and don't
16 require the physician to look at it. But to me,
17 parsing out these three groups and acting like
18 they're very separate, when in the way the clinician
19 is using it, it is really blurred among that. I was
20 just wondering how that was arrived at.

21 DR. PETRICK: So this is Nick Petrick. And
22 I'm, I guess, not completely clear what we're talking
23 about. What are the three different groups that
24 you're referring to? I'm not even sure --

25 DR. LEITCH: Using a CAD device just to

1 pick up anything, which is what we're talking about
2 today, using a CAD device to establish a diagnosis,
3 and using a CAD device to screen studies so that you
4 decide which study a radiologist needs to look at.
5 That was on our thing, three categories.

6 DR. PETRICK: Okay, yeah, that's what I
7 understand what the categories are.

8 DR. LEITCH: Yes.

9 DR. PETRICK: And I'll try to speak. Maybe
10 other people can speak to this as well. I think,
11 again, we're trying to look at the difference in risk
12 associated with it, and maybe what you're saying is
13 you don't see a difference in the risk associated
14 with any of those devices. What we have seen is a
15 number of devices that come in in this initial
16 category, which is prompting devices.

17 We have devices in the Agency that have
18 come in for Pap smears that actually triage
19 outpatients. It says the pathologist doesn't need to
20 look at these. These are normal women. They never
21 see the case. We look at that as a different risk
22 category, where the device is doing a different test
23 which is eliminating part of clinical practice for at
24 least some subset of patients. And so we may
25 approach it in a somewhat similar manner, but it also

1 may be different in how to approach regulating that
2 device.

3 And likewise with the diagnostic
4 information. Again, you're trying to give some
5 additional information to give further influence to
6 the clinician's decision, and that, again, may have
7 some impact on the risk associated with that device.
8 If you score a lesion and you know, you score it
9 correctly, that might be very helpful. If for
10 whatever reason you score it wrong, that may again
11 lead to the clinician making a different decision if
12 they just saw the marking. And I think that that's
13 sort of the reason that we're looking at it. We're
14 trying to, I think, start at the lowest level of what
15 a device might do and hopefully work up from there on
16 how to regulate them.

17 DR. D'ORSI: Any other questions or
18 comments? Yes.

19 DR. CARRINO: The question I have is, since
20 we're focusing on CAde today, does that mean we're
21 going to be back for the other two at some other
22 point for the methodological issues? Or should we be
23 thinking a little bit more broader?

24 MS. MORRIS: Well, actually one of the
25 questions that we have is kind of future

1 considerations at the classification of, for this
2 point in time, the CADe devices. And so as I'm
3 hearing you, I'm just kind of speaking off the top of
4 my head.

5 But if there is belief by the Panel that
6 there really is not distinction between CADe and CADx
7 and that they need -- that the risk associated with
8 it is high enough that we need to consider the
9 classification of these devices, not in a separate
10 bucket but together, that's important information
11 that we can take back and then consider about the
12 classification of these devices and whether or not
13 they should be in the same classification, whether
14 that be that they go all to Class II or whether they
15 all go to Class III. And that's all based on the
16 risk/benefit. So that's more of a future
17 consideration, but we would certainly like to hear
18 comments, and that's why we added that question.

19 DR. D'ORSI: I'd like everybody to weigh in
20 on this. At least in mammography, and that's the one
21 I have the most experience in, it's a binary type of
22 exam. One is screening where you're throwing that
23 out and you pull in possible sources of malignancy,
24 possible. A Dx needs a much larger database to now
25 analyze the stuff that the net brought in. So it's

1 not really the same function and the same test and
2 the same algorithm and the same percentages it's
3 going to give you. One set of percentages is, is
4 this a focal, three-dimensional finding?

5 The second one says, okay, that's a focal,
6 three-dimensional finding. Now, what's the chance
7 that this focal, three-dimensional finding is benign
8 or malignant? So that, at least in mammo, is the way
9 it works. And for mammo, I think that's an excellent
10 separation. I can't speak for CAD colonoscopy or
11 lung, and maybe we can hear from everybody on that.

12 DR. KIM: Well, for colonography, the issue
13 is not so much trying to find the histology but to
14 find a polyp, and basically it's just a morphologic
15 projection into the colon, and it could represent
16 either adenomatous or non-adenomatous or even early
17 cancer. And so the purpose, I think, the helpful
18 purpose of CAD and CT colonography is to point out
19 areas and add redundancy to your search.

20 Because of what we know with CTC
21 interpretation, one of the reasons why we do so well
22 is that we interrogate the dataset from different
23 viewpoints, looking at it from 2-D, 3-D, and you go
24 through the same datasets several times. And the
25 nice thing about CAD is it gives you an extra or a

1 different way to look at it and an extra
2 interrogation of the data. And so it's more just of
3 a detection as opposed to any diagnosis.

4 DR. D'ORSI: Yes.

5 DR. STEIER: For lung, at least in my
6 opinion, it would be a little bit different. We have
7 no trouble finding things on CAT scan and PET scan.
8 The problem is are they significant or not? You
9 know, at least the CAT scans I see, it's very unusual
10 to have a chest CT scan that doesn't have something
11 on it. So it's really more of an is-it-significant-
12 or-not issue, so maybe the CADx piece rather than the
13 finding. We can find a million things. You know,
14 again, scar tissue, granuloma, old, healed fibrosis,
15 you know, a whole host of things. But it really
16 becomes a significant issue as to whether it's
17 significant or not. Now, fortunately we're blessed
18 with PET scan, you know, which helps us a lot. Is it
19 active or not? And if the PET scan lights up, we're
20 a lot more aggressive. And if the PET scan doesn't
21 light up, we're not. But it's really a significance
22 issue more than a detection issue, in my opinion.

23 DR. D'ORSI: So if I'm hearing what the
24 people said, that the CADE for mammo and colonoscopy
25 is truly an E, but maybe not so much so for lung.

1 DR. STEIER: Yes. Yeah, that would be my
2 own opinion. Other people might differ.

3 DR. D'ORSI: Can people kind of weigh in on
4 that as we just go around? Dr. Swerdlow.

5 DR. SWERDLOW: Coming from the colon point
6 of view, I certainly agree with Dr. Kim. The only
7 issue, which is somewhat analogous to mammography, is
8 that, well, primarily we're looking for polyps, but
9 we also have the issue of flat lesions and some small
10 cancers, and that is somewhat analogous to
11 calcifications versus distortions versus mass in
12 mammography. And so we need to look at CAD systems
13 sort of with the independent morphology there.

14 DR. D'ORSI: Would you like to comment?

15 DR. JIANG: Well, from a technical point of
16 view, I think there's a separation between detection
17 and analysis. But I think I want to defer this to
18 the clinicians.

19 DR. D'ORSI: Dr. Tourassi.

20 DR. TOURASSI: Yeah, the same comment here.
21 But it is rather clear from the discussion and from
22 our experience that even with a CADe device, there is
23 some of the X component there hidden somehow. So in
24 the end, it's how the device is labeled, if it serves
25 the purpose for which it is labeled. This is where

1 we need to focus on. If the labeling is, is it
2 something actionable, as you said, let's define
3 actionable. Is it something suspicious? Let's
4 define that. Is it cancer? We will work with that.
5 Whatever the labeling is, we will have to focus on
6 that and derive the rules based on that.

7 DR. D'ORSI: Dr. Glassman.

8 DR. GLASSMAN: I agree. There is a
9 discrimination function in deciding how many CADe
10 marks to put on an image, and that, by definition, is
11 a CADx function. But it's at such a low level that
12 it's not trying to subdivide benign and malignant,
13 and therefore I think that the current devices are
14 truly CADe, although in the future I expect we'll see
15 CADx devices.

16 DR. D'ORSI: Dr. Dodd.

17 DR. DODD: So I guess I thought a
18 distinction was if I began to put a probability of
19 malignancy or the probability that this is an
20 actionable finding, then that begins to push this
21 over into a CADx function. And so, you know, even
22 though you may view this as a CADe, the minute I
23 start putting the likelihood of this on it, then it
24 falls into a different bend.

25 I think the relevant question for the Panel

1 as we move through the day will be what different
2 analysis questions does this CADx pose, and will they
3 be sufficiently different to warrant a different
4 convening of this meeting? And at this point, you
5 know, I do think there are different analysis
6 questions posed, but maybe by the end of the day, we
7 can highlight some of the distinctions.

8 DR. D'ORSI: Dr. Zhou.

9 DR. ZHOU: I think it will also depend on
10 the labeling of the device to say what's the intended
11 use that'll aid the physician to decide the
12 treatment, to decide the diagnosis. So I think that
13 either analysis or the design should be consistent
14 with what's intended use in the labeling.

15 DR. D'ORSI: Dr. Payne.

16 DR. PAYNE: I don't have any additional
17 comments.

18 DR. D'ORSI: Dr. Mittal.

19 DR. MITTAL: I think that CADx devices have
20 advanced functionality to CAde devices because all
21 the screening devices really have the functionality
22 to diagnose, so it's very difficult to differentiate.
23 So essentially for all three indications, whether for
24 mammography, colonography or lung cancer, it's an
25 advanced function of CAde.

1 DR. D'ORSI: Dr. Ziskin.

2 DR. ZISKIN: I think there is a distinction
3 and it's primarily an intent. It is the intent just
4 to put a mark on a film, and which has to be
5 interpreted and read by the radiologist, whereas a
6 CADx could be used as a final diagnosis, even. But I
7 see a distinction. Although most things could have
8 some CADx aspect to it, it's the intent, I think,
9 that is the bottom line, and also the labeling, as
10 was brought up.

11 DR. D'ORSI: Dr. Rosenberg.

12 DR. ROSENBERG: Yeah. I mean, I think most
13 of the devices now, the purpose is to draw attention
14 to something, which really puts it in the CADE
15 category. But there might be the additional
16 functionality if there were some indication of level
17 of concern, but that doesn't seem to be what we're
18 discussing now.

19 DR. D'ORSI: Dr. Abbey.

20 DR. ABBEY: Yeah, they seem different to
21 me, and they ought to be evaluated differently.

22 DR. D'ORSI: Dr. Lin.

23 DR. LIN: Yeah, I think I will agree with
24 Dr. Kim. With colon cancer screening, it's a little
25 different from with the lung and with the breast in

1 that there's a readily available and relatively safe
2 follow-up procedure than would be done with just
3 colonoscopy. In fact, colonoscopy itself is heavily
4 used as a primary screening tool. So in this
5 situation, the CAD device is really functioning much
6 more as a CADe device rather than a CADx device.

7 Now, having said that, it does -- the
8 virtual colonoscopy does report size, and there's a
9 strong correlation between size and whether or not
10 the polyp is going to be clinically significant, in
11 other words, precancerous. So there is also an
12 aspect of sort of the CADx aspect to this, but I
13 think it mainly functions as a CADe device.

14 DR. D'ORSI: Anything to add, Dr. Leitch?

15 DR. LEITCH: I would agree about the
16 pulmonary stuff, that it probably complicates things
17 more than aids in diagnosis. I do agree that they
18 are different in terms of indicating the lesion.
19 But, again, when the physician's thinking about it,
20 if you say a study has, you know, this false negative
21 and sensitivity issue, the physician's kind of
22 thinking about the value of it for that sort of false
23 negative and false positive. That's what the
24 clinician is thinking about, and that's what
25 you're -- in fact, that's what a lot of this data was

1 talking about, you know, that you pursue something
2 that doesn't need to be pursued.

3 And so that's why I think it's blurred
4 because when the clinician is -- when you say
5 sensitivity and false negatives and false positives,
6 they're thinking with respect to cancer, not with
7 respect to did I see a mark that looked like
8 something benign, and I don't need to -- you know, I
9 see the mark and I can just discount it. So I think
10 in the clinical application of this, that's what we
11 have to think about.

12 DR. D'ORSI: Dr. Seabert.

13 DR. SEIBERT: Seibert.

14 DR. D'ORSI: Seibert. Excuse me.

15 DR. SEIBERT: See, we're different.

16 DR. D'ORSI: I'm sorry.

17 DR. SEIBERT: Well, I agree with what
18 has --

19 DR. D'ORSI: I'll call you Tony.

20 DR. SEIBERT: They're -- Tony, that's
21 better -- different devices. I think labeling and
22 intent are the basic premises that have been
23 discussed many times.

24 DR. D'ORSI: Dr. Bourland.

25 DR. BOURLAND: I agree with these comments.

1 It seems like labeling and intent. But I do think
2 step one is find a region or volume, and I think,
3 under current definition, that is CADe. So even CADx
4 would have that. And the blur, of course, is adding
5 additional functions onto CADe starts equaling CADx.

6 DR. D'ORSI: Dr. Kim, anything to add?

7 DR. KIM: Not at the moment.

8 DR. D'ORSI: Dr. Carrino.

9 DR. CARRINO: Yes, I think CADe and CADx
10 can be different, but often they're the same, meaning
11 that the distinction gets blurred -- can get blurred
12 fairly quickly. And so if you're looking at, like,
13 the standalone systems, when you're calculating the
14 sensitivity and specificity, you're then assuming 100
15 percent probability for that system if we're using
16 the sensitivity and specificity to represent the
17 diagnosis of a pathology. If we're just using
18 sensitivity and specificity for the identification of
19 a finding or a feature, then that's a different
20 study. So does the system identify as many findings
21 as the reader? And so we have to clarify that. My
22 concern comes along with the methodological issues.
23 You know, are there similar methodological issues or
24 are there different methodologies? And, of course,
25 the logistical and organizational aspect of getting a

1 group of people this size together to kind of think
2 about these things and not, you know, focus on E and
3 then, you know, what about X?

4 So I do think keeping the strata, the
5 differences between detection, diagnosis, and then
6 the third part, screening. I know a number of
7 countries that are understaffed with radiologists who
8 would love to have a screening tool to say this chest
9 radiograph is completely normal and they don't need
10 to have a radiologist look at it. So I think those
11 three distinctions are important, but we should think
12 broadly about, you know, keep the end, you know,
13 start with the end in mind and kind of look toward
14 the whole package while we're focusing on CADE.

15 DR. D'ORSI: Dr. Duehring.

16 DR. DUEHRING: Once again, you know, I have
17 to support the thought that there is a distinct
18 difference and we should be concerned with the
19 labeled intent. Although the labeled intent and the
20 intent of the end user may vary, I think that we have
21 to be concerned with the intent of putting it on the
22 market.

23 DR. D'ORSI: Mr. Uzenoff.

24 MR. UZENOFF: Yeah, I think that
25 industry -- a position in general would be that it's

1 not advantageous to lump everything together, and to
2 the extent these buckets make sense, and especially
3 the labeled intent should have a strong bearing on
4 the degree of regulation, consistent with the level
5 of concern.

6 So whether it's two or three buckets or
7 maybe there's one name, but if the regulatory regime
8 is flexible and appreciates and recognizes and can
9 accommodate different levels of concern, that would
10 be an important issue. And also importantly, as this
11 deliberation is going on, is to have some regulation
12 that's clear and issued and that industry knows what
13 to do because there is -- so that they can act on it.

14 DR. D'ORSI: So if I'm hearing the whole
15 discussion, it sounds like we should probably keep
16 these separate, even though there may be blurring
17 uses with diagnosis. But for the purposes of what
18 we're charged with today, we're talking about a
19 detection-only device. Is that fair? Thank you.
20 Let's break for lunch. We'll come back at -- sorry,
21 Dr. Dodd.

22 DR. DODD: Sorry. I know, I learned
23 yesterday, if you don't ask a question now, you might
24 not be able to, yesterday. I just want to follow up.
25 Since nobody's asked a question about the reuse of

1 test data, could I ask Dr. Petrick a question? I
2 just want to make sure I understand what you're
3 proposing, and on Page 44 there's a pretty clear
4 description, I think. Is this a setting where
5 there's no reader study involved? Is it really for
6 standalone only or --

7 DR. PETRICK: So that would be a scenario
8 for standalone. It gets a little more complicated
9 than what you might -- in what might be possibilities
10 when readers are involved. If you get all new raters
11 for a particular study, then obviously they would've
12 never seen those cases before. If there was no
13 readjustment to the CAD, they wouldn't have been able
14 to learn anything from that particular dataset. So
15 there may be this option that you could reuse that
16 dataset for a different study. Now, exactly when you
17 do that, I'm not completely clear, but that's another
18 option.

19 DR. DODD: Right. And then, would you
20 recommend some limitation on the number of times the
21 data could be reused that's relative to the amount of
22 samples that are added to it?

23 DR. PETRICK: Yeah, and that's one of the
24 issues is, based on the number of cases that are
25 reused, in some sense you can continue the cycle

1 moving forward, and we have some work that we're
2 trying to work on in the laboratory now to look at
3 that in at least sort of idealized situations, to say
4 how many cases that you actually want to add in order
5 to control for some of that bias.

6 But I think if you have large numbers of
7 those variations and you're randomly sampling from
8 that dataset, or some reasonable number of cases,
9 anyway, then it is some of these things that
10 potentially could be perpetual in the sense that it's
11 constantly renewing itself, that you could use it
12 multiple times.

13 DR. DODD: And then, how about multiple
14 comparisons adjustment?

15 DR. PETRICK: So that's a question of,
16 statistically, what should you do? And at this
17 point, I guess I don't have a great recommendation.
18 What we've seen before in CAD in the literature and
19 in companies that have used data over again is that
20 they don't do an adjustment for it. But once you're
21 adjusting the data, what should the right adjustment
22 be? It's a little bit unclear exactly what that
23 should be. So it's a difficult question.

24 DR. DODD: Thank you.

25 DR. SEIBERT: One more quick question.

1 DR. D'ORSI: Dr. Seibert.

2 DR. SEIBERT: Dr. Petrick, how is the data
3 normalized? Let's say we have digital -- full-field
4 digital mammography and you get it. If the dataset
5 comes from one manufacturer, how does another
6 manufacturer use it? Obviously, you say you use raw
7 data, but raw data isn't necessarily raw data from
8 the sensitivity metric point of view. Could you
9 comment on that?

10 DR. PETRICK: So there are two
11 possibilities. One is you normalize the data that's
12 coming out of the different systems to some sort of
13 standard that is the input to your particular CAD
14 algorithm. That means there'll probably be some
15 variations between manufacturers and especially in
16 the noise characteristics for that system. That may
17 be very small and insignificantly related to your
18 algorithm or it may make a big impact. It's hard to
19 know. The other approach is that you develop
20 basically similar algorithms but slightly different
21 modifications for each individual piece of hardware
22 that comes in.

23 DR. D'ORSI: All right, let's break for
24 lunch. We'll come back at 1:00. You'll still have
25 chances to ask questions after the afternoon

1 speakers.

2 (Whereupon, a lunch recess was taken.)

3

4

5

6

7

8

9

10

11

12

13

14

15

16

17

18

19

20

21

22

23

24

25

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

(1:00 p.m.)

DR. D'ORSI: We have a big agenda, so let's start. We will now proceed with the Open Public Hearing. Public attendees are given an opportunity to address the Panel to present data, information, or views relevant to the meeting agenda. I'm going read a disclosure into the record.

Both the Food and Drug Administration and the public believe in a transparent process for information gathering and decision making. To ensure such transparency at the Open Public Hearing session of the Advisory Committee meeting, FDA believes that it is important to understand the context of an individual's presentation.

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

For example, this financial information may include a company's or a group's payment of your travel, lodging, or other expenses in connection with your attendance at the meeting. Likewise, FDA

Free State Reporting, Inc.
1378 Cape Saint Claire Road
Annapolis, MD 21409
(410) 974-0947

1 encourages you, at the beginning of your statement,
2 to advise the Committee if you do not have any such
3 financial relationships. If you choose not to
4 address this issue of financial relationships at the
5 beginning of your statement, it will not preclude you
6 from speaking.

7 The Panel will be given an opportunity to
8 ask questions of the public presenters at the
9 conclusion of the Open Public Hearing. If recognized
10 by a Panel member, please approach a podium to answer
11 the questions.

12 I would like to remind public observers at
13 this meeting that public attendees may not
14 participate except at the specific request of the
15 Chair.

16 Prior to the meeting, we received 10 formal
17 requests to speak during today's Open Public Hearing
18 session and can only accommodate these individuals.
19 As I call your name, please come forward to the
20 microphone. We ask that you speak clearly into the
21 microphone to allow the transcriptionist to provide
22 an accurate record of this meeting. You will have
23 five minutes for your remarks. When you begin to
24 speak, the green light will appear. The light on the
25 timer will turn yellow to warn the speaker when there

1 is one minute remaining. In the interest of fairness
2 to the other participants, we ask you to conclude
3 your statements within the five-minute timeframe.

4 The first speaker is Dr. John Deluca. Just
5 state your name and you can start. Thank you.

6 DR. DELUCA: Okay, good afternoon. I'm
7 John Deluca. I am Vice President of Regulatory
8 Affairs and Quality Assurance for iCAD. I am an
9 employee of the company. The company paid my airfare
10 and hotel for this meeting.

11 iCAD is a developer and manufacturer of CAD
12 medical devices. So today I want to present some
13 comments on the new CAD guidance, and I also want to
14 touch on another major area here, informed consent
15 for image collection, which is not necessarily in the
16 CAD guidance now, but it has a direct correlation to
17 clinical considerations for CAD devices.

18 So some general comments. We welcome the
19 efforts of FDA in issuing in this guidance. We hope
20 it's going to be a catalyst for change, particularly
21 in the area of regulatory submission review at the
22 Radiology Branch. CAD manufacturers have been
23 experiencing a gridlock in FDA review of submissions.
24 In particular, iCAD has experienced unreasonable
25 delays in our submissions, and I think the FDA

1 themselves told you today, the reason for that is
2 that they can't agree on how to handle CAD
3 submissions. All of this has caused new products and
4 improvements not getting to the market and our
5 patients in a timely manner. In general, the areas
6 of concern with the new CAD guidance is that it only
7 addresses new or significantly modified 510(k) CAD
8 devices and not Class III CAD devices. Secondly, it
9 lacks a robust testing and submission paradigm, and
10 I'll talk a little bit more about that in a second.

11 If we look at the 510(k) CAD guidance, it
12 provides examples where modifications to cleared
13 510(k) CAD devices may result in new submissions.
14 The guidance should focus on testing and clinical
15 considerations and device classifications for CAD
16 devices and not necessarily modifications to CAD
17 devices.

18 I think it's been clear the last two days
19 that there is an existing guidance that FDA issued
20 back in 1997, deciding when to submit a 510(k) for a
21 change to existing devices. This document was issued
22 in 1997. It has served industry well. It has served
23 FDA very well. It has a detailed flowchart to assess
24 modification to 510(k) devices, and it's clear that a
25 510(k) holder is the best-qualified person to make

1 that assessment.

2 With respect to Class III CAD devices,
3 there is no guidance or paradigm on when to submit a
4 PMA supplement for any type of Class III device, as
5 there is for the 510(k) devices. And that's the
6 guidance document I just spoke about in the slide
7 previously. So the current CAD device document does
8 not answer key questions -- key concerns of Class III
9 CAD manufacturers, and I'm going to run through a
10 couple of these. And some of these have been brought
11 up today, earlier, with the FDA questions to the
12 Panel.

13 How do Class III CAD manufacturers handle
14 incremental changes to Class III devices? For
15 instance, modifications to mammographic CAD software
16 interface to accept new digital DR/CR images with no
17 change to the core CAD algorithm.

18 For approved CAD devices that do not have
19 reader studies today, the Panel must understand that
20 there are at least two mammo CAD devices that have
21 been approved through the PMA process without reader
22 studies. For these types of devices, is standalone
23 testing still acceptable for incremental changes?
24 And the basis of approval for these mammo CADs was
25 standalone in the original PMA. For CAD devices with

1 valid reader studies, when do incremental performance
2 changes demonstrated through standalone testing
3 trigger another reader study? Industry needs a clear
4 paradigm issued in a timely manner to move forward.

5 Yesterday, the MITA folks put a chart
6 looking at risk and changes to devices. I want to
7 present one that I put together myself. So let me
8 just clearly say industry needs a testing paradigm
9 based on science, safety, efficacy, but it also has
10 to be proportional to the type of change. So if you
11 look at this rather simplistic chart, on the left
12 side you have changes that go from low risk to high.
13 And one change might be workflow, you're going to
14 make a simple workflow change in your CAD product.
15 For this we're suggesting that simple verification
16 testing, that that function works, would be
17 necessary.

18 If you're going to make a software
19 interface change, this might be considered a
20 moderate-risk change. This might be as I just
21 identified before. We sell our product to a number
22 of mammographic companies, and we make simple changes
23 to accept their images. So in this case,
24 verification testing might be appropriate, as well as
25 standalone testing.

Free State Reporting, Inc.
1378 Cape Saint Claire Road
Annapolis, MD 21409
(410) 974-0947

1 Now, if you make core algorithm changes,
2 you're going to make software changes where the
3 sensitivity or the specificity may change, this might
4 be considered a high risk, and all three types of
5 changes -- I'm sorry -- all three types of testing
6 paradigms might be appropriate for that.

7 I wanted to spend a moment here because I
8 took some notes earlier from the sessions because I
9 did not know exactly what FDA was going to say on
10 some of this on the reader studies.

11 DR. D'ORSI: Can we please just wrap it up
12 a little? We're on a tight schedule.

13 DR. DELUCA: Okay.

14 DR. D'ORSI: Thank you. You can just --

15 DR. DELUCA: All right, let me just go
16 through a couple more slides. So let me just cover
17 informed consent very, very quickly here. Image
18 collection is absolutely critical for CAD
19 development. It's the lifeblood of developing a CAD
20 system. We use images to develop, train, and test
21 the software. iCAD believes that informed consent
22 should be waived for de-identified, retrospective
23 data collection.

24 This is where we collect data from
25 institutions where the patient is long gone. Their

Free State Reporting, Inc.
1378 Cape Saint Claire Road
Annapolis, MD 21409
(410) 974-0947

1 image is there, their case records are there, and
2 those are the images and information we would like to
3 develop our product. FDA has consistently maintained
4 that informed consent can never be waived for
5 specimens that are non-identifiable, and further,
6 they do not allow IRBs to decide whether or not to
7 waive informed consent. This has had a significant
8 impact on the collection of data and the development
9 of CAD products.

10 So in closing, I would like FDA to
11 seriously consider exercising enforcement discretion
12 as to the informed consent requirements for de-
13 identified, retrospective image collection, as they
14 have for in vitro diagnostic leftover specimens.

15 I have one slide here that I'll just
16 quickly go through and summarize for the new
17 guidance.

18 DR. D'ORSI: We can read that as you go
19 down. Thank you.

20 DR. DELUCA: Okay, thank you.

21 DR. D'ORSI: The next speaker is
22 Dr. Maryellen Giger. Please try and stay within the
23 time limits so we can get adequate discussion time.
24 Thank you.

25 DR. GIGER: Okay. I am Maryellen Giger

Free State Reporting, Inc.
1378 Cape Saint Claire Road
Annapolis, MD 21409
(410) 974-0947

1 from the University of Chicago. I am here
2 representing myself. Some disclosures. My research
3 is supported by various grants from NIH, DoD, DOE,
4 and the University of Chicago. My conflict of
5 interest is here. I'm a stockholder in R2 Hologic,
6 and I receive royalties, and these dribble down
7 through the University from Hologic, GE Medical,
8 Median, River, Crane, Mitsubishi, and Toshiba. The
9 other conflict of interest, I'm the current President
10 of the American Association of Physicists in
11 Medicine, and I do mention them in my talk. However,
12 I'm not representing them here. I'm representing
13 myself only.

14 So as a CAD researcher, I'm concerned about
15 the timeliness and consistency of the translation of
16 CAD developments to clinical use. Computers are
17 everywhere. They're becoming a part of all walks in
18 our lives. What is important here is how do we
19 further the progress of CAD research, evaluate new
20 devices, expedite the process so the CAD can be
21 incorporated into clinical practice in a timely
22 manner? And how do we keep realizing over and over
23 again that the radiologist's training is very
24 important in how these systems are being used?

25 So in thinking about how to evaluate a new

Free State Reporting, Inc.
1378 Cape Saint Claire Road
Annapolis, MD 21409
(410) 974-0947

1 CAdE system, well, first we wanted a least burdensome
2 approach because if it's too burdensome, folks won't
3 come back with their improved models and just leave
4 their old ones out there and patients won't benefit
5 from new and improved systems. We want
6 standardization of testing, including scoring method
7 and ground truth. We want to maintain the integrity
8 of a test set and avoid reuse of data for testing.
9 We want to allow for a case mix that includes both
10 enrichment with different lesion types, different
11 cancer prevalence, to try to match the clinical
12 content. We want to allow for reproducibility
13 measures, and we want to realize the reader mix may
14 be the biggest problem here and that in ultimate
15 clinical practice, the user will change with
16 different CAD systems. So potentially we might need
17 to take the user out of this evaluation. In this I'm
18 talking about standalone.

19 So what I'm suggesting here is a two-stage
20 method of evaluation where we can consider here
21 double reading. If we can show that CAdE, that is,
22 that CAdE is similar to double reading, then
23 potentially we can move on to our standalone
24 evaluation. So the two-stage method is presented
25 here.

1 Stage 1. We need to determine a
2 performance level standard for CAD based on published
3 data and a cooperative study. For example, with
4 double reading, if we can show that CADe is
5 equivalent to double reading, and in this study use
6 radiologists trained in CAD usage with different CAD
7 systems, and perform -- and this would be a group
8 effort with industry, academia, et al. -- an ACRIN-
9 like study. The output from the study would tell
10 us -- and this would be one big, large reader
11 study -- would tell us what is needed as the minimum
12 bar, what is the minimum level of a computer
13 standalone performance in terms of detection
14 sensitivity and false positive marks per image. If
15 we knew that minimum bar, then we could move on to
16 the second stage, and this is where it would help
17 expedite the movement over to clinical practice.

18 We could evaluate the systems, and let me
19 move to the next slide. We want to evaluate them as
20 standalone, and this would be similar to the testing
21 of image acquisition devices, where you can accept
22 something based on, say, image quality metrics of MTF
23 signal-to-noise ratio. And here the test result
24 could be obtained for various populations and look at
25 sensitivity false marks per image.

Free State Reporting, Inc.
1378 Cape Saint Claire Road
Annapolis, MD 21409
(410) 974-0947

1 And the way to do this is with an
2 independent technology assessment institute. The
3 institute would have a protected and sufficiently
4 large database, and you can vary the enrichment of
5 this dataset. The dataset would be so large that the
6 integrity of the test would remain because when you
7 want it to test your device, it was brought to the
8 institute, randomly drawn, cases would come from this
9 test set to match the distribution of cases you need
10 in the cancer prevalence. This independent institute
11 would only give the manufacturer the test results,
12 such as sensitivity and false positive mark, and thus
13 the company could not train to the test set.
14 Evaluation method would eliminate the variation of
15 the radiologist mindset, skill, level of training, et
16 al. Retesting of the CAD system by the technology
17 assessment institute would yield measures of
18 reproducibility, since when you went in again, you
19 would use basically a different set of test cases.

20 The black box knowledge from the
21 manufacturer is maintained and output from this
22 evaluation could be given to the FDA for use. This
23 institute could have an oversight from a scientific
24 organization such as the APM, and one would expect it
25 could be used in FDA evaluations for different --

Free State Reporting, Inc.
1378 Cape Saint Claire Road
Annapolis, MD 21409
(410) 974-0947

1 both screen-film and full-field digital mammography,
2 in order to test the label and include the test
3 scores in this label.

4 So, in summary, this would give us the
5 least burdensome approach to demonstrate substantial
6 equivalence. It would standardize the testing,
7 including scoring method and ground truth. It would
8 maintain the integrity of the test set and avoid
9 reuse of data for testing. It would allow for a case
10 mix of enrichment and matching to whatever the
11 clinical content prevalence that you would want, and
12 includes random samples of the population. It would
13 give you reproducibility metrics. It would avoid the
14 reader mix problems because this institute approach
15 would eliminate reader variability bias, training
16 level, et al. And the clinical utility question on
17 this would've been shown in the first stage of this
18 two-stage process.

19 Things we need to consider, though, is the
20 role of the user interface. If they substantially
21 change their user interface, this may not work. And
22 also different types of false positives may be --
23 even though two companies may have the same amount of
24 false positives per image, they may have different
25 types of false positives, but that could potentially

1 come out in the post-market evaluation.

2 So please consider the role of a technology
3 assessment institute in pre- and post-market
4 evaluations for CAdE, and this would also be useful
5 for CAdx. Thank you.

6 DR. D'ORSI: Thank you very much. The next
7 speaker is Dr. Zuley.

8 DR. ZULEY: Good afternoon. My name is
9 Rita Zuley. I'm an Associate Professor of Radiology
10 at the University of Pittsburgh. I am here
11 representing the American College of Radiology and
12 the Society of Breast Imaging. They have paid for my
13 trip here.

14 So I want to give some clinical
15 perspective. CAD is one input that we use as
16 radiologists to determine clinical action. We also,
17 for example, look at information for magnification
18 views, or changes when there's slice thickness
19 variation in CT or detector changes or protocols in
20 MR or transducer frequencies with ultrasound. This
21 is just one thing that we use.

22 FDA's role is to determine the safety and
23 effectiveness of a device, not the quality and
24 effectiveness of the radiologist using that device.
25 For MRM studies that we're speaking about, the

Free State Reporting, Inc.
1378 Cape Saint Claire Road
Annapolis, MD 21409
(410) 974-0947

1 testing, the interaction of the radiologist with the
2 CAD includes many variables that are outside the
3 control of the CAD device. The inter- and intra-
4 observer variability is so great that we would need
5 very large-scale studies for this to be appropriate.

6 We heard about learning effect earlier
7 today, and there was a study already discussed from
8 Elmore looking at learning effect. And so we had
9 reader recall rates before CAD was introduced into a
10 set of readers at 6.2 percent. The recall rate went
11 up as the radiologist became familiar with the CAD,
12 and once they became accustomed to that CAD
13 algorithm, the recall rate went back to normal.

14 So when do we do reader studies with this
15 experience? If we don't and the radiologists don't
16 have training, the arm that looks at the new ROC or
17 the new CAD will be negatively impacted. So do we
18 wait two years from a new product being introduced to
19 do such a study?

20 Now, let's think about the keep all the
21 positives rule. The first thing you learn when
22 you're given a CAD is don't let the CAD stop you from
23 recalling something that you saw on the film.
24 Therefore, if you follow that rule and you're a good
25 radiologist, there's no way you're going to have

1 false negatives introduced into the safety and
2 effectiveness of the CAD. That would be the fault of
3 the radiologist, not the device.

4 Now, let's look at mammography, which is --
5 I've been using CADe since 1998. It was the first
6 products for CADe introduced in 1998. They are the
7 oldest and most studied CAD products on the market.
8 So why is it that mammography CAD is classed as a
9 Class III device when lung and CT colonography CADs
10 are Class II? There is valid scientific evidence for
11 mammography CAD. There's a large body of literature
12 supporting the safety and effectiveness more than CT
13 and lung. And so is this question true? No.

14 If we look at safety, is breast cancer a
15 higher risk or has a higher death rate disease than
16 lung or colon cancer? No. Again, looking at safety,
17 is recall at mammography because of a false positive
18 mark or additional views or six-month follow-up
19 higher radiation exposure to the patient than a
20 follow-up CT scan? No. Is breast biopsy more
21 invasive than lung or colon biopsy or colonoscopy
22 with sedation? No.

23 The effectiveness of CADe for mammo has
24 been shown through multiple studies. We know that we
25 can find smaller and earlier cancers with CAD. If we

1 look at the range of literature out there, the one
2 that showed the least benefit actually came from my
3 home institution, where we only saw 1.7 percent
4 improvement using sub-specialists like myself. There
5 are groups out there showing almost a 20-percent
6 benefit in private practice. No study has shown a
7 decrease in detection with CAD mammography, and most
8 mammograms are read and interpreted in this country
9 by private practitioners, who would yield the most
10 benefit from a CAD device.

11 So if a new CAD device can show that cancer
12 detection sensitivity and false marking rate are
13 similar to predicate devices already approved through
14 a cohort of cases, that should suffice for its 510(k)
15 approval without reader studies. And if two FFDM
16 systems are substantively equivalent, the CAD
17 algorithm testing should be standalone testing as
18 well.

19 So, in summary, safety and effectiveness of
20 CADe mammography is well documented and should be
21 reclassified as a Class II device. Current
22 guidelines will stop CADe development, and therefore
23 there will be no further benefits to patient care.
24 Thank you very much.

25 DR. D'ORSI: Thank you. The next speaker

Free State Reporting, Inc.
1378 Cape Saint Claire Road
Annapolis, MD 21409
(410) 974-0947

1 is Dr. Robert Nishikawa.

2 DR. NISHIKAWA: I'm Bob Nishikawa from the
3 University of Chicago. These are my financial
4 interactions with different companies, but I'm here
5 representing myself.

6 Commenting on both guidance documents, I
7 think where the research is clear, the guidance
8 documents have a reasonable approach to market.
9 However, when there's uncertainty, the approach is
10 there on the side of more rather than less. And this
11 produces a document that is extremely burdensome, and
12 I feel that if this document goes into effect as is,
13 it'll kill new and innovative products in CADE.

14 So I understand that the FDA is in a no-win
15 situation. If they're too lenient, non-effective
16 products go to market and there are consequences of
17 that. But if they're too restrictive, effective
18 products will be delayed, and that's going to result
19 in more than companies are going to be upset.
20 Healthcare will be compromised, and in the case of
21 CAD, there may be women who die from breast cancer
22 who would not otherwise have died if CAD was used.
23 So I'm going to give a couple examples of where I
24 think the document is overly burdensome.

25 The first one I was going to talk about was

Free State Reporting, Inc.
1378 Cape Saint Claire Road
Annapolis, MD 21409
(410) 974-0947

1 reuse of test data, but I think Nick Petrick covered
2 that quite well and what he described, I think, is a
3 reasonable approach. I just want to emphasize to the
4 Panel that collecting datasets are extremely
5 difficult. It's not easy. It's almost as hard as
6 making the algorithm. In some cases it's harder than
7 making the algorithm. It's not an easy thing to do.
8 So collecting datasets for anything, observer studies
9 or anything, is difficult. And I think the approach
10 that Nick described will overcome the over-fitting
11 possibility that Dr. Gwise gave.

12 And then I'd like to comment on the
13 proposed observer study. For a PMA, there were three
14 different arms, read CAD not used arm, a sequential
15 read, and then a concurrent read. Three reads when
16 only one is needed is overly burdensome. Costs are
17 higher, it's much, much harder to recruit
18 radiologists, and the time to complete the study is
19 much longer. And these go up by more than a factor
20 of three.

21 In the guidance documents it says a reading
22 scenario should be consistent with the intended use
23 of the device, and therefore a concurrent or
24 simultaneous read is not needed because the CAD is
25 labeled as a second reader. I understand there's a

1 possibility that people use it as concurrent reading,
2 but there's no data, clinical or otherwise, to
3 suggest that it's good or bad. And I believe that
4 it's overly burdensome to a manufacturer to answer a
5 research question in their submission that goes
6 beyond their labeling claims.

7 We don't need this first observer study.
8 Two different studies have shown that the first
9 component of a sequential read is essentially the
10 same as independent reading without CAD. And so we
11 only need a single observer study.

12 There's some question about post-approval
13 studies. I don't think these are needed. The
14 clinical studies show that CAD has a comparable
15 increase in sensitivity and recall rate, and if you
16 look at the appropriate studies, which I'll define in
17 a second, the increase in sensitivity is about nine
18 percent and the positive predicative value stays
19 about the same. And if you ask a radiologist, here's
20 a technology that you can prove that your sensitivity
21 and positive predicative value stay the same, they're
22 going to be for it.

23 Okay. So now you have to -- to get my
24 result, you have to look at the way the different
25 clinical studies are performed. And if you look

1 at -- if you read this paper that I wrote with
2 Lorenzo Pesce, we show that a historical controls
3 method for collecting whether CAD is effective or not
4 is not an effective way. So you have to kick out
5 those studies, that leaves seven left, and you can
6 get that result I showed.

7 And I just want to finish with one thing.
8 If radiologists perform comparably on different
9 digital systems and the standalone performance on
10 those -- of CAD on those different digital systems
11 are the same or comparable, I don't see why you have
12 to do an observer study for every single
13 manufacturer.

14 Thank you.

15 DR. D'ORSI: Thank you. The next speaker
16 is Stephen Vastagh.

17 MR. VASTAGH: Good afternoon. I don't have
18 a presentation for my talk. I'll just try to close
19 this. Very good.

20 Mr. Chairman, distinguished members of the
21 Panel, FDA staff, my name is Stephen Vastagh, and I'm
22 the Director of Industry Programs at the Medical
23 Imaging and Technology Alliance, which is a trade
24 group of manufacturers of medical imaging and
25 radiation therapy equipment manufacturers, and as

1 such, I am paid by the Alliance, which is a trade
2 group which is financed by the manufacturers.

3 I will be followed by three colleagues who
4 will be speaking specifically to the issues in the
5 guidances. I'd like to make just a couple of general
6 comments.

7 First, we ask you to note that the comments
8 that will be made this afternoon do not represent the
9 totality of the issues that industry is interested in
10 commenting and will comment in the written
11 submissions. It's just the availability of the time
12 between the announcement of the guidance and this
13 hearing, this Panel meeting, that did not allow a
14 complete listing and discussion of the issues.

15 Second, the FDA is posing nine complex
16 questions to you, the Panel, which in fact encompass
17 some 30 separate questions. These questions were
18 announced just two workdays in this Panel meeting. I
19 want to share with you that the manufacturers'
20 representatives, more than a dozen of them, which
21 include inventors and developers of CAD products,
22 have spent more than 20 hours together over the past
23 two days, and still they could not answer some of
24 these questions without further research. Thus, we
25 fully appreciate how challenging these questions are

1 for the members of this Panel.

2 However, your answers are critical for all
3 the stakeholders in the clinical and scientific
4 community because they'll greatly influence the
5 future course of action by FDA on these guidances and
6 on the review of CAD products for many years.
7 Perhaps the timing for responding to these questions
8 would be more appropriate with greater notice and
9 with the benefit of input from the public comments,
10 which will not be on the public record until January
11 19th, 2010, and they will be the first of the public
12 comments on these CAD guidances.

13 Now, I would like to review the -- some
14 statistics about CAD product submissions and
15 approvals. There has not been in -- according to our
16 review, there has not been a CAD product cleared or
17 approved in the United States since October 18th,
18 2006. It's now the end of 2009. And it's not for
19 the lack of trying. Since 2006, we estimate that
20 there were about 10 to 20 attempts made by
21 manufacturers to submit applications for clearing or
22 approving new CAD products. To compare, there were
23 10 CADE applications cleared from 2003 to 2006.

24 It has become clear, at yesterday's FFDM
25 Panel meeting, that the delays within FDA were caused

1 by the split within the professional staff. A
2 handful of them already has been able tie up a large
3 portion of the new medical imaging technologies,
4 mammography CAD image processing, for the last two or
5 three years by insisting that every or nearly every
6 submission requires clinical studies, which resulted
7 in non-submission or non-approval of applications.

8 In turn, this discourages investment,
9 research, and new product development. In addition,
10 new direction such as imaging as biomarkers also
11 suffer indirectly. This also resulted in limiting
12 access of patients to newer technologies. It caused
13 economic hardship, driving some companies to the
14 brink because no products are approved for sale.

15 The lack of new products has also stopped
16 new research by academic institutions. This dry
17 spell in research will have impact in future years.
18 It has also impacted jobs in this industry. It's
19 amazing that so few persons have been able to cause
20 all of this. It's truly amazing. Ladies and
21 gentlemen, the question is, is it right?

22 Lastly, I'd like make this comment, that it
23 should really not be considered a foregone conclusion
24 that industry, the CAD industry will continue to
25 develop CAD technologies for use in the United

1 States. Without an enormous investment by small and
2 large companies alike, without significant interest
3 from medical futurists, CAD would not exist to the
4 extent it does today.

5 But CAD will not continue to progress as a
6 science, as an area of research interest, or as a
7 growing series of more and more sophisticated
8 products without products actually reaching
9 healthcare providers, being used in healthcare, and
10 benefiting patients.

11 Thank you very much for your time.

12 DR. D'ORSI: Thank you. You're welcome.

13 And the next speaker is Steve Slavens with GE.

14 MR. SLAVENS: Good afternoon. I'm
15 Stephen Slavens with GE Healthcare. I come here
16 sponsored by GE Healthcare, but I represent the NEMA
17 CAD committee as chair of that committee. MITA would
18 like to make some observations from our review of the
19 guidances in the period of time that they've been
20 available. It includes our look in the last few days
21 at the questions you'll be asked today, and we
22 haven't fully reflected on what we heard prior to
23 this morning, so I will take the opportunity to
24 comment at a later moment.

25 Let's remind ourselves of how CAD is

Free State Reporting, Inc.
1378 Cape Saint Claire Road
Annapolis, MD 21409
(410) 974-0947

1 typically used. The most common clinical protocol is
2 that the physician reads the image and forms a first
3 opinion. The physician then triggers the CAD device,
4 displays and reviews the regions indicated by the
5 CADe, and then forms a final opinion. The CADe is
6 thus an adjunctive decision-making tool and does not
7 directly result in biopsy, for example. That appears
8 to be one of the major risk concerns. And at this
9 point, of course, we would point out that a highly
10 trained radiologist acts as the final learned
11 decision maker.

12 Let's look at a rule that is again commonly
13 applied in CADe. The physician should always -- and
14 this is covered by manufacturers' labeling -- read
15 the case prior to the display of the CADe marks, and
16 then, secondly, we advise that the physician should
17 never not work up a finding if she or he is concerned
18 that the CADe mark failed to mark it.

19 Let's take a look at some of what we heard
20 yesterday. There is a requirement that is listed in
21 the document for multiply scanning a patient.
22 Drs. Pisano and Zuley yesterday expressed that in the
23 medical community they disapprove of double exposure
24 unless it's medically indicated. And this has the
25 detrimental effect to us as sponsors of studies, that

1 it's very difficult to get patients enrolled because
2 we get very little support from the medical community
3 in those trials.

4 FDA additionally does not require that the
5 underlying modality that acquires the image go
6 through double imaging. So it's odd to us to see a
7 requirement that the post-processing software needs
8 to go through such a multiple scan. And then there
9 are alternatives that exist, such as the use of
10 phantom or simulations. We believe that multiple
11 scans should, thus, not be required under any
12 circumstances solely for CAD.

13 Let's look a moment at the definition
14 because it raises some concerns. It's a long
15 definition. There are three CADEs that have been
16 announced that have been defined. But underlined are
17 two aspects of those definitions that have not been
18 articulated. The intent of the guidance, as we heard
19 it earlier, is to be generalizable over all future
20 CADs, and if you read that, I think you will have a
21 reaction as we have had, that it blurs the
22 distinction between where does an advanced windowing
23 or leveling algorithm advance all the way up to CADx?
24 This blurs the line, and it's so vague that we need a
25 more precise definition.

1 Comparison with predicate device is an
2 important issue to us. The FDA suggested a 510(k)
3 product performance have a comparison with the
4 predicate. But we have difficulty complying with
5 that because, as has been noted earlier today, the
6 predicate device may be of a competitor's and thus
7 unavailable, and the predicate dataset may not be
8 available for our use. And we may not even know
9 enough about the composition of the dataset in order
10 to try to replicate it.

11 The guidance structuring causes us problems
12 also. It is confusing because it interweaves the
13 comments about requirements for a 510(k) and a PMA.
14 What would help us out would be to make it less
15 burdensome, clearly separating and clarifying the
16 difference in the requirements between the two
17 regulatory regimes.

18 Determination of intended use, according to
19 the guidance that has been multiply cited in the last
20 two days, it instructs that the manufacturers define
21 the intended use and then have the opportunity to
22 contraindicate other uses for which they have not
23 validated device performance. The intended use
24 statement and the data provided in the labeling will
25 reflect how the CAde device should be used.

Free State Reporting, Inc.
1378 Cape Saint Claire Road
Annapolis, MD 21409
(410) 974-0947

1 But a concern appears to be off-label use,
2 and it is reflected, as you see, when the FDA speaks
3 of the encouragement of reading scenarios and
4 specifically concurrent reads that are outside the
5 manufacturer's intended use. We have not yet seen
6 that FDA has justified this expectation that
7 manufacturers would need to test outside the
8 contraindicated use. We believe that this is not
9 least burdensome.

10 So as a way to kind of help things going
11 forward, we spoke yesterday at the FFDM meeting of a
12 paradigm that looks at the risks and the technologies
13 and suggestions of what might be a reasonable way of
14 performing testing on that. While you look at that,
15 I will introduce our next speaker, then, Terry
16 Sweeney from Philips.

17 MR. SWEENEY: Thank you. This is
18 Terry Sweeney, and I'm with Philips Healthcare,
19 representing MITA here today. I'm the Vice President
20 of Clinical Affairs for Philips.

21 As we look at the balanced, risk-based
22 evaluation proposal here -- and again, it is aligned
23 with what we suggested yesterday -- as a tiered stage
24 approach towards evaluation of product changes and
25 new product introductions, as to what category they

1 may fall into and what type of evaluation of those
2 products should take place.

3 If we look at the lowest level category
4 highlighted in yellow here, under Case 1, we're
5 talking about a minor modification to an existing
6 device that's already been cleared for market. These
7 may be minor changes of such that we would be
8 evaluating using a standalone approach to test and
9 evaluate the product, and those modifications is what
10 we would suggest for a minor change to the product
11 itself.

12 Under Case 2, this is now a new device but
13 is a similar device to predicate devices that are
14 already on the market. Again, we would suggest that
15 standalone testing is sufficient for these new CADE
16 devices if the product has a direct comparison to the
17 predicate devices and demonstrates equivalent or
18 better performance using the standalone approach.

19 Looking at the highest category here of
20 potential risk and maybe the introduction of new
21 technologies or new indications for use that have
22 never been seen or evaluated before by the FDA, we
23 would again recommend a standalone evaluation with
24 appropriate clinical data supplied, as necessary,
25 based on the nature of the change or the technology

1 being introduced to the marketplace.

2 In looking at standalone testing, FDA, in
3 their guidance document, represents a hypothetical
4 situation in which the new CAdE identifies additional
5 abnormalities that are not detected by the predicate
6 device, but misses some of the abnormalities that
7 were detected by the predicate device. NEMA believes
8 that unless there's a large disparity in the true
9 positive CAdE findings, that reader variability will
10 have a much bigger effect, as some of the other
11 speakers have already talked about today, than the
12 actual algorithm variability itself. So we believe,
13 again, the standalone performance evaluation is
14 sufficient under these cases.

15 In looking at the definition of minor
16 modification, which I think the Panel will be
17 addressing today, MITA considers minor types of
18 changes to be along these lines. Software
19 environmental modifications do not change the
20 underlying CAD algorithm. These kind of changes
21 occur all the time at manufacturing facilities, such
22 as operating systems or compilers that are used to
23 develop the software, sometimes modifications are
24 made to that type of equipment.

25 Design changes to the surrounding

Free State Reporting, Inc.
1378 Cape Saint Claire Road
Annapolis, MD 21409
(410) 974-0947

1 applications, such as DICOM input/output. These are,
2 again, DICOM standards that we comply with, are user
3 interfaces that the user may have easier interaction
4 based on changes and modifications to the equipment,
5 computer hardware changes, or retraining of the
6 algorithm, based on expanded database. So, again, we
7 are always trying to improve the performance of our
8 products and continued retraining and improvement
9 should not be considered a significant or major
10 change to the product.

11 Further, under minor modifications,
12 industry asserts that any minor modification will be
13 followed by software testing conducted in-house to
14 establish similarity with the prior device
15 technology, and also the standalone algorithm testing
16 will ensure that there's a similarity with the prior
17 device's capabilities. And this would then be
18 documented into our quality management systems, which
19 are required to record the evidence and validation of
20 these types of changes. Minor modifications do not
21 result in changes to sensitivity or specificity
22 claims.

23 We also are looking at the use of -- a
24 single use of test datasets. There's been several
25 discussions earlier today, and MITA is very

1 encouraged by what they heard from the FDA today
2 regarding the potential to continue to reuse some of
3 the datasets and perhaps, with minor modifications,
4 continue to use them forward into the future. So we
5 are willing to work with the FDA to develop the
6 methodology and what it would take to actually modify
7 those datasets to continue to be able to use them
8 going forward.

9 And looking at single-use datasets, if each
10 algorithm improvement that we did want to make to a
11 product required a whole new dataset, then we'd have
12 a dramatic problem in that we would not be able to
13 acquire enough patient files to continually keep up
14 with the update on the changes and the flow of the
15 development processes that we all have. And as such,
16 we would have to develop a new database for each and
17 every one of these. It would prohibitive and could
18 not go forward with the level of change that would be
19 desirable. So therefore, again, the input from the
20 FDA is very encouraging today, that reuse of data may
21 be allowed under certain circumstances.

22 I'd like to introduce now Julian Marshall
23 to discuss the powering of subgroups. Thank you for
24 your attention.

25 MR. MARSHALL: Good afternoon. I'm

Free State Reporting, Inc.
1378 Cape Saint Claire Road
Annapolis, MD 21409
(410) 974-0947

1 Julian Marshall. I am an employee of Hologic, and
2 I'm here representing MITA. Hologic did pay for my
3 trip.

4 Regarding pairing studies for subgroups,
5 the FDA, in the guidance, posed is there a minimum
6 number of cancers that should be included in their
7 clinical study to ensure that the entire spectrum of
8 cancer is represented? MITA's answer is no. While
9 CADe manufacturers strive to collect large, broad
10 datasets, it's not possible to ensure that a dataset
11 contains the entire spectrum of cancer. In addition,
12 particularly where stress test datasets are required,
13 the full spectrum of cancer may not be represented.

14 In addition, FDA poses, should a CADe
15 device's clinical performance assessment be powered
16 so that statistically significant results can be
17 obtained for the clinically relevant subgroups? And
18 MITA believes it really should not be necessary to
19 power the study in order to objectively measure
20 performance on subgroups unless the manufacturer
21 proposes to make such claims.

22 Regarding new acquisition devices,
23 standalone performance testing is sufficient on a
24 database that is scaled to show CADe non-inferiority,
25 collected ideally using the new acquisition device.

1 But also we should realize there can be contributions
2 from simulated lesions on normal images from the new
3 acquisition device.

4 Regarding trade secrets, industry spends
5 millions of dollars developing proprietary CADE
6 algorithms. These devices provide a useful clinical
7 function, and ensuring that testing is sufficient to
8 establish safety and effectiveness or performance
9 measures is the responsibility of the manufacturer
10 with the oversight of the FDA. The testing data
11 should be allowed to speak for itself. But exactly
12 how CADE devices provide that clinical function is
13 not relevant to their performance.

14 The list of algorithm details requested by
15 FDA spans three pages and includes sufficient
16 disclosure to allow complete reverse engineering of
17 the algorithm. For example, how an algorithm
18 determines selection of seed points of region
19 segmentation is not important to understanding the
20 safety and effectiveness of a device. Providing this
21 level of detail, which is not necessary for device
22 evaluation, is extraordinarily time consuming and
23 would not be least burdensome. In addition, we're
24 concerned about inadvertent or inappropriate
25 disclosure of trade secret information, and there

1 have been cases in the past where information about
2 vendor products has leaked out of the FDA.

3 Decisions regarding the need for additional
4 clinical evaluation should be made utilizing the same
5 logic applied to other Class II 510(k) devices. CAdE
6 should be recognized and risk-managed as the
7 adjunctive decision-making tool for competent
8 physicians. The Agency should consider the use of
9 scientific literature reviews with comparative
10 analysis with predicate devices as a method of
11 reducing the clinical burdens imposed upon the
12 manufacturer even for minor changes.

13 We have other observations, and I'll let
14 you read those. Thank you for your time.

15 DR. D'ORSI: Thank you. The final speaker
16 is Mr. Morgan Nields.

17 MR. NIELDS: Thank you. I don't have any
18 slides, and I just have a few brief comments. As a
19 matter of disclosures, my financial disclosures are
20 on record with my talk yesterday, but with respect to
21 the CAD industry, I have no involvement whatsoever,
22 have no investments in a company that does make CAD
23 products, nor is it likely to make CAD products, from
24 what I've been listening to.

25 (Laughter.)

Free State Reporting, Inc.
1378 Cape Saint Claire Road
Annapolis, MD 21409
(410) 974-0947

1 MR. NIELDS: I just have three things to
2 comment on, PMA, databases, and regulatory
3 guidelines. With respect to PMA, right now in the
4 case of breast devices, well, yesterday you discussed
5 the issues of what's required for 510(k) for full-
6 field digital mammography that was proposed by the
7 FDA three and a half years ago. I suggest we get on
8 to it, as the ACR suggested, and reclassify these
9 algorithms. There's no reason they should be in the
10 PMA category.

11 The database issue is an important one, and
12 every company has made reference to that. But much
13 like the tissue banks that the NCI has funded that
14 has spawned the creation of innumerable discoveries
15 in the human genome and proteomics which has allowed
16 a lot of new biologics to be developed to treat
17 cancer, I think this is similar. I think the federal
18 government has responsibility to create those
19 databases. Dr. Maryellen Giger made reference to an
20 institute. However it's organized, the data needs to
21 be collected on a national basis. So these types of
22 tests can be done using those national databases.

23 With respect to regulatory guidelines, you
24 heard yesterday, I think, all the questions that were
25 being asked. It was all about ROC curves. Well, the

1 ROC curve, every radiologist has their own ROC curve,
2 and the devices that they're using aren't really
3 changing their ROC curve. It's trying to help them
4 improve their ROC curve. And I think the FDA's
5 regulatory mindset is to test this. These are valid
6 scientific questions. It's the wrong time and the
7 wrong arena. These tests should be done in the
8 clinic, published by academic institutions and others
9 who have an interest in trying to improve the
10 technology to improve patient care.

11 And lastly, if CAD is important -- and I
12 believe it is. I'm an observer. I look at the data.
13 I think it's important. You will not have a CAD
14 industry. I've been associated in the imaging
15 industry for more years than I can image, 30-
16 something. I've never heard an industry statement
17 that I heard today. I've never heard words like
18 that. You will not have a CAD industry. It doesn't
19 matter if you're a startup, a million, billion dollar
20 company, a hundred billion dollar company. If it's
21 uneconomical, it's not going to happen. And right
22 now, this is on the threshold, on the tipping point
23 of not having these products available at all.

24 Thank you.

25 DR. D'ORSI: Thank you. Thank you to all

Free State Reporting, Inc.
1378 Cape Saint Claire Road
Annapolis, MD 21409
(410) 974-0947

1 the speakers. Does anyone on the Panel have any
2 questions for any of the speakers? Yes, Dr. Abbey.

3 DR. ABBEY: For Dr. Giger. You brought up
4 the idea of this national institute, and I'm hearing
5 a lot of requests that we proceed in a very timely
6 fashion. How fast do you think an institute could be
7 up and running and have enough cases that the FDA
8 could call upon it for?

9 DR. GIGER: Well, I would hope for such an
10 institute, that we would get participation from the
11 government to start, either a grant or a contract
12 similar to ACRIN was started. Who would've thought
13 we would be doing large-scale imaging studies and
14 also industry contributing a certain portion of it.
15 And if also industry, at least in the first pass,
16 gave their images, and academia.

17 I'll be optimistic and say I would think if
18 we really had funding right away and started working,
19 probably within a year we could have a skeletal part.
20 So maybe for one disease type computer-aided
21 detection, say, mammography CADe, if we had enough
22 buy-in from industry, academia, and the government to
23 set it up. And we would have to, of course -- so
24 that's my answer.

25 DR. CARRINO: As a follow-up, do you think

Free State Reporting, Inc.
1378 Cape Saint Claire Road
Annapolis, MD 21409
(410) 974-0947

1 it would be suitable for those who have NIH funding
2 to, you know, do imaging research, to have the NIH
3 require them to submit their imaging datasets and
4 that'd be a possible repository?

5 DR. GIGER: Yes, I think so because
6 currently, for NIH grants now, we have to have a
7 data-sharing clause in our grant to get funded, and
8 those would go to the institute. I think a concern
9 with the institute, once it's done, it would be
10 great, and there's going to be some growing pains,
11 and one of it is the quality of the truth, which I
12 had listed in a slide but I was running out of time,
13 it went too fast. We would not just collect the
14 data, but we also need to verify the truth, and then
15 we could start some comparisons.

16 DR. D'ORSI: Dr. Bourland.

17 DR. BOURLAND: And so the follow-up, then,
18 is to the MITA group, which maybe you can answer or
19 not. But, yes, you would agree with a validated,
20 robust database accessible to all with truth thereof
21 known, et cetera?

22 MR. SWEENEY: Terry Sweeney, Philips
23 Healthcare, again for MITA. Yes, we would agree to
24 such a database being established and support such an
25 activity. Again, one of the issues is informed

1 consent, though, and being able to get that informed
2 consent to be shared. Sometimes they're very limited
3 and very specific about where the informed consent
4 may be to a specific company. So if we wanted to
5 share across all companies, we would have to get,
6 again, a waiver on the informed consent issue, where
7 that data could be shared across the whole of the
8 industry.

9 DR. BOURLAND: And if appropriate, can I
10 direct something to the FDA, whether they would --

11 DR. D'ORSI: Yes, sure.

12 DR. BOURLAND: Would that be all right with
13 FDA?

14 DR. D'ORSI: Yeah.

15 DR. BOURLAND: And then would they be
16 willing to let IRBs handle the waiver for
17 retrospective studies in particular?

18 MS. MORRIS: The issue of informed consent
19 is a very sticky wicket. That may require some
20 regulatory changes. So we couldn't give an
21 endorsement of that if it was dependent upon the
22 informed consent issue. This is a topic of interest
23 with the Agency, and it's trying to address it. But
24 at this point in time, we haven't come to resolution
25 with respect to our statute and regulation, the

1 requirements.

2 But the concept that's been discussed and
3 if there is cooperation with the industry, certainly
4 FDA would be interested in partnering and trying to
5 move things forward, as long as it addresses our
6 statutory requirements with the regulations.

7 DR. D'ORSI: Yes, Dr. Leitch.

8 DR. LEITCH: I think this is for
9 Stephen Slavens, who was talking about comparison
10 with predicate device not being available to test
11 against. So what would you test against? What would
12 be your ideal study, then?

13 MR. SLAVENS: Well, the study that was
14 described for PMA, essentially with the CAD on and
15 CAD off. So, for example, we may have a device
16 that's already indicated for use to assess a specific
17 disease type, and then we would do the readings prior
18 to and then with the markers on.

19 DR. LEITCH: Okay.

20 MR. SLAVENS: That would be the nature of
21 our study.

22 DR. LEITCH: So a reader study, though?

23 MR. SLAVENS: Correct.

24 DR. LEITCH: Okay, not a standalone study?

25 MR. SLAVENS: Correct.

1 DR. LEITCH: Okay, okay.

2 DR. D'ORSI: Dr. Mittal.

3 DR. MITTAL: I have a question for a MITA
4 representative. Will the industry be willing to
5 support the concept of funding of the institute? I
6 think it's an interesting concept to have an
7 institute for this purpose, if the federal government
8 is not willing to give you any grant.

9 MR. MARSHALL: This is Julian Marshall, and
10 maybe this is just my opinion because we haven't had
11 this discussion within MITA or within my company.
12 But I think you have to be mindful that there are
13 companies of all sizes that want to participate in
14 this commercial market. And so however an
15 organization got structured, the cost would have to
16 be reasonably borne out by organizations of very
17 different sizes. So that's one cautionary note. And
18 then secondly, if there were no government funding
19 available, then the question is what would it cost?
20 And I think we'd have to evaluate those numbers for
21 ourselves. So with no estimates, I don't think that
22 can be answered.

23 DR. D'ORSI: Mr. Uzenoff.

24 MR. UZENOFF: Bob Uzenoff. Yeah,
25 Dr. Mittal, there is some ongoing history of exactly

Free State Reporting, Inc.
1378 Cape Saint Claire Road
Annapolis, MD 21409
(410) 974-0947

1 that type of thing going on between industry and the
2 National Institute of Health, through the Foundation
3 for the National Institute of Health, for an
4 intermediary, and that's in the area of chest and
5 chest CT, a program for truthing, sharing images and
6 truthing. So not that that is a prediction of the
7 future, but where there's a benefit and there's a
8 nice structured arrangement there, there's some
9 record of that happening.

10 DR. D'ORSI: Yes, Dr. Bourland.

11 DR. BOURLAND: A question for MITA again on
12 algorithm and proprietary information like that. So
13 the question is, is there a comfort level with
14 stating, for instance, the type of algorithm or in
15 general, for instance, that might be, for instance, a
16 user guide so a user would have a handle? Because
17 sometimes this does have no limits of a confidence
18 interval for the user, for instance. And so is there
19 a balance that would be acceptable to the industry
20 group?

21 MR. MARSHALL: Yes. This is Julian
22 Marshall. And already in our submissions, at least
23 with our PMA device, we provide block diagrams and
24 information about the building blocks within. That's
25 not the issue. It's the minutiae of what's being

1 requested that was the objection that we made.

2 DR. D'ORSI: Thank you. We're going to go
3 right to the questions at this point, and I want to
4 turn over the podium, the speaker, to Janine Morris,
5 who's going to describe a little different method
6 that we're going to use, and I'll let her describe
7 what's being done.

8 MS. MORRIS: Okay, thank you very much.
9 This is Janine Morris with FDA. Because of the
10 complexity of the questions and the amount of time we
11 have today, I made an attempt to try to condense
12 things, and I want to provide you an overview of what
13 these questions are trying to elucidate with respect
14 to these guidance documents.

15 My first comment is I want to bring the
16 attention to the Panel as well as the audience that
17 these guidance documents are under review for
18 comment, and we would certainly seek everyone's
19 comments to be submitted to the FDA so that we can
20 continue this discussion and make sure that we
21 develop guidance documents that are useful to the
22 industry, to FDA review staff, and get the products
23 that are needed by the clinicians out onto the market
24 with a reasonable assurance of safety and
25 effectiveness.

1 As you'll see, if you've looked through
2 these guidance documents, they're a significant
3 departure from the past review practices of CAD
4 devices. CAD was an evolving technology that kind of
5 sped up before really the Agency was able to
6 understand the impact. So it should be recognized
7 that it is a departure, and it is that we're seeking
8 Panel input in terms of the adequacy of the raising
9 the bar.

10 And so these questions are going to be
11 focusing on our bigger issues with respect to that,
12 and it's focusing on the clinical requirements. If
13 there are issues after the comment period about other
14 sections of the guidance, certainly we would be
15 addressing those internally. But this is where we
16 feel are going to be the greatest areas of concern.

17 So the questions that have been outlined
18 can be broken up into three parts. So Questions 1
19 through 5 are really dealing with the pre-market
20 requirements for CAD devices. I will be going
21 through and kind of condensing those and combining
22 different parts to try to help facilitate that.

23 The second part are post-market
24 considerations. Once we determine for, you know,
25 substantial equivalence as well as a reasonable

1 assurance of safety and effectiveness, sometimes
2 there are additional questions that need to be
3 answered in a post-market setting, and we want to
4 raise this to your awareness and whether or not that
5 should be considered as well. And those are
6 Questions 6, 7, and 9.

7 And then, finally, there's a section that I
8 consider future considerations because they're very
9 complex issues, and we're basically seeking the input
10 from the Panel of whether we should further explore
11 this and that the Agency would try to find -- put
12 into a process of considering this. And these issues
13 really cover Question 8 as well as, I believe, 5(c).
14 That's dealing with the second readers and the actual
15 user performance of the CADs. So those are the basic
16 decision points in these questions.

17 The first question is what I want to start
18 with and get a general discussion. So if Robert
19 could put up Question 1. So we're going to have a
20 discussion of a general nature about the guidance
21 documents, and we're going to be asking you to give,
22 you know, are there big areas in which we've missed
23 or that we're off the mark? This is to give you an
24 opportunity to give your high-level, important
25 questions or deficiencies about the guidance.

1 After Question 1, it gets a little more
2 complicated. So I'm just going to give you an
3 overview of how I've condensed things. So what are
4 FDA's big questions to the Panel with respect to
5 these guidance documents? Well, the first one is, is
6 I'm going to take the presumption of condensing what
7 these two guidances are saying, that clinical data
8 will likely be needed for all new CAD devices as well
9 as modifications that we expect would impact reader
10 performance.

11 So earlier there was a discussion about, if
12 there are minor modifications to a CAD device that's
13 under a 510(k), it's a Class II device and there are
14 minor modifications that do not impact reader
15 performance, then there is an option for industry to
16 decide that they do not need to submit a 510(k)
17 because it fits the paradigm of minor modifications.
18 So anything beyond that, that would impact reader
19 performance, the guidance is suggesting that clinical
20 performance data would be necessary to assess a new
21 CAD and to assess any modifications to a CAD that
22 warranted a higher, you know, level of risk, of
23 impacting user performance.

24 So, essentially, what we're asking you as
25 the Panel is do you agree with this paradigm? Do you

1 agree that for these type of devices, whether they're
2 Class II or Class III, that this multi-center, multi-
3 reader study approach is necessary for the risk and
4 benefit of the device? So if we all agree with that,
5 then the way the guidance is currently written can
6 proceed.

7 If we don't agree with that, then it gets
8 into several questions that we've flushed out. So
9 Questions 2(b) and 5(a) talk about when a reader
10 study would not be necessary with respect to changes.
11 So if we need to go to those questions to discuss an
12 alternative, we can refer to those and discuss it
13 further.

14 The next big area is, when we are asking
15 for clinical data, when can we use it for
16 generalizability? So that is, when can we take a
17 clinical study and generalize across subgroups? So
18 this could be a generalize across for a mammo CAD,
19 microcalcifications and masses, or for lung CAD, the
20 different type of lesions. It also would be because
21 of the advancement and the use of CAD, you know, that
22 for CT and FFDM, that there are all of these existing
23 products on the market in which CAD can be used with.
24 So when is it appropriate, when clinical studies are
25 done, that the data can be generalizable across those

1 devices?

2 So we have a new colon CAD used for CT, and
3 they come in and they want to present a clinical
4 study. Is it generalizable across all CT systems?
5 Or is it necessary to have it evaluated on each CT on
6 its own? The same for FFDM. It's going to be a
7 question of what can be considered generalizable in
8 the clinical -- from the clinical study.

9 So that is a very big question because
10 we're changing our direction, and we have to consider
11 what's already on the market and what is going to
12 come on the market in the future. Now, these cover
13 Questions 3(c), 3(d), and 4(a) through (d) and 5(b).
14 And if we need to get into more specifics and
15 examples, we can refer to those questions.

16 The next area is we agree on situations
17 where we feel that we need a clinical study. There
18 are certain areas where we need further
19 clarification. Those areas are reuse of data, which
20 is covered under 3(a); which is an appropriate
21 control to use, which is under 2(a); the endpoints
22 that would be considered under the study, which ones
23 are the important ones to power the study on? Are
24 you going to power it on the ROC curve, or are you
25 going to power it on sensitivity/specificity, or are

1 you going to power it on all of them, or they have
2 equal weight? That's being addressed in 3(c).

3 Finally, it's a question of reader
4 characteristics, how many readers, the
5 characteristics of those readers and the proportions.
6 And if we're able to get that, that would be
7 addressed under 3(b). So basically these are our
8 important clinical considerations for these studies
9 because that is what's going to dictate how large and
10 complex these studies might be.

11 Finally, as we said earlier, that there are
12 some post-market considerations that we want to
13 discuss in general, and that's covered under
14 Questions 6, 7, and 9. And then the future
15 considerations, that is where we want just to a
16 heads-up of whether or not we need to further examine
17 this issue of classification for these devices and
18 the issue of whether or not the Agency needs to
19 address the actual use of the devices. And that
20 alone could involve several Panel meetings, but we
21 really are looking for a go/no go of this is an
22 important issue that needs to be addressed by the
23 Agency.

24 Okay. So I'll start with Question 1, and
25 I'll have Robert read Question 1, and then,

Free State Reporting, Inc.
1378 Cape Saint Claire Road
Annapolis, MD 21409
(410) 974-0947

1 Dr. D'Orsi, if you could just solicit input from the
2 Panel.

3 DR. OCHS: So, again, just to summarize,
4 earlier we heard presentations on the following
5 viewpoints: with respect to pre-market requirements
6 for CAD devices, such as imaging issues, statistical
7 issues, clinical issues, and post-approval
8 considerations.

9 Again, the following questions are intended
10 to elicit your thoughts and recommendations on the
11 scientific and clinical content described in the CAD
12 guidance documents. And your clinical and scientific
13 input will assist the Agency in determining how these
14 guidance documents should be applied in the pre-
15 market review of CAD devices, whether that be in
16 determining substantial equivalence for Class II CAD
17 devices or demonstrating a reasonable assurance of
18 safety and effectiveness for Class III CAD devices.

19 So our first question today is:
20 Considering the input provided in the March 2008
21 Panel meeting, has the Agency adequately addressed in
22 these two draft documents the major points of
23 discussion and recommendations for a pre-market
24 review, such as comparison of the device description,
25 device standalone performance testing, clinical

1 performance testing, and labeling? And please
2 describe any areas of concern that should be
3 clarified. Identify and describe areas that should
4 be modified, removed, or added, and provide your
5 rationale for these changes.

6 MS. MORRIS: So I'm just going to interrupt
7 here, and as an example from Question 5(e), it
8 states, Do you believe that the draft guidance
9 documents adequately explain the clinical meaning of
10 area under the ROC curve? Do you believe that the
11 draft guidances adequately reflect the use of
12 alternative metrics? So that's just an example for
13 you on how you might provide feedback in general
14 about the guidance areas where we need to make
15 improvement, clarification, where we've made
16 omissions that are important to you.

17 DR. D'ORSI: Thank you. Again, remember,
18 this is an overview, and we'll get into all the
19 detail with the further questions. So does anybody
20 have in their perusal any gross omissions or cloudy
21 issues that they see in these two guidance documents,
22 the two draft guidance documents? Yes, Dr. Leitch.

23 DR. LEITCH: I think this is a question
24 that's come up earlier. When you're looking at the
25 PMA versus the 510(k), the designation of the device

1 is into which category, the separation of breast
2 versus colon and lung, sort of remains unclear to me,
3 and I think it is unclear to the manufacturers as
4 well. And so I think that's -- it's not clear from
5 these documents how that is decided.

6 MS. MORRIS: Okay. So the guidance
7 document. We currently have CAD devices that are
8 Class II and reviewed under 510(k). And so one of
9 the guidance documents that references the
10 performance for 510(k) devices just addresses that,
11 and it's the ones that we currently recognize as
12 remaining in Class II, and within that guidance it
13 talks about when clinical performance is needed. The
14 clinical guidance document is for any device that the
15 Agency determines needs clinical performance. So it
16 can be a Class III device and a Class II device.

17 Now, this doesn't exactly address your
18 question because what I want to focus on is your
19 reaction to the actual guidance documents, not
20 necessarily how we're going to regulate or how should
21 we regulate all CAD devices or CADe devices, whether
22 they be in Class II or Class III, because that's a
23 much larger discussion. But if that is an important
24 one, we will proceed and do that in a future Panel
25 meeting. But for now, the majority of colon CAD and

1 lung CAD devices are Class II, unless someone comes
2 in with something that has such a different
3 technological characteristic that we would find it
4 not substantially equivalent or there is a new
5 intended use, that they go from a CADe to a CADx
6 because now they're looking at that. And then the
7 mammography CAD still remains in Class III. So if I
8 haven't answered your question, try repeating.

9 DR. LEITCH: I mean, I guess when you're
10 trying to figure out the magnitude of the studies
11 that are required, that matters to some degree. If
12 you always have breast in III, you know, and you
13 always have colon and lung in II, you know, the way
14 you think about the trials that are necessary, it
15 seems like that would be different. The magnitude of
16 proof that's required.

17 MS. MORRIS: Can you elaborate on why they
18 might be different?

19 DR. LEITCH: Because you make them
20 different. You say the requirements are different.

21 MS. MORRIS: Okay. So for Class III
22 devices, we're trying to establish a reasonable
23 assurance of safety and effectiveness. So
24 mammography CAD were the first ones that I remember
25 that we recognized as being CAD devices. And so

1 that's what led it into a Class III device. For
2 colon and lung, it was an evolution over PAC systems,
3 archiving systems, and it's crossed a line. If the
4 Panel believes that the risks associated with those
5 devices are the same as the risks associated with
6 mammo CAD, then we should consider moving those to a
7 Class III category.

8 (Laughter.)

9 DR. LEITCH: I don't think that was the way
10 they wanted them to go.

11 MS. MORRIS: Well, yeah, I said that to be
12 more provocative than anything. But currently we're
13 trying to develop a guidance that would be applied to
14 what we currently have in-house with FDA, meaning
15 what we have cleared under Class II and what we have
16 approved under Class III and trying to provide
17 industry with some guidelines on how to move forward.
18 What is the level of evidence we would need for these
19 devices? And so far, I am not seeing a big
20 distinction between the two groups. And if that is
21 not the correct direction to take, then we need to
22 hear from the Panel about that.

23 DR. D'ORSI: Dr. Glassman.

24 DR. GLASSMAN: I would, I think, echo the
25 fact that the distinction between breast and other

1 organ system CADs is regulatory and artificial and I
2 think is an impediment to lightening the load, at
3 least on the breast CAD companies, in terms of
4 submissions, because unless I've got it wrong, you
5 can't use equivalence to a predicate for a Class III
6 device.

7 And, therefore, for many instances where I
8 think we want to lessen the burden, the Class II
9 process seems to me reasonable for something which is
10 fairly well understood, fairly well researched, and
11 is -- you meant to be provocative, and I guess I will
12 too. I think breast belongs in Class II and the
13 sooner that the Agency can get it there, the sooner
14 the regulatory burden on the breast CAD manufacturers
15 will decrease.

16 MS. MORRIS: And I certainly appreciate
17 these comments because these are the issues that
18 we're struggling with internally. But that is our
19 latter question, the question at the end, because
20 right now the urgency is what do we do this minute?
21 How do we regulate the devices as they're classified
22 now? What level of evidence do we need so that
23 industry can move forward and so that you have
24 devices that are available to you? So that's what
25 the purpose of these guidance documents are. If we

1 want to discuss changing classifications, whether
2 it's going up in classification or down in
3 classification, the Agency will explore that, if that
4 is important. But for today's meeting, we only want
5 to focus on now so that we can move forward and that
6 we're not stalemated.

7 DR. D'ORSI: Yeah, Dr. Zhou.

8 DR. ZHOU: So in regard to the document, I
9 have two comments. About the reused test data, I
10 feel like right now there's no sound scientific
11 method to analyze reused data, so that you should
12 strike out from the document, because you should tell
13 that clearly there is no valid way to analyze that
14 kind of data. I don't feel like that is made clear
15 in the document.

16 The second point I want to make is, since
17 you asked about the ROC, I think the interpretation
18 of the AUC may not be totally correct. I think
19 currently they say the AUC is the major for
20 separation between disease and non-disease
21 distribution. Actually they are not. ROC use is not
22 separating of distributions. So that probably needs
23 to also make that clear.

24 DR. D'ORSI: Mr. Uzenoff.

25 MR. UZENOFF: Thank you. As Industry

Free State Reporting, Inc.
1378 Cape Saint Claire Road
Annapolis, MD 21409
(410) 974-0947

1 Representative, I would just like to recall, you're
2 asking for a high level of comments on the guidance,
3 and there would be five of them that I would like to
4 mention, and they're in a document that's been
5 submitted for the docket and distributed in your
6 packets this morning. But they are the requirement
7 for multiple scans of patients, reuse of data, and we
8 heard from one of the industry speakers about some
9 good and interesting news about possibilities on that
10 today from Dr. Petrick, concern about disclosure of
11 trade secrets in the pre-market requirements, the
12 labeling intended use situation that Mr. Marshall
13 talked about, and then, finally, in the area of
14 clarity, the distinction between the 510(k) and the
15 PMA requirements. And they're detailed in -- they're
16 in further detail in the letter in front of you.

17 MS. MORRIS: Thank you.

18 DR. D'ORSI: I think we'll get to some of
19 those in the discussions. Yeah, Dr. Jiang.

20 DR. JIANG: Yeah, I'd like to get a
21 clarification. I heard it said that these documents
22 are nonbinding, whereas the FFDM documents were
23 binding, from yesterday. I wonder what that exactly
24 means. I get a sense that FDA does require clinical
25 studies in many of the situations that have been

1 spoken to. So can we get a clarification? Are we
2 really requiring companies to do these things, or
3 this is really not binding, just suggestions?

4 MS. MORRIS: So what's meant for general
5 guidance, when we state that they're not binding, is
6 it's the Agency's attempt to state our expectations
7 for us to make a decision. Our mandate is to make a
8 decision as to whether or not a Class II device is
9 substantially equivalent, you know, to a legally
10 marketed Class II device of the same type. Or, for a
11 Class III device, whether or not there's a reasonable
12 assurance of safety and effectiveness. We have to
13 make that decision and we need data to make that
14 decision. We're driven by what data industry
15 provides to us.

16 And so there is a level in which we get to
17 a point of what is adequate to make those decisions.
18 So we would put together a construct of what we think
19 would be necessary. So the 510(k) guidance lists out
20 what we feel would be necessary for a company to
21 demonstrate substantial equivalence to another -- of
22 their device to a legally marketed device. And
23 that's what our current thinking is.

24 But if the industry would come in with a
25 completely different paradigm or another way of

1 assessing it, then it's incumbent upon the Agency to
2 consider that. And if it addresses our questions for
3 safety and effectiveness and substantial equivalence,
4 then we would adopt that. It's the same concept that
5 I raised yesterday with standards. Just because the
6 FDA doesn't recognize a standard, it doesn't mean
7 that the standard isn't valuable and can assist us in
8 making that decision. So that's what's considered
9 nonbinding. Does that answer your question?

10 DR. JIANG: Yes, thank you.

11 DR. D'ORSI: Okay, just as an aside, it
12 almost may be easier for a manufacturer to
13 demonstrate standalone than a comparative study with
14 CAD. It's just my own thinking. Let's go on now
15 with the basic questions. Do you want to read the
16 first one, Janine?

17 MS. MORRIS: Okay. So as I stated before,
18 one of our big questions -- and I'm just going to put
19 it out to the Panel to basically agree or disagree
20 that when industry wants to submit or get marketing
21 clearance or approval for a new CAD device or a
22 modification to their existing CAD device that we
23 would expect would impact reader performance, a
24 reader study would be necessary.

25 So considering what's out on the market,

Free State Reporting, Inc.
1378 Cape Saint Claire Road
Annapolis, MD 21409
(410) 974-0947

1 considering what the relative risk/benefit of these
2 devices are, do you agree with this concept? Because
3 that's what the guidances are essentially stating.
4 If you don't agree with this concept, then we should
5 explore further of when the clinical data would not
6 be needed. And I can cover that with a couple of
7 questions that we've drafted. So it would be nice
8 just to go around the table and find out whether
9 there is agreement or not.

10 DR. D'ORSI: Let's just start down the
11 table with the -- this is basically a question about
12 the need for clinical studies in the scenario that's
13 covered in this question. So why don't we just get
14 started with that, and let's go down and we'll start
15 with Mr. Uzenoff.

16 MR. UZENOFF: I think, yes, there would be
17 situations where there would not be the need for
18 reader studies.

19 DR. DUEHRING: I also agree that it's not
20 pertinent in every situation to have a reader study.

21 DR. CARRINO: Yes, I also agree, within the
22 context of CADe alone, that standalone evaluation
23 could suffice.

24 DR. D'ORSI: Let me make one clarification
25 here. The thinking behind this is if a manufacturer

1 brings their device to the FDA, that almost by
2 definition it's not a minor change. So that's kind
3 of the thinking that's going on underneath this
4 question. Do you want to restart, then, with that
5 little tidbit? No, okay.

6 DR. CARRINO: I think what we're getting to
7 is we're really separating CADE from CADx, and CADx
8 is that the user has to be involved in CADx. CADE is
9 an assistant tool and to validating the -- I don't
10 necessarily need to have the reader involved because
11 my variability -- our variability as a tribe of
12 radiologists could be much greater than what's
13 introduced by the system.

14 So I think being able to have a tool that's
15 available to me as a second opinion could be done
16 with some of the standalone information. And I think
17 the discussion from this morning and this afternoon
18 balanced each other and really brought together those
19 points and helped us refocus our thinking.

20 DR. KIM: I would agree with John, as well.
21 I think especially, certainly when it's outlined in
22 the FDA guidance documents, this sets the bar and
23 it's a very ideal bar. But when you weigh it against
24 what's least burdensome and what is practical, I
25 think you have to consider other things.

1 And one of the questions I have
2 specifically for CAD colonography is that reader
3 performance studies with a high number of readers and
4 a high number of cases, a large number of cases, is
5 very difficult because each exam takes anywhere from
6 10 to 20 minutes. And so I can't foresee having
7 large enough numbers to account for reader variation.
8 And so that diminishes the impact of reader
9 performance studies because you don't know how much
10 of it is related to just random variation. And so if
11 that's the case, and if that's one of the reasons
12 really impacting industry, I think it's something we
13 need to look at closely and realize that it's not
14 perfect, but does it fulfill the level that the FDA
15 would need?

16 DR. BOURLAND: I agree with the previous
17 comments because in terms of these documents being
18 guidance, then there is the opportunity for the
19 vendor to bring forth an argument that states and
20 proves the functionality of the device and the
21 fidelity of that.

22 DR. SEIBERT: I also concur with respect to
23 what has been previously said. A tiered approach, I
24 like that. Certainly, sometimes you have to go back
25 with a 510(k) for -- just because there's some

1 modifications that could have an impact but they're
2 not that level that would require a reader study. So
3 that's my opinion.

4 DR. LEITCH: I agree, there are
5 circumstances when standalone is fine, and I think
6 that this separating exactly what we're talking
7 about, if it's detection versus diagnosis is key and
8 it has to be separated in the minds of the clinicians
9 when they use the device. If there is a claim that
10 is very much reader-impacted on the labeling, then
11 you have to have a reader study that would validate
12 that.

13 DR. LIN: I also agree. My understanding
14 is that these CAD devices are now being marketed as
15 second reader devices, and in this situation, I think
16 we don't necessarily need to have clinical
17 assessments for every case.

18 DR. ABBEY: So I think that the FDA needs
19 to have the ability to request observer performance
20 studies when they're necessary and I don't -- but I
21 don't think they should necessarily be mandatory
22 either. There are a lot of -- it really depends on
23 the nature of what's changing. And so if you can
24 have the spectrum, I think that's the most
25 advantageous way. And then guidance has to be clear

1 on what trips, the various steps in that. But so not
2 necessary, but maybe, but perhaps.

3 DR. ROSENBERG: Yeah, I also like the idea
4 of the tiered approach, and has been stated here and
5 we all know, there's a large variation in radiologist
6 readers in the population of radiologists. So if you
7 really wanted a valid reader study, it would be a
8 very complicated thing to demand.

9 DR. ZISKIN: Certainly there are times when
10 the reader studies are not necessary.

11 DR. MITTAL: I agree. I don't think reader
12 studies are necessary in every situation. However, I
13 will suggest to the Agency that they come up with a
14 strict definition of what is a minor modification.

15 DR. D'ORSI: That's us.

16 (Laughter.)

17 DR. PAYNE: Well stated there. I will
18 concur with the previous comments. I think, with
19 regard to the CADe devices, I think, because they're
20 just -- I guess I think of them -- and I'm a medical
21 physicist, of course, and I don't have to worry about
22 any of this, but at least as a clinician, but they're
23 primarily -- I look at them as -- my perception is
24 they're marking devices and as such, I think,
25 requiring more stringent testing when all you're

1 looking for is marking is not necessary. Thank you.

2 DR. ZHOU: I think it will depend. I mean,
3 if the device has enough change there in their
4 program or in their system so that they impact the
5 diagnostic or detection accuracy and also the change
6 in accuracy by the readers, I feel like if that's the
7 case, then they need to have a clinical study. But I
8 want to point out one thing even for the CADe: this
9 device does not act alone. You need to have a
10 radiologist to read it. I mean, that was a question
11 in the morning, about to say if we're able to
12 actually demonstrate that, then we don't need the
13 radiologist. But that's probably not a good idea.
14 So if you're able to tell the patient, to say what is
15 the diagnostic or detection accuracy of CADe, there's
16 a human factor in it. So if you don't do the reader
17 study, I feel like you might miss that part of the
18 information which might be important.

19 DR. DODD: So I'll agree that there are
20 circumstances that we would not need a reader study.
21 But along the lines of Craig's statement, I think we
22 need to allow for there being a requirement for
23 reader studies in some cases.

24 DR. GLASSMAN: First, I'd like to say that
25 the Chairman of the Panel of March 4th and 5th, 2008

1 should be executed.

2 (Laughter.)

3 DR. GLASSMAN: At that time we unanimously
4 agreed that the reader studies were necessary in all
5 instances, and I think that everybody must remember
6 that the Agency has taken that to heart and now we're
7 changing. Now, having said that, I would agree that
8 reader studies are not necessary in every instance,
9 and in fact, other than the Class III instances,
10 there may only be very few times when they're really
11 necessary.

12 DR. TOURASSI: I agree with the rest of
13 Panel. I believe that when a new or modified device
14 is presented, if it has the same intended use, the
15 same labeling, and it conveys the second opinion in a
16 similar matter as the predicate device, then
17 comparison of standalone performance should be
18 sufficient.

19 DR. JIANG: I think, across the board, a
20 requirement of reader studies is not only not
21 necessary, it's also not warranted. I think that's a
22 way to learn how CADe works. As we gain more
23 experience about that, there will be less and less
24 instances where we require readers studies, and I
25 think we should use them judiciously. We should use

1 it when there is specific concerns that the reader
2 performance might be affected. I can think of many
3 instances you would expect the reader performance not
4 to be affected. So I agree with the rest of the
5 group, that it should not be required across the
6 board.

7 DR. SWERDLOW: I also agree. Also,
8 hopefully, further over time, as the datasets for the
9 standalone become more and more robust, the need for
10 a clinical study would even further diminish so that
11 over time it may become a relatively moot issue.

12 DR. STEIER: It's not that I don't trust
13 radiologists, but I kind of read my studies, CAT
14 scans, on my patients, anyway. So I guess that makes
15 me my own second reader. As for Dr. Glassman's
16 comment, how surprising for a group of radiology
17 people to change their mind about something.

18 (Laughter.)

19 DR. D'ORSI: No comment.

20 DR. STEIER: I agree that reader studies
21 would not always be necessary. Thank you.

22 DR. D'ORSI: So if I'm sensing the group,
23 the Panel here, it seems like a sort of almost
24 unanimous vote that clinical data will not likely be
25 needed for all new CAD devices as well as

1 modifications.

2 Having that, let's read Question 2(b) and
3 we'll go into some issues where we can separate when
4 and when not.

5 DR. OCHS: I'll first read the big overview
6 for Question 2. Then I'll go to Question 2(b). And
7 there's a lot of examples given in Question 2(b), but
8 I'll just read Question 2(b) so you have more time
9 for discussion.

10 Under Section 6 of the 510(k) draft
11 guidance, the Agency states that a clinical
12 performance assessment will usually be necessary to
13 demonstrate substantial equivalence to a predicate
14 CADe device. A clinical performance assessment is
15 expected for all original PMAs. The clinical
16 performance guidance was developed to provide
17 recommendations for designing a reader study to
18 support either a 510(k) or a PMA.

19 And then we have a question on control
20 arms, which is Question 2(a), but I'll go to Question
21 2(b). So standalone performance.

22 Please describe under what conditions the
23 Agency should consider accepting standalone
24 performance in lieu of clinical performance data for
25 a CADe device. And then as a means for discussion,

Free State Reporting, Inc.
1378 Cape Saint Claire Road
Annapolis, MD 21409
(410) 974-0947

1 we have provided the following examples.

2 And then they range from, you know,
3 manufacturers and different changes and such. Do you
4 want me to read every single example or --

5 MS. MORRIS: I will make a stab at giving
6 you a high-level picture of each of these examples.
7 And then if we need to get into a little bit more
8 detail, we can read them through. But I'll try to
9 capture the gist of it.

10 So the first example is a manufacturer of a
11 new CAD device. And it was discussed earlier, I
12 think, by one of the industry reps in the open
13 session. But this is where you're looking at the
14 standalone performance. The Agency is looking at the
15 standalone performance and the new device, against
16 the predicate device, usually their own, is using the
17 same database of cases, using the same truthing, the
18 same scoring methods, but the abnormalities are
19 coming out a little bit different, meaning that we're
20 picking up additional abnormalities, but different
21 from their predicate device.

22 So (i) and (ii) are similar. The first
23 case is you're using the same database, the same
24 truthing, the same scoring, and if these false
25 positive marks have a different pattern, do we need

1 to go to clinical performance testing? The second
2 part is using a different database of cases, with
3 different truthing and different scoring
4 methodologies. So it's a slight nuance. So taking
5 that first part where you're looking at the outcomes
6 with different false positive.

7 DR. D'ORSI: Let's open that up.

8 Dr. Glassman.

9 DR. GLASSMAN: As long as they were
10 different false positives and not different true
11 positives, I think that would be acceptable. In the
12 days of film-screen mammography and CAD, when we
13 would run the films through the CAD device several
14 times, we would get different markings for the same
15 film and the same device. So there's a certain
16 amount of normal variation. As long as the new
17 device did not miss true positives, I would not be
18 worried by that difference.

19 MS. MORRIS: Does it matter whether or not
20 it's used when you're comparing the same truthing
21 versus different truthing or the same databases
22 versus different databases?

23 DR. GLASSMAN: It wouldn't matter to me.

24 DR. D'ORSI: What you've described here is
25 actually an improvement. If you look at it, it's not

1 knocking out any true positives, but it's dropping
2 false positives. So that's sort of a no-brainer
3 there, to me, anyway. How do people feel about that?
4 If I read this correctly, if the device that they're
5 testing drops some findings out, but no true positive
6 findings, and drops false positives, that's a gain.
7 So that should be no problem.

8 DR. ZHOU: So they have the false positive.
9 But you also put the burden on the patient, right?
10 Because the patient don't have -- let's say suppose
11 that's the case in the clinical setting. So the
12 patient don't have a cancer and they say, yes, they
13 have a cancer, that puts -- does that put burden on
14 the patient?

15 DR. D'ORSI: Well, remember, this is a
16 detection device, so the only burden it would put on
17 with mammography, anyway, it may be different with
18 chest, is a burden of an exam to clarify the validity
19 of that finding, and that would be it. So when it's
20 not found, there is nothing to deal with.

21 DR. ZHOU: An interesting question would be
22 actually is that false positive mistake could
23 translate into false positive in the patient part of
24 it because here it is standalone. They look at the
25 detection part of it. So if there's a link there,

1 they might have some problem. But like you say, even
2 do they have false positive for detection part, but
3 when you get to the patient part, it doesn't matter.
4 Then I don't --

5 DR. D'ORSI: I mean, what you say is
6 strictly correct that, you know, there may be a
7 downside. But in this scenario, this should be, my
8 God, I would jump for joy if I got this data on a
9 CAD. A new product, anyway. Yes, Dr. Tourassi.

10 DR. TOURASSI: I think the critical issue
11 between the two scenarios was the database, if the
12 comparison is on the same set of cases or not. So
13 for scenario number one, clear cut. It's obvious to
14 tell the difference. For scenario number two, it's
15 not. Because if the comparison is done on two
16 different datasets for the selection of the cases, in
17 one case it was simple, and the makeup of the
18 different manifestations of breast cancer, let's say,
19 is very different from one set to another, then
20 standalone comparison performance comparison wouldn't
21 mean as much. So for scenario number two, there
22 needs to be some control mechanism to ensure that
23 these two different databases have the same level of
24 difficulty, the same type of proportion, at least the
25 same type of different signs of disease and so forth.

1 MS. MORRIS: So for clarification, can I
2 just get clarification --

3 DR. D'ORSI: Sure.

4 MS. MORRIS: -- on that comment? So they
5 could not necessarily be the identical, but they
6 could be comparable?

7 DR. TOURASSI: Correct. For example, for
8 breast cancer, we know that the CAD devices do so
9 much better with calcifications than with masses. So
10 if the comparison is done on a database that is
11 overwhelmingly full of calcifications for the new
12 device, then we're biasing favorably. So we have to
13 be sure about this.

14 DR. D'ORSI: And I think this is what
15 Dr. Giger was aiming at with a separate, usable
16 database for all standalone CADs. That would be the
17 ideal because there's no cognitive thinking on a CAD
18 device. So if you shove the same cases in, with the
19 proviso that no one can really get a hold of these to
20 train their new device to these cases, that's the
21 ideal. That's great. And the only reason that you
22 are changing is because the validity of training to
23 that dataset can't be regulated. But the ideal is an
24 identical dataset for the standalone. How do people
25 feel about that? Yeah, Dr. Bourland.

1 DR. BOURLAND: Well, the other part of the
2 question, then, also is the analysis portion and
3 whether that's the same or not. And so the question
4 is, is there a relationship between the two
5 databases, for instance, statistically speaking, in
6 terms of the distribution of diseases and the types
7 of cases?

8 And then also analysis. There are
9 different ways to do analyses. One can just be as
10 valid as the second. Perhaps there's a relationship
11 between them that needs to be demonstrated. If it's
12 an inappropriate analysis, then the answer would be
13 no. You can be wrong.

14 DR. D'ORSI: Yes, Dr. Dodd.

15 DR. DODD: Can I just get a little
16 clarification on what types of different truthing
17 we're talking about? Truthing is pretty critical on
18 my mind. Are we talking about maybe one case where
19 we had biopsy verified cancers versus, you know,
20 trying to repeat the detections that a panel of
21 expert radiologists got? I mean, that might be a
22 case where there would be different levels of
23 concern.

24 MS. MORRIS: Can I flip it around in terms
25 of, you know, if we've accepted a certain method of

1 truthing in the past and the industry's able to
2 demonstrate that it's a comparable truthing, you
3 know, the same type, that that would be acceptable?
4 Or can you identify when it wouldn't be the same,
5 that we would have to reexamine?

6 DR. DODD: I guess my example I just gave
7 might be one where I would feel concerned. You know,
8 if you're comparing a panel of expert radiologists as
9 one truth versus biopsy verified cases.

10 MS. MORRIS: So to put it different, that
11 isn't as significant, that you have a different
12 dataset and you have a different set of readers, but
13 they're still comparable to the original dataset and
14 the original group of readers.

15 DR. DODD: Sure. I think at that point you
16 have to begin to incorporate some of the uncertainty
17 with the truthing into your analysis.

18 MS. MORRIS: Okay.

19 DR. D'ORSI: Dr. Rosenberg.

20 DR. ROSENBERG: Yeah. No, I think the
21 comparability is important, but I think we also want
22 to make sure, as now we've gone from film-screen to
23 digital mammography, that the databases that are
24 being used are relevant to the cases that are
25 actually being -- the CAD is being applied to. So I

1 don't think we want to put too large a barrier to
2 having updated data that makes the algorithms valid
3 for the cases they're being used for. So I think we
4 have to be careful about the datasets being relevant
5 to the current populations. I guess that's not
6 really an answer.

7 DR. D'ORSI: Well, it's a good statement.
8 Dr. Jiang.

9 DR. JIANG: Yeah, I just want to follow up
10 on that. I think that's very important. Dr. Petrick
11 was talking about progressively changing part of the
12 dataset, deleting some and adding some more. I think
13 that's very important as the image quality or what we
14 understand as the image itself progresses. For
15 example, earlier days, you know, we wouldn't use
16 1970s mammograms today.

17 DR. ABBEY: But in that light, so now
18 you're going to get a comparison with, say, 20
19 percent different images. You're not going to get --
20 you're going to get disagreement. You're not going
21 to know, if you would've got agreement, if the cases
22 had all been the same. So it raises a problem, I
23 think, for the FDA to say, well, what's the standard
24 by which we say no, this is disagreement, or what is
25 this agreement? So at least maybe you then say,

Free State Reporting, Inc.
1378 Cape Saint Claire Road
Annapolis, MD 21409
(410) 974-0947

1 okay, whatever portion are the same, you have to
2 demonstrate agreement there. But at least on Part
3 (i) here, my understanding was is that some of the
4 abnormals are being missed and new ones are being
5 gained. So it didn't seem to meet your criteria,
6 Dr. Glassman's criteria, that you don't miss any one
7 that you got before. Now, you're missing some that
8 you got before, but you're getting ones that you
9 didn't.

10 MS. MORRIS: Can I just clarify?

11 DR. D'ORSI: Go ahead, Janine.

12 MS. MORRIS: If you read the actual part,
13 you're right, it's a little bit different scenario.
14 So I was trying to stage it. So the first stage is
15 you're getting all the truths, but there may be a
16 different mix of false positives. And the next stage
17 I'm going to take to your level of different truth.

18 DR. ZISKIN: I have a general question.

19 DR. D'ORSI: Yes, I'm sorry.

20 DR. ZISKIN: I have sort of a general
21 question. I would be better able to answer questions
22 of when clinical studies are necessary if I
23 understood better what's the downside? What are we
24 trying to protect? In other words, if all the CADE
25 just is input into the radiologist, well, has bad

1 information been given to him where we know that his
2 performance has been decreased because of that? And
3 if I knew more about those situations, then I'd say,
4 okay, we've got to be very careful in those
5 situations to really be sure of ourselves. But in
6 the other aspects, when the performance of the
7 radiologist is not decreased at all, I don't see what
8 the downside is on this. And so at this moment,
9 unless I heard more on that side, I would say, as
10 long as the physical measurements on a standalone
11 thing are equivalent, at least, I don't have any
12 problem with not having any requirement for the
13 clinical studies.

14 MS. MORRIS: The operative word is
15 equivalent, though, and it has to do with assessing
16 the standalone performance and if the differences --
17 you're going to have differences. What is the
18 clinical impact of those differences? Does it
19 require us to go to the next level, having an MPMC
20 study?

21 DR. D'ORSI: If I can take a stab at
22 just -- first of all, it was very difficult. Usually
23 what happens, whatever you call performance, how well
24 you pick out disease and how well you leave normal
25 women alone, let's say that's your performance level,

1 usually there's almost always a tradeoff. Usually
2 sensitivity goes up, specificity goes down, and vice
3 versa. Any test that raises both of them is superb.
4 But all the articles written, that's the tradeoff.
5 So the thing is what are you using this CAD for? And
6 you don't want what? You don't want false negatives
7 in most of these, and I assume in chest as well. So
8 if you increase your false positives, what's going to
9 happen to your false negatives is they're going to go
10 down. If you decrease your false positives, your
11 false negatives will generally go up. So the price
12 you pay for subtle detection and the lack of false
13 negatives is the false positive. So you have to be
14 very careful how you balance that.

15 DR. ZISKIN: I just presume the way that
16 you were speaking, you were speaking about the
17 radiologist and his balancing of positive and
18 negative, and I was referring to the CADE. And if
19 the CADE is giving you information, let's say it's
20 good information or bad information, will it ever
21 decrease your performance?

22 DR. D'ORSI: It could. For example, let's
23 say the manufacturer says, okay, I'm going to throw
24 on 17 more marks. This is hyperbole. But I'm going
25 to pick up one cancer in that. So now the reader has

1 17 marks to look at for one malignancy. So it's a
2 value judgment. Are those 17 marks going to dissuade
3 a reader from looking at the rest of the mammogram
4 clearly, thereby maybe missing more? Are there going
5 to be more women recalled, which will decrease their
6 performance? It's very difficult to say. So maybe
7 one new subtle cancer is a price for 17, but what
8 happens is you're getting up to that far end of an
9 ROC curve where a huge increase in the negative side
10 is only going to give you a small value on the
11 positive side. So yes, it can affect performance.

12 DR. ABBEY: But didn't you just describe
13 concurrent reading? I mean, you just described --
14 because you just read through it to come up with your
15 read. You turn the CAD on, we'll have a million
16 marks.

17 UNIDENTIFIED SPEAKER: Microphone.

18 DR. ABBEY: Now I'm done talking.

19 (Laughter.)

20 DR. D'ORSI: Well, I don't want to
21 monopolize this because I shouldn't.

22 DR. STEIER: Can I?

23 DR. D'ORSI: Yes.

24 DR. STEIER: I would like to just comment
25 on what Dr. Ziskin said, and what was presented

1 before is this is really just an adjunct and whether
2 it's taking the films in front of the bright light --
3 do we still do that, in front of the hot light there
4 and hold it up and down?

5 DR. D'ORSI: Not if you want to pick a
6 monitor up at home.

7 DR. STEIER: Now, you have to pick up the
8 whole monitor, right? Or turn the film upside down
9 or around or increase the sensitivity on the screen,
10 on the PAC system, or turn up the contrast or not.
11 There's lots of things we do in an adjunctive fashion
12 to increase the radiologists' or other clinicians'
13 ability to read X-rays and studies.

14 So if this is really just an adjunct to
15 what the radiologist is responsible for, and it's
16 still the radiologist who's determining the ultimate
17 read and thereby influencing the treating clinician
18 and how he treats the patient, you know, I think
19 that's an important point. If it's used correctly.

20 If it's going to be used -- if you guys are
21 going to violate the label and use it concurrently,
22 you know, that's a whole different kettle of fish, I
23 guess. But if it's used properly in addition to the
24 regular read, again, I'm not sure what it is, what
25 problem it is that we're solving.

1 DR. D'ORSI: Well, again, I don't want
2 to -- it's going to sound like D'Orsi meeting here,
3 so I don't want that. I want to get opinions from
4 around the table. But to quickly -- I'm sorry -- to
5 quickly answer that, there's a lot of studies that
6 show the more marks that are on a film, your
7 performance can drop. I mean, there's clear-cut
8 studies on that. So it does have -- it's not just
9 something to ignore. So yes, it's only a device, but
10 if that device is altering how you look and how you
11 think and other factors of an interpreter which is
12 cognitive, yes, it can affect performance.

13 DR. STEIER: Okay, I would say, if that's
14 true, then a standard should be set for an acceptable
15 number of false positives, and some yardstick should
16 be set.

17 DR. D'ORSI: There you go.

18 DR. ZHOU: Well, because now we're missing
19 a point. I mean, we're talking about whether we need
20 the reader study. All I hear is that the quality of
21 the film himself does not -- has some relationship
22 with the reader accuracy. So if you don't conduct a
23 reader study, how do you know what the impact of
24 missing some abnormality by one machine and the other
25 machine picks up different abnormality, how does that

1 affect reader accuracy without a study?

2 DR. D'ORSI: Yes, Dr. Kim.

3 DR. KIM: I would say that, yes, in an
4 ideal world, for every little change, you'd want to
5 do as much as you could possibly to get the highest
6 level of certainty. But I think our job is to sort
7 of balance it with what is acceptable in terms of
8 effectiveness and safety and yet practical to allow
9 continued interest in this field, I guess, by
10 academics and CAD companies. And so I think we have
11 to decide and make tradeoffs, just like we do in
12 radiology every day, when you call lesion or you
13 don't, you know, where are you going to set your
14 sensitivity and specificity?

15 So, you know, I would say that looking at
16 it specifically from a CAD colonography
17 perspective -- and I'm sorry, I don't really have
18 that much experience in mammography -- a reader
19 performance study is really labor intensive, just
20 because this exam takes a long time to read on the
21 front end. And so I think we have to be very
22 judicious on when we apply a reader performance
23 study, and I would say that it probably needs to be
24 done on the front end, just to make sure there isn't
25 some unintended interaction that really depresses

1 performance.

2 But once it's been established at that
3 first study, as they make upgrades and request a 510
4 clearance because they feel the upgrades are valid to
5 be a 510 sort of pathway, then we should just go
6 ahead and just use standalone at that point.
7 Obviously that's not perfect. Could there be
8 something unintended? Sure. But I think that's
9 something we have to balance.

10 DR. D'ORSI: Yes.

11 DR. LEITCH: I think in terms of the marks
12 depressing performance, it sort of depends on the
13 site, I think. In the colon, where you have a large
14 surface area that you've got to look at, you kind of
15 appreciate having the marks, whereas on a mammogram,
16 if the marks cover up that tiny surface area that you
17 have to look at, then it's discouraging and you're
18 not going to do it. And with lung, because there are
19 so many granulomas, it also is the more marks is a
20 more disconcerting thing on the examination. So it
21 may depend on the organ site, how much the marks can
22 have an outcome on the performance.

23 DR. D'ORSI: All right, let's get a
24 consensus on (i) and (ii). Do you agree with (i) on
25 standalone and do you agree with scenario -- I'm

1 sorry. 2(b). Do you agree with that scenario as a
2 standalone? Let's just get a --

3 UNIDENTIFIED SPEAKER: 2(b)(ii)?

4 DR. D'ORSI: 2(b)(ii). 2(b), 2(b), dot,
5 dot. Can we start?

6 MS. MORRIS: Can I just clarify a little
7 bit? We're going to break this up to make it as
8 simple as possible, and I hope I'm not making it
9 worse. If I make it worse, tell me. So 2(b)(i) and
10 (ii) is talking about looking at standalone
11 performance, and we see differences in the standalone
12 performance between the subject device and the
13 predicate device.

14 And so the first scenario is, is that you
15 have different false positives, but there's no -- you
16 haven't missed any of the true positives. And I've
17 already heard for (ii) that there are issues with a
18 different database and a different truthing and
19 different scoring methodologies, so it needs to be
20 comparable. So if you want to add to that, that's
21 fine. So that's the first scenario.

22 The second scenario is when we start
23 changing the true positives, where you have a
24 tradeoff of -- and this is where I'm going to read
25 the question of 2(b)(i), is a new CAD identifies

1 additional abnormalities that are not detected by the
2 predicate, but misses some abnormalities that were
3 detected by the predicate, but the new CAD has fewer
4 false positives. So that's a different dynamic. I
5 was starting with a very simple scenario that
6 should've been a no-brainer, I hoped, and then now I
7 added complexity to the second part.

8 DR. DODD: So are we responding to (ii)?

9 DR. D'ORSI: Are you sure of that?

10 MS. MORRIS: Am I sure? No.

11 DR. D'ORSI: No, no, I'm sorry. I'm
12 reading 2(b)(i). That sounds like the no-brainer
13 one.

14 DR. ZHOU: No.

15 DR. D'ORSI: No?

16 DR. ZHOU: No, no, no, no.

17 MS. MORRIS: When I first described it, I
18 was simplifying it.

19 DR. D'ORSI: Oh, got you. Okay, I'm sorry.
20 That is not the no-brainer.

21 DR. ZHOU: No.

22 DR. D'ORSI: Okay.

23 DR. ZHOU: It was different.

24 DR. D'ORSI: Okay, it's the deeper problem.
25 Okay, great.

1 DR. GLASSMAN: Carl, a point of
2 clarification.

3 DR. D'ORSI: Yes.

4 DR. GLASSMAN: Are the abnormalities that
5 you're talking about true positives or are they false
6 positives that were just picked up by the CAD?

7 MS. MORRIS: For the sake of discussion,
8 I'm going to say they're true positives.

9 DR. D'ORSI: Yes.

10 DR. ROSENBERG: To clarify, what you're
11 saying is the sensitivity is the same but different
12 lesions are made and missed and the specificity is
13 better.

14 DR. ZHOU: Can I? You can't talk about
15 sensitivity here. They talk about location. So
16 location ROC is not -- because you have different
17 locations, different disease status.

18 DR. D'ORSI: All right, we're getting a
19 little mired down here.

20 MS. MORRIS: Right. Again, this is
21 standalone performance, and we're looking at the
22 performance of the device with a dataset that we have
23 truthing on and we're comparing to. So a company has
24 a device on the market already. Now, they've made a
25 difference to it. They want to introduce this new

1 model. They do standalone performance testing and
2 now the characteristics, the findings from the
3 standalone testing, are a little bit different. And
4 we need to make a judgment call. Does this raise the
5 clinical impact? Do we need an MRMC study now?

6 DR. D'ORSI: Basically, when is a
7 standalone not enough as a comparison?

8 DR. DODD: I'm sorry, can I clarify one
9 thing? Because we're having some confusion still
10 about the example. The abnormalities that are missed
11 on the new device, are they -- I was taking that to
12 mean you might miss an abnormality in image one but
13 catch an abnormality in image two. We're not talking
14 about different abnormalities within the same image
15 that are being -- or are we trying to get -- because
16 I don't know that we want to get into a location-
17 specific discussion right now because I think that
18 adds a level of complexity that we don't have the
19 time to get into.

20 DR. D'ORSI: Basically, what's being asked,
21 the CAD work on per image, period. They don't verify
22 one finding to another. So they're per image
23 findings. So you throw up a film, and your predicate
24 has X, Y, it has two true positives and two false
25 positives, and your predicate has two true positives

1 and one false positive. Do you need anything else
2 with that? You don't need a clinical test for that.

3 DR. TOURASSI: But this is not the scenario
4 that they were describing.

5 DR. D'ORSI: No, that's the easy one.

6 DR. TOURASSI: That's the easy one. Oh,
7 yeah.

8 DR. D'ORSI: Now, we're going to the hard
9 one.

10 DR. TOURASSI: Okay.

11 DR. D'ORSI: Okay.

12 UNIDENTIFIED SPEAKER: The false negative
13 one? Is that where we're going?

14 DR. TOURASSI: Now we're going to the
15 difficult.

16 DR. D'ORSI: So let's try and think
17 globally. What would kick a standalone test into a
18 clinical arena? What would do that for you, where
19 you would say that's not enough? Well, let's take
20 these examples here. The second one -- what was that
21 example you gave, Janine?

22 MS. MORRIS: Could you repeat that?

23 DR. D'ORSI: No, the example of the -- we
24 just went over the no-brainer. And if we can get to
25 the next harder level.

1 DR. ZHOU: For the first one is -- my
2 understanding is the two missed different abnormality
3 and now the new one gets more false positives.

4 DR. D'ORSI: No, no. That's what we're
5 going to into now. You're right, the first one was
6 where it actually got better. Now, go ahead, Janine.

7 MS. MORRIS: All right. You know, if it's
8 easier for us not to have these scenarios, if you can
9 agree on when something is okay and when you would
10 have a concern, that might be a better way at getting
11 at these questions. These examples were derived to
12 help facilitate the discussion. It's not that we
13 need an absolute. We need a kind of guidance on
14 where should we draw the line in the sand. I can
15 clarify one more time, but I don't want to add to the
16 confusion.

17 DR. D'ORSI: Okay, let me try. If a device
18 has a drop in true positives, would that kick off a
19 clinical study from a standalone?

20 DR. GLASSMAN: Yes.

21 DR. D'ORSI: If a device had an increase in
22 false positives, would that kick off a clinical
23 study?

24 DR. GLASSMAN: It depends on how bad an
25 increase.

1 DR. D'ORSI: Okay. So we kind of agree on
2 the sensitivity side, is that correct, that if it's
3 dropping out true positives --

4 DR. TOURASSI: I have a comment here.

5 DR. D'ORSI: Yes.

6 DR. TOURASSI: Even for the sensitivity,
7 the whole idea is to determine non-inferiority. So
8 there is some kind of confidence interval assigned to
9 everything.

10 DR. D'ORSI: Right.

11 DR. TOURASSI: So have to keep that in
12 mind, not just raw value, it missed one lesion.

13 DR. D'ORSI: Right. But we're asking, is
14 that enough to kick it up? We're not saying -- we're
15 saying that now we have to verify that information
16 with a clinical study. In other words, what you're
17 saying -- in other words, when is a standalone data
18 enough to kick it up into a clinical study?

19 DR. TOURASSI: What I was suggesting is
20 that we need to go with a statistical route. So if
21 it violates, if it reduces sensitivity beyond the
22 acceptable confidence interval that we expect, then,
23 of course, a clinical assessment study.

24 DR. D'ORSI: All right.

25 DR. STEIER: A question. What would the

1 Panel consider an acceptable -- based on the products
2 that are out there now -- acceptable sensitivity and
3 specificity? Ninety percent? Ninety-five percent?

4 DR. D'ORSI: I don't think that's the exact
5 question. The exact question, I think, may be a
6 little restrictive. She says when it's kicked up to
7 a worrisome level, and I guess the only way you can
8 get that is with a statistical margin. Do we kick it
9 up on any change in true positive? How's that?

10 DR. STEIER: Well, I was talking with staff
11 and just on a higher level of view, you know, what
12 would we consider adequate sensitivity and
13 specificity for one of these products? And if you
14 know that and if you can establish that --

15 DR. D'ORSI: I don't know.

16 DR. STEIER: -- I think then you have
17 something you can compare any new product to. And if
18 it's above that, it's good, and if it's below that,
19 it's not good.

20 DR. D'ORSI: Yeah, that's easy to say, but
21 who knows.

22 UNIDENTIFIED SPEAKER: It depends. It
23 depends on what you're doing.

24 DR. TOURASSI: Yeah, and it takes it to a
25 completely different direction. There is a predicate

1 device. This is the assumption. The clinical
2 assessment study has been done. It has established
3 that it improves radiologist performance with a
4 certain sensitivity and specificity. So this is the
5 starting point for us to decide the delta value of
6 the non-inferiority test.

7 DR. STEIER: Okay. So we agree. You're
8 just not saying what the number is.

9 DR. TOURASSI: That's what we have to
10 decide.

11 DR. DODD: Well, I don't know that we're
12 being asked to set the delta value. I mean, I'm
13 assuming that we will do some statistical evaluation
14 to make sure that we are comfortable with the
15 sensitivity value dropping significantly outside of
16 that margin. But I don't know. Is the FDA asking us
17 to establish such a boundary?

18 DR. D'ORSI: No. Let's go around with this
19 scenario. Number one, the true positives go down
20 outside the limits of the predicate, the false
21 positives go outside the limits of the predicate.
22 We'll go around. Which one, if any, would trigger a
23 clinical study? Bob.

24 MR. UZENOFF: I'm going to have to abstain.
25 I haven't discussed this at this level of detail with

1 the industry people. So --

2 DR. D'ORSI: Okay.

3 MR. UZENOFF: -- I'm reluctant to weigh in
4 on that.

5 DR. D'ORSI: All right. Can I abstain,
6 too? No.

7 (Laughter.)

8 DR. DUEHRING: I believe if it's losing a
9 significant amount of true positives, then it has to
10 go to a higher level. So if it's giving me false
11 negatives, it's going to go to a different level.

12 DR. D'ORSI: Dr. Carrino. What about the
13 other scenario? Try and answer both scenarios.

14 DR. CARRINO: Well, I'll try and make it
15 broad. I get uncomfortable when things change about
16 five percent. And so that's a number that we kind of
17 use for a variety of things, clinical predication
18 models, et cetera. I'm in favor of Dr. Tourassi's
19 approach with some kind of statistical. So if you're
20 looking at sensitivity/specificity pairs, because now
21 we're talking about different portions of the product
22 changing, I think we'd have to look at the
23 sensitivity/specificity pairs, some statistical test
24 and use our -- you know, what we typically do
25 clinically. We may accept a certain p-value.

1 Sometimes it's 5, sometimes it's 10, depending on
2 what the task is.

3 DR. D'ORSI: Okay, Dr. Kim.

4 DR. KIM: I would say, for both situations,
5 I would say that you wouldn't need any reader
6 performance study. On the flip side, if you had the
7 same sensitivity and you had different lesions that
8 were being marked and you could tell that from your
9 datasets, which if we're going to go the route of
10 replacing datasets, that might not be possible to
11 know exactly if for the same sensitivity you're
12 marking different lesions, I would say, in that
13 instance, just again trying to balance what is
14 practical or not, that you would not need a reader
15 study in that instance.

16 DR. D'ORSI: Dr. Bourland.

17 DR. BOURLAND: So if the overall
18 performance became poor, then I think, as the
19 company, I would probably go back to the drawing
20 board and try to figure out what's going on before
21 needing a reader study because it just doesn't make
22 sense to me that it would -- both things would get
23 worse, why I would want to bring that forward, unless
24 it's a much faster device and it does the computation
25 much quicker.

1 But under the scenario -- I mean,
2 otherwise, I think it is the statistical bounds which
3 went out of that, and I sort of like the five percent
4 rule, too. For physics, we try to do two or three
5 percent. But if you throw people in there, it
6 becomes 5 or 10 percent or even more. But that would
7 be appropriate.

8 DR. D'ORSI: So just to clarify, this is
9 not a drop in two parameters. This is one either
10 remaining the same or actually getting better and one
11 dropping and vice versa.

12 DR. BOURLAND: But what I think I heard was
13 the true positive was decreasing.

14 DR. D'ORSI: That's one scenario.

15 DR. BOURLAND: Yeah.

16 DR. D'ORSI: And the other one would be the
17 false positives increasing.

18 DR. BOURLAND: Okay. So still the
19 boundary --

20 DR. D'ORSI: Okay.

21 DR. BOURLAND: -- relative to the stat.

22 DR. SEIBERT: Well, I concur. I think the
23 boundaries are a good idea. I'm less concerned with
24 the false positives as opposed to a dropping of a
25 true positive.

1 DR. LEITCH: So I would agree, the clinical
2 studies would be appropriate if the drop was
3 significant in one parameter.

4 DR. LIN: I think if there's a drop in true
5 positives, we are obligated to run clinical studies,
6 especially if the drop has a statistical significance
7 because then, by definition, it's not substantially
8 equivalent by standalone studies.

9 DR. D'ORSI: Dr. Abbey.

10 DR. ABBEY: A drop in either parameter,
11 yes, since I don't really believe that will go
12 forward. If different positives are identified, then
13 I think it's very critical which positives they are,
14 things that would've been detected anyway versus
15 things that aren't. So it's not so clear to me in
16 the case that is actually specified here with the
17 question here.

18 DR. D'ORSI: Dr. Rosenberg.

19 DR. ROSENBERG: Yeah, I agree with the
20 other comments. If it's not substantially
21 equivalent, it doesn't go forward. If there's minor
22 changes in which positives are identified, that's
23 almost to be expected. But what constitutes a minor?
24 Maybe it's the few percent rule that seems
25 reasonable. But one would expect some minor changes

1 would be typical when you change the algorithm.

2 DR. D'ORSI: Dr. Ziskin.

3 DR. ZISKIN: Yeah, a drop in the false
4 positive detection rate below a statistical
5 variability, I think, does require the clinical
6 study. The false positives I'm not as much as
7 concerned about. If it really increased a very large
8 amount, that would be somewhat suspicious, and I
9 guess then I would sort of bend for the clinical
10 studies also.

11 DR. D'ORSI: Dr. Mittal.

12 DR. MITTAL: In both scenarios of
13 specificity and sensitivity, if it's in the negative
14 direction, it should trigger a clinical study.

15 DR. D'ORSI: Dr. Payne.

16 DR. PAYNE: Well, I liked Dan Bourland's
17 comments in terms of -- I think to some extent we
18 might be beating the dead horse in terms of if in the
19 testing it showed that the new device was, well,
20 certainly significantly inferior to the predicate
21 device, you're not going to go forward. So I'm not
22 so sure that -- and I understand the question, but
23 I'm not so sure that it's going to actually raise its
24 head. So I think, otherwise, the other comments
25 stand.

1 DR. D'ORSI: Dr. Zhou.

2 DR. ZHOU: Here I wanted to clarify. I
3 mean, the sensitivity and specificity here have
4 different meanings than a standard sensitivity,
5 because the correct diagnosis, not only you identify
6 presence or absence of the lesion but also location
7 has to be right. So false positives actually can --
8 including the one you identified, if you say there is
9 a presence of lesion but give the wrong location,
10 that's actually bad too. So I mean, not just looking
11 at sensitivity, you also look at specificity also
12 important.

13 But I was just thinking that's a no-
14 brainer. So if they drop, we should go to the
15 clinical. But probably a pharmaceutical company
16 don't want to do that -- the device company don't
17 want to do that. That's not good. But I'm thinking
18 suppose that's better. Suppose actually there's
19 better sensitivity. I mean, does that mean they're
20 good? Suppose they're better. That means the new
21 device does something different than the old device,
22 either good way or bad way. So in order to actually
23 confirm whether the new device is doing it better, I
24 wonder still, do we need to actually go to clinical
25 or not? If these are different.

1 DR. D'ORSI: Right. Yeah, Dr. Dodd.

2 DR. DODD: So I don't have too much to add,
3 but I want to follow up a little bit on what I think
4 Dr. Abbey was saying. I don't know exactly who. But
5 it does depend -- in a case where you're identifying
6 different positives, the question about whether or
7 not you go on to a clinical study will depend on the
8 clinical impact of those findings, right? So if this
9 is a relatively common, hard-to-identify, high-risk
10 type of lesion, then you might -- you know, you have
11 to balance those tradeoffs, right?

12 The other thing is I might be willing to
13 accept -- I hope I don't regret saying this, but,
14 yeah, some increase in the false positive rates but
15 with a bound without doing a clinical study. So I
16 like the idea. I don't know if it was Dr. Steier who
17 mentioned that, of setting a bound. And I think it
18 will depend on the particular clinical application.
19 But if we understood the real relationship between
20 the number of markings and how it impacts the reader
21 performance, then it's possible we could set a bound
22 on the number of false positive markings we're
23 willing to accept with an increase in sensitivity,
24 without going on to a clinical study.

25 DR. GLASSMAN: If there is a drop in the

Free State Reporting, Inc.
1378 Cape Saint Claire Road
Annapolis, MD 21409
(410) 974-0947

1 true positives beyond the statistical bound, I think
2 that this is a non-equivalent device and the process
3 stops and you don't need a reader study. If there's
4 an increase in the false negatives above the bound
5 without a concomitant increase in the true positives,
6 it's a non-equivalent device and the process stops
7 and you don't need a reader study. However, if you
8 get an increase in true positives with an increase in
9 false positives, then I think you need a reader study
10 to see whether the false positives denigrate that
11 true positive result in the real world.

12 DR. D'ORSI: Dr. Tourassi.

13 DR. TOURASSI: This answer strongly depends
14 on the application, and we're supposed to give a more
15 general view of the problem. I do believe that we
16 have to go with the statistical test, a clinically
17 acceptable bound for the disease, the organ, and then
18 if either/or sensitivity, specificity, or whatever
19 the performance metrics are exceed the bounds of the
20 non-inferiority test, then clinical assessment.

21 DR. D'ORSI: Dr. Jiang.

22 DR. JIANG: I like Dr. Bourland's idea. I
23 think that sort of reflects what's going to happen in
24 practice. But I also want to suggest that we should
25 recognize the fact that things are statistical. If

1 you have two CAde devices, unless they're identical
2 in every single way, you're not going to
3 necessarily -- they have the same sensitivity but
4 you're not going to necessarily detect the same
5 lesions in every image. We know this from film
6 mammograms. You'd run the same CAD device. You
7 don't always detect the same lesion.

8 So I think when we talk about that, the
9 question of what lesion gets missed and what lesion
10 you picked up is important. I think there are things
11 that we can do to look at that. For example, if you
12 pick out very obvious things, high-contrast things,
13 those are probably not good. But on the other hand,
14 if you're picking up more subtle things, that should
15 be good. So I think that argues against a fix, a
16 threshold, but more careful looking at not
17 necessarily triggering a reader study but look at
18 that first before triggering a reader study.

19 DR. D'ORSI: Dr. Swerdlow.

20 DR. SWERDLOW: I was having the same
21 comment in private over here as Dr. Bourland when he
22 was saying it. So essentially I agree, but I was
23 able to think of a scenario where that might not
24 truly apply, which is if the predicate had a
25 sufficiently high number of false positives and the

1 new test dropped that to significantly improved
2 reader performance overall, then maybe it would be
3 worthwhile. But if what you're going to talk about
4 then is incorporating improving your performance,
5 you've got to prove that with the appropriate study.

6 DR. D'ORSI: Dr. --

7 DR. STEIER: Steier.

8 DR. D'ORSI: Thank you.

9 DR. STEIER: Much like a lesion on a
10 mammogram, this sounds a little bit like a moving
11 target. But I do think we need a, you know,
12 transparent scientific way to evaluate what we think
13 are good CAD devices or not. Much like spiral CT,
14 where we know have 95 percent sensitivity and around
15 93 percent sensitivity and like that, there are
16 standards that can be developed and individualized
17 for different studies. At least we have a target,
18 something that's transparent, something that people
19 can work with and maybe take some of the subjectivity
20 out of it.

21 DR. D'ORSI: We're going to take a break,
22 and I hesitate to ask this question. Does the FDA
23 have a sense of what they need with this?

24 (Laughter.)

25 MS. MORRIS: I personally have a sense of

Free State Reporting, Inc.
1378 Cape Saint Claire Road
Annapolis, MD 21409
(410) 974-0947

1 what we need. You've given us a great body of
2 different scenarios to consider. And there will
3 still be a lot of work for the Agency to do, but I
4 think that you've given us some direction that will
5 help a lot.

6 DR. D'ORSI: I think that's all we can do
7 because these are very complex. Basically, you don't
8 want a lot of false negatives.

9 MR. SWINK: I'd like everyone to check
10 their travel arrangements when they leave the room.
11 We're going to keep you here a little longer.

12 DR. D'ORSI: Let's take a 10-minute break.
13 Is that okay?

14 (Off the record.)

15 (On the record.)

16 MS. MORRIS: Well, let's say that I'm going
17 to summarize what I heard, and I want to hear if
18 we're in agreement with this summary of what I heard
19 from the Panel. So as you can see, you can probably
20 understand now why we're having so much problem
21 ourselves trying to weigh in on these different
22 scenarios that might come before the Agency.

23 So what I heard, and I want to make clear
24 so that it can assist industry as well, is that for a
25 CAD device that is coming under 510(k), if they came

1 in with standalone performance, where we have the
2 standalone performance of the device subject to
3 510(k) and the predicate, now that predicate could be
4 their device that's on the market or it could be
5 another manufacturer's device. Now, how they get
6 that data, that's up to them. But that is an option
7 for them to do.

8 So we are comparing apples to apples, the
9 same dataset, the same methodology, the same
10 truthing, so that if we have that comparison data of
11 apples to apples and the outcome of that standalone
12 performance is within an acceptable margin of
13 variance, then they don't need to go on to a reader
14 study.

15 If it falls above or below this standard,
16 the Agency will go back and try to establish that
17 based on what we've already seen on the market. Then
18 it's a question of, well, if it's severe, we're going
19 to say not substantially equivalent. If it's
20 questionable, maybe we'll defer them and give them an
21 option to go to a clinical study.

22 So when we're saying it could be a device
23 that's been modified or it could be a new device, no
24 matter what, it's the option of comparing with a
25 predicate that's on the market and looking at the

1 standalone performance using the same set of
2 criteria. Is that what everyone understood and would
3 agree to that?

4 DR. D'ORSI: Yes. Let's raise our hands.
5 Yes. Okay. Janine, does that kind of take --
6 include 5(a) in this, sort of?

7 MS. MORRIS: Excuse me. Yeah, 5(a) was
8 just basically asking what you could identify as a
9 minor modification. So in terms of when clinical
10 data is needed, I think you've given us a good sense
11 of where we have to be careful. There's just one
12 other example I would like to have described and see
13 if there would be a great deal of disparity in your
14 comments on how to do that. And so this would be
15 2(b)(iv), and it deals with when it's a difference in
16 prompt. So Robert's going to read that for us.

17 DR. OCHS: Okay. A manufacturer previously
18 received clearance for a CAD device that can serve as
19 a predicate device. The intent to provide to the
20 intended user a different prompt format from that of
21 their cleared device, such as findings are now marked
22 with a circle rather than an arrow. The prompt
23 format is the only change made; the CAD algorithms
24 were not changed or modified in any manner. Should
25 they perform another clinical assessment? So this is

1 just a change in prompt.

2 DR. D'ORSI: Statements? Questions? Yes.

3 DR. CARRINO: My inherent answer would be
4 no, but there was something presented this morning
5 that said there was a study that I'm not familiar
6 with that said prompt changes has a significance as a
7 performance.

8 DR. D'ORSI: I do remember that. Any other
9 comments?

10 DR. ABBEY: Yeah. So there were two
11 studies presented. One was an older one by Karpinski
12 (ph.), and there was a newer one that I'm less
13 familiar with. And the older one, at least as I
14 recollect, says that the less contained the prompt
15 is, the lower the performance. What I don't know
16 about it and what I don't recollect is the
17 methodology used there and whether those differences
18 are overcome with a fair amount of experience. To my
19 knowledge, that point has not be addressed, but I
20 don't know for sure. I hesitate to comment. And
21 maybe that could be that somebody else does know.

22 DR. D'ORSI: Dr. Jiang.

23 DR. JIANG: I think it would depend on what
24 kind of change it is. If somebody changes from an
25 arrow to a circle that's big enough to encompass the

1 entire image, that's going to be pointless, right?

2 (Laughter.)

3 DR. JIANG: But if it's just really just a
4 simple change of the mark, I don't think it's
5 necessary.

6 DR. D'ORSI: Dr. Glassman.

7 DR. GLASSMAN: I would agree that it's not
8 necessary as long as it is a reasonable change, not
9 encompassing the entire image or something silly. Of
10 course, the question of why a manufacturer would
11 change a prompt to something that was silly is, I
12 guess, another issue.

13 DR. D'ORSI: Mr. Uzenoff.

14 MR. UZENOFF: Bob Uzenoff. Yeah, a
15 question in the PAC system can be a little more
16 complicated in that the output from CAD may go to a
17 PAC system, and it may be the PACS vendor or the user
18 of the PACS that configures their PACS how to display
19 the marks.

20 So you could have a situation where the
21 PACS manufacturer doesn't change the mark, but the
22 user changes the mark on their own. And you can also
23 have the situation where the PACS -- or the CAD
24 manufacturer does change the mark, as in this
25 example, but it's not changed to the user's display

1 because the PACS is configured to display positives
2 or negatives a certain way. So there's another layer
3 of complexity on that question. So I think probably
4 it's best -- I would suggest not to try to regulate
5 that. Maybe the answer is no.

6 DR. D'ORSI: Yes, Dr. Zhou.

7 DR. ZHOU: I'm wondering, since the
8 manufacturer thinks maybe the change they make is
9 significant enough and that's why they submit this to
10 the FDA. I remember I heard from the FDA, even if
11 it's a minor change, they don't have to do it
12 anything, right? So I'm assuming -- so that's a
13 significant change, right? That's the reason they
14 submit this to the FDA.

15 MS. MORRIS: I'll clarify that. If they
16 change the prompt and if they're submitting it, yeah,
17 it suggests to me that perhaps it could change the
18 user performance, and I guess my only default would
19 be if we questioned the impact of that change, we
20 could always seek Panel input.

21 (Laughter.)

22 DR. ABBEY: Could I make one more comment
23 on that?

24 DR. D'ORSI: Yeah.

25 DR. ABBEY: There's two studies there. I'm

1 not familiar with the more recent one, and that may
2 pertain more directly to this. I just don't know.
3 So that's the only caveat I would put there.

4 DR. D'ORSI: I think in general, if the
5 mark is not out of bounds with marking a lesion nor
6 covering the lesion, then I would say no, that that's
7 not necessary to do any reader study. How do people
8 feel about that?

9 UNIDENTIFIED SPEAKER: I would agree.

10 DR. D'ORSI: Okay. So can we go now to,
11 Janine, your next point?

12 MS. MORRIS: The next point is about
13 generalizability of studies. So when clinical data
14 is required, can it be generalizable across subgroups
15 and then also across devices? And so it covers a
16 number of different questions. We can start with,
17 let's see, Question (c) and (d), 3(c) and 3(d).

18 DR. OCHS: Okay, I'll start with 3(d) and
19 3(c).

20 MS. MORRIS: Yes.

21 DR. OCHS: A manufacturer's CAD device is
22 designed to detect abnormalities on mammograms. Is
23 there a minimum number of cancers that should be
24 included in their clinical study to ensure that the
25 entire spectrum of cancer is represented? Should

1 their clinical performance assessment be powered so
2 that statistically significant results can be
3 obtained for the clinically relevant subgroups of
4 cancer manifesting as microcalcification clusters and
5 cancers manifesting as masses? Does the answer
6 depend on whether or not we have prior experience
7 that CAD devices do not perform well in one of the
8 subgroups, such as masses?

9 MS. MORRIS: And then (d) is just taking a
10 similar example, but it's for lung nodules.

11 DR. OCHS: So for Question 3(d): Would
12 your answers to Item (c) apply to other CAD devices?
13 For example, a CAD device that's designed to detect
14 lung nodules. Should their clinical performance
15 assessment be powered so that statistically
16 significant results can be obtained for the
17 clinically relevant subgroups of lesions, such as
18 nodules near the mediastinum versus the peripheral
19 lung fields? Or would this only be expected if the
20 manufacturer proposed to make such claims in their
21 labeling?

22 MS. MORRIS: So, in general, we're trying
23 to find out that if we've been provided with a reader
24 study that's performed across these subgroups, would
25 it be generalizable enough to accept it as long as

1 they didn't make specific claims? Or on some of
2 these examples, like the issue with mammography, is
3 the concern between microcalcifications and masses,
4 when we know that there is a better performance with
5 microcalcifications? Do we need to make that
6 distinction and get more robust data for both
7 subgroups? And with lung, is it whether or not it's
8 needed to have that statistical significance in the
9 subgroups, or is it only necessary if they make
10 claims between these subgroups?

11 DR. D'ORSI: Thank you. Yes.

12 DR. STEIER: Yeah, at least for lung, I
13 would say only if they make claims related to that in
14 their labeling because, you know, I don't know
15 clinically of what significance that would be. But
16 if they made the claim, then I would make them
17 support it.

18 DR. D'ORSI: Yeah, Dr. Glassman.

19 DR. GLASSMAN: I would agree with that,
20 although I think that the examples of the cancers or
21 lung nodules in the dataset that they use to test
22 their machine should represent a spectrum of
23 abnormalities. I've been impressed when FDA staff
24 goes through this. I mean, they go through it in
25 such detail, if they were to see a lung nodule

1 submission that had no mediastinal nodules, that
2 would be a flag. But given that they were there but
3 the labeling was general, then statistics just for
4 the general case would be fine.

5 DR. D'ORSI: Dr. Zhou.

6 DR. ZHOU: I can agree with that, except I
7 think if we know in advance in another study that
8 maybe the -- like in this example, CAD device may not
9 perform well in some of the subgroups, even if we
10 know in advance in the literature. So maybe it's
11 worth it to consider testing a sample.

12 DR. D'ORSI: Any other discussion? So it
13 sounds as if it can be generalizable. Is that the
14 consensus of everybody? Just yes? Okay.

15 DR. ZISKIN: Yes, if it were limited,
16 though. I mean, you couldn't take some algorithm
17 for, let's say, in the breast and say this is also
18 good for the polyps in the colon. So not that degree
19 of generalizability.

20 MS. MORRIS: Within the actual clinical
21 use, meaning that we're studying a CAD device for
22 colon and, you know, there may be different findings,
23 that it's generalizable within that clinical use.
24 And we were just giving examples. One example was
25 lung. One example was for mammography. We're not

1 mixing.

2 DR. ZISKIN: Okay.

3 DR. ROSENBERG: I assume that the clinical
4 test cases are representative. They're not cherry-
5 picked.

6 MS. MORRIS: That would be my assumption
7 too.

8 DR. D'ORSI: Okay, good. Do you want to
9 take the next one?

10 MS. MORRIS: Okay. So the next group of
11 examples that we have that relate to the
12 generalizability of the clinical study is under 4.
13 So we're in this situation where, all right, 4(a)
14 deals with CT CAD and 4(c) and (d) deal with breast
15 CAD. So let me start with CT. But they're similar
16 in nature. And Robert, do you want to read 4(a)?

17 DR. OCHS: Okay.

18 MS. MORRIS: And maybe the first part too.

19 DR. OCHS: Yeah, I'll read the
20 introduction.

21 So the two draft guidance documents, when
22 finalized, will represent a change from our past
23 approach and thinking concerning the performance data
24 requirements for CAD devices. Many CAD devices are
25 currently on the market, as are a wide range of

1 medical device equipment for generating images to
2 which CAD is applied. Please discuss the conditions
3 in which clinical performance assessments should be
4 conducted for devices under review for the first
5 time, such as for devices new or previously cleared
6 with changes, to provide adequate assurance that the
7 CAD performance data are generalizable across medical
8 imaging devices. For the purpose of the discussion,
9 we have provided the following questions and
10 examples.

11 So 4(a): A manufacturer's CT CAD device is
12 intended to be used on a variety of CT devices.
13 There are a large number of CT systems currently on
14 the market. How should a clinical study be designed
15 to demonstrate that the CAD performance is
16 generalizable across all CT systems? Should the
17 study design include every type of CT system with
18 which the CAD device is intended to be used?

19 MS. MORRIS: So I'll just summarize again.
20 This is a brand new CAD coming on the market, and so
21 they're going to do a clinical study, and they're
22 going to use it on a CT. So is it adequate to just
23 have it done on one of the CTs that are on the market
24 and it's generalizable across all of the ones that
25 are on the market? Should it be a mixture of CTs for

1 it to be generalizable? Or does it have to be data
2 that supports each system that's on the market?

3 DR. SEIBERT: I've got a question. What is
4 the labeling that the manufacturer would provide for
5 that? Would it be generalizable labeling or would it
6 be for specific CT scanners? Because they do have
7 different qualities and capabilities.

8 MS. MORRIS: Certainly, based on the data
9 that we would get, we would ask the industry to have
10 appropriate labeling to match that. So I guess
11 that's a good point. In your comments to that, refer
12 to whether it should specifically be labeled as such.

13 DR. D'ORSI: Dr. Seibert.

14 DR. SEIBERT: In my opinion, then, I think
15 the manufacturers, at the minimum, should label to
16 what CT scanners their device is going to be
17 applicable for, for the different CT scanners.

18 DR. STEIER: Sure. And perhaps they could
19 provide documentation of that, if requested.

20 DR. D'ORSI: Yeah, Dr. Zhou.

21 DR. ZHOU: Well, I just try to answer the
22 questions. I think this is kind of straightforward.
23 Statistical point of view is just random sample
24 because you want to generate your results to all CTs.
25 So the only way to do that is to random sample. I

1 don't know what number is from that population of CT
2 and do the study.

3 DR. PAYNE: Dr. D'Orsi?

4 DR. D'ORSI: Yes, I'm sorry, Dr. Payne.

5 DR. PAYNE: Well, I think you're asking
6 almost an impossible question. I mean, I do do CT.

7 DR. D'ORSI: We specialize in that.

8 DR. PAYNE: I do CT, and I think Tony and
9 Daniel. CT has become very complicated. There are
10 all kinds of algorithms, you know, sharp ones and not
11 so sharp. We have all kinds of reconstruction
12 techniques. We have, you know, spiral scanning. We
13 reconstruct their images two, three, four, five, six,
14 eight times. You go into a department now, and
15 you'll acquire maybe one or two datasets, but you'll
16 review them under several different reconstruction
17 algorithms.

18 So I think just saying that you're going to
19 do it over a number of CTs, I don't think that's
20 going to answer the question at all. You know, it's
21 just too complicated. I understand your dilemma, but
22 I don't know how to solve your problem.

23 DR. D'ORSI: Yes, Dr. Swerdlow.

24 DR. SWERDLOW: Yeah, given the vast number
25 of, to use a current analogy, mix and models of CTs

1 out there, that strikes me as being virtually
2 impossible. That'd be like asking you to design a
3 tire that fits every make and model of car currently
4 on the road in this country. I have no doubt that in
5 some parts of the country, there are still some
6 people using non-spiral CTs for limited applications,
7 and I can't imagine trying to design a product for
8 something that old. I think it might be reasonable
9 to say CTs manufactured from such a date forward and
10 that we've tested them on, you know, the
11 manufacturers that are responsible for some
12 percentage of the market penetration in this country.
13 But to try and do them all is ridiculous.

14 DR. D'ORSI: Mr. Uzenoff.

15 MR. UZENOFF: I'm Bob Uzenoff. It is a
16 complex issue, and some of the things that come to
17 mind are that by the time a device is approved and
18 the images have been collected, the CT scanner may be
19 discontinued and new CT scanners may be on the
20 market.

21 What I would suggest to deal with that is
22 to ask the manufacturer, in their labeling, as I
23 think that they would, to indicate what scanners the
24 CAD system had been tested on and to offer a comment
25 on why they feel that their CAD product is

1 generalizable. They're the best to know what the CAD
2 system is sensitive to, and so they can support or
3 add comments about the generalizability. It may not
4 be generalizable or not, but there's -- you know,
5 there's a spectrum of generalizability, and they're
6 in the best position to comment on that.

7 DR. D'ORSI: So is it fair to say that they
8 should clearly discuss what CT units fit their CAD
9 device, and if they're silly enough to say
10 everything, they have to prove it for everything? Is
11 that fair?

12 DR. BOURLAND: They're somehow going to
13 have to specify the parameters for image acquisition,
14 which, as we're hearing, can be large.

15 DR. D'ORSI: Does that help?

16 MS. MORRIS: Yes, it does. It's as we
17 suspected that, you know, it's up to the companies to
18 come in and make certain statements and claims, and
19 the burden is upon the industry to provide the data
20 that supports those.

21 So if they come in with a new CAD device
22 and the data demonstrates adequate performance in CT
23 A and CT B and they label it accordingly, we think
24 that that would be suitable for it to go on the
25 market. And I like the idea of some description or

1 justification about generalizability specific to the
2 characteristics because, I agree, it's the industry
3 that's going to know those products the best and
4 where the vulnerabilities are.

5 DR. D'ORSI: Okay, let's go to the next
6 point, then.

7 MS. MORRIS: So the next difficulty is in
8 the area of mammography CAD. (c) and (d) are just
9 two variations. So I'll have Robert read (c) first.

10 DR. OCHS: A manufacturer has a new breast
11 CAD device and would like to market it for use with
12 all legally marketed full-field digital mammography
13 systems. How should the clinical study be designed
14 to demonstrate that the CAD performance is
15 generalizable across all legally marketed FFDM?
16 Should clinical studies with each legally marketed
17 FFDM be required?

18 MS. MORRIS: So it's the same as the CAD --
19 the CT example, but we're just going to mammography
20 CAD and whether or not there would be a different
21 answer since this is a Class III PMA device.

22 DR. D'ORSI: Dr. Glassman.

23 DR. GLASSMAN: I think that if the device
24 passed standalone testing for the additional full-
25 field digital mammography unit, that that should be

1 satisfactory. If the device failed the bench
2 testing, because it's already a legally salable
3 device, a multi-reader, multi-case study would be an
4 option to withdrawing the unit for that claim and to
5 see if they can prove equivalency.

6 MS. MORRIS: Okay, I'm confused a little.
7 Let me clarify again. This is for a Class III
8 mammography CAD device, brand new, and let's say
9 there are currently five FFDMs currently on the
10 market. So how do we want the clinical data? Do we
11 want it to be representative across all of them or
12 restrict the clinical data to whatever the company
13 comes through to demonstrate?

14 DR. GLASSMAN: I'm sorry. I misunderstood
15 the question. For a new Class III, clinical data
16 would be necessary. I think it would depend on the
17 labeling. If the manufacturer labeled it for a unit
18 or two units, they would have to show data for those
19 units, and if not, they would have to show data for
20 any unit that they wanted to use it with.

21 DR. D'ORSI: Dr. Rosenberg.

22 DR. ROSENBERG: I guess one question is do
23 they have to power the study for the different units?
24 And I think that's -- well, right. And that's a
25 problem. But we know that the full-field digital

1 units are at least subtly different in the data that
2 they acquire. And maybe the physicists could help us
3 with how much that might matter.

4 DR. D'ORSI: Dr. Seibert.

5 DR. SEIBERT: Well, certainly each of the
6 full-field digital units do have a different
7 sensitometry or response for the input exposure
8 relative to the output digital values, and that's
9 going to have an impact on how the individual -- the
10 Class III CAD device would act on that system. I
11 think the manufacturers are savvy enough, however, to
12 be able to understand what that so-called for-
13 presentation data is, and they can get it into the
14 space in which their particular full-field or their
15 CAD device functions. And it would have to be that
16 they would, in their labeling, say yes, we are
17 actually proficient with Fuji and with GE and with
18 Siemens, so on and so forth, and Hologic. And if
19 they did that, then I think that would be an
20 appropriate thing that they would have on their
21 labeling.

22 DR. D'ORSI: Dr. Tourassi.

23 DR. TOURASSI: I think another acceptable
24 scenario would be to have the clinical assessment
25 study done on one FFDM manufacturer with a full MPMC

1 design, but then have additional validation studies
2 of standalone performance for the additional
3 manufacturers that they target in their labeling.

4 DR. D'ORSI: Dr. Jiang.

5 DR. JIANG: Yeah, just to follow up on
6 that, I would agree with that approach. The question
7 in my mind is will the radiologist performance differ
8 on different manufacturers' FFDM? I mean, we approve
9 these FFDM machines to say they're equivalent, right?
10 So that would, to me, mean the radiologists'
11 performance are similar. Then, if that's true, why
12 would we expect the computer-aided performance to
13 differ if the standalone are also similar across the
14 different devices? So to me standalone should
15 suffice.

16 DR. GLASSMAN: I'd like to withdraw my
17 comment and agree with the two people to my right.

18 (Laughter.)

19 DR. D'ORSI: Dr. Zhou.

20 DR. ZHOU: But that hasn't been established
21 yet, to say the five existing devices have similar
22 performance across all the radiologists. That has
23 been established scientifically? I don't think so,
24 right? But DMIST. Well, I don't see that. They do
25 report in the papers.

1 DR. JIANG: Isn't that also the 510(k)
2 clearance process for FFDM machines? Am I mistaken
3 here?

4 MS. MORRIS: We're talking about a mammo
5 CAD, which would be a III. Yeah.

6 DR. JIANG: So what I'm confused is, are we
7 questioning the similarities between the images from
8 different FFDM machines and the radiologist
9 performance from those machines without CAD? I think
10 if there is no difference there, then I don't see a
11 reason for us to question why CAD would differ.

12 DR. D'ORSI: Dr. Seibert.

13 DR. SEIBERT: Well, the only reason it
14 would differ is because of the fact that they do have
15 a different sensitometric response from the input
16 data going in to the digital values going out. For
17 instance, Fuji has a logarithmic space whereas GE, as
18 an example, has a linear space, in terms of
19 conversion of incident exposure to output digital
20 values. So I know that the CAD manufacturers know
21 that. They have to know that so that they can
22 normalize that input, that DICOM, not for-
23 presentation, for-processing data, so that they can
24 make sure that the algorithms will handle that
25 information correctly.

1 DR. D'ORSI: If I'm hearing correctly, I
2 think they have to also label which devices their CAD
3 can be placed on and not have it generalizable,
4 similar to CT but probably not as many permutations
5 as CT.

6 MS. MORRIS: I also heard that the
7 standalone performance testing of all -- each -- you
8 know, if you can do that -- is also demonstrable to
9 help with bridging to the clinical. Did I understand
10 that correctly? No?

11 DR. D'ORSI: Say that again. I think
12 that's correct, what you just said, yeah.

13 MS. MORRIS: Okay.

14 DR. D'ORSI: If the standalone is the same.

15 MS. MORRIS: So if they wanted to label
16 their device to be used with five FFDMs, but their
17 clinical study was done on one, as long as the
18 standalone performance for five were within a margin
19 of acceptability, then that would be acceptable.

20 DR. D'ORSI: I think that's valid, yeah.
21 As everybody knows, these units are not looking at
22 what we're looking at. They're looking at the look-
23 up tables. So they don't have to deal with monitors
24 and brightness and contrast and anything. So if the
25 raw data, the look-up tables, are substantially the

1 same, then I agree, then they would have to label
2 which ones they were going to use it for, for the
3 reader study, which would bring in a lot of other
4 factors that the machines don't deal with. Is that
5 fair?

6 DR. SEIBERT: I think that there would be a
7 least burdensome approach, from my perspective, if
8 you do a full study on one system and then you apply
9 the standalone only on the others, with the
10 appropriate modification of the input data.

11 DR. JIANG: I think --

12 DR. D'ORSI: Yes, Dr. Jiang.

13 DR. JIANG: -- that would address the
14 concern Dr. Seibert raised earlier about the
15 different physical characteristics, the differences
16 between the devices. You would need that to know it.

17 DR. BOURLAND: A brief comment here. So I
18 think this maybe goes back to the CT. It's just
19 maybe the point you get is, well, these are all DICOM
20 images, therefore they'll work. And maybe I don't
21 know if you get that type of presentation or not.
22 Then that certainly goes a long way, but there are
23 issues relative to acquisition and the meaning of the
24 intensity values.

25 DR. D'ORSI: Consensus? Standalone?

1 Substantially the same? Then we can have a one-
2 reader study. Is that the consensus? Good, okay.

3 MS. MORRIS: Okay. So on (d), I think you
4 may have answered this. It's flipping it around. So
5 Robert, would you read (d)?

6 DR. OCHS: So a manufacturer has a breast
7 CAD device approved for use with a specific legally
8 marketed FFDM system based on a robust MPMC study.
9 They would like to market it for use with additional
10 legally marketed FFDM systems. Is a clinical
11 performance assessment necessary to assess the CADe
12 for use with a new FFDM or is standalone performance
13 data sufficient to demonstrate comparable performance
14 based on the specifications of the device?

15 DR. D'ORSI: We kind of answered that yes.
16 Okay. All right, let's go the next point.

17 MS. MORRIS: Okay, the next point is --

18 DR. D'ORSI: It's 5(b), I think.

19 MS. MORRIS: Yeah, 5(b). So it's again
20 specific to mammo CAD and the fact that there are
21 different -- and I'm going to have Robert read 5(b).

22 DR. OCHS: Do you want me to read
23 Question (i), too?

24 MS. MORRIS: Let me read, myself, while
25 you're reading the first part.

1 DR. OCHS: Oh, okay, okay.

2 MS. MORRIS: And then I'll decide if we
3 have to go into --

4 DR. OCHS: Okay.

5 MS. MORRIS: -- the subparts.

6 DR. OCHS: Okay. So for 5(b):
7 Mammographic CAD devices contain separate and
8 distinct algorithms that detect masses versus
9 microcalcifications. The following questions seek
10 input on whether this distinction should have
11 regulatory significance.

12 MS. MORRIS: Okay, go ahead and read the
13 first one.

14 DR. OCHS: Okay. Not enough. If a
15 regulatory submission for an original mammography CAD
16 device reveals that reader performance does not show
17 safety and effectiveness separately for masses and
18 microcalcifications, such as, suppose safety and
19 effectiveness is shown for microcalcifications but
20 not for masses, should the indications for use
21 specify that the device is only indicated for the
22 detection of microcalcifications? If yes, do you
23 believe that the mass detection portion of the device
24 should be disabled or removed?

25 DR. D'ORSI: Comments? Discussion?

1 Dr. Glassman.

2 DR. GLASSMAN: Yes and yes.

3 DR. D'ORSI: That's fast. Okay. Comments?

4 Dr. Leitch.

5 DR. LEITCH: I guess it would be the degree
6 of difference, you know. I mean, if it's somewhat
7 less for detection of masses but it does detect to
8 some degree, you could indicate that lower
9 performance and still allow it to exist.

10 DR. D'ORSI: Discussion? Any other
11 discussion? What's the sense of some -- any other
12 discussion on this? Basically, we're saying that if
13 there is -- does not show safety and effectiveness,
14 that's pretty severe. For masses and calcifications,
15 would you allow it to be placed for both or would you
16 give restrictions on whatever it's not safe and
17 effective for? Is that right?

18 MS. MORRIS: Yes, but in order to do that,
19 you would have to have the study powered for both
20 subgroups, correct?

21 DR. D'ORSI: Yes, in order to get the
22 premise. Is that what you're saying?

23 DR. CARRINO: Right.

24 DR. D'ORSI: Yes.

25 DR. CARRINO: Not being a mammographer, I

1 think those are --

2 DR. D'ORSI: Oh, I'm sorry, yeah.

3 DR. CARRINO: Not being a mammographer,
4 those are two fundamentally different tasks, right,
5 looking at calcifications versus masses? So they
6 should be two different -- basically, they're two
7 different CADes, in my musculoskeletal opinion.

8 (Laughter.)

9 DR. D'ORSI: Dr. Jiang. Yeah.

10 DR. JIANG: I guess the question is not too
11 clear in my mind. One side of it has to do with the
12 labeling. Are we talking about breast cancers or are
13 we talking about masses versus calcifications in the
14 labeling? I don't think people are labeling it that
15 way. So if we're talking about breast cancer, and
16 then the question would be, for example,
17 calcifications do better. Masses do not as well.
18 Then the question, to me, would be, for masses, with
19 the current performance, would we expect it reduces
20 radiologist performance? Or was it a question that
21 it may improve performance, but we just can't -- we
22 just don't know it for sure? So if it's the latter
23 scenario where we expect the improvement but we're
24 not sure, I'm not as concerned about that. But if
25 we're really concerned about there's a possibly of

1 reduction of performance, then I think we should look
2 at it separately.

3 DR. D'ORSI: Yeah, just a point of
4 clarification.

5 DR. LIN: I think we should --

6 DR. D'ORSI: Oh, yeah, maybe this will
7 help. Reader performance does not show safety and
8 effectiveness. That's the premise of this statement;
9 is that right? Okay.

10 DR. LIN: I think it's not clear whether or
11 not it was just not tested and not powered enough to
12 show safety and effectiveness or whether or not it
13 showed that it was unsafe and ineffective in this
14 question. That's very important.

15 And also, the other thing we need to
16 remember is that if these devices are used properly,
17 i.e., as a second reader, then it really should not
18 deleteriously affect the radiologist performance,
19 even if the accuracy for mass detection is not high.
20 So in other words, it shouldn't adversely affect the
21 performance of the radiologist. The radiologist
22 should've read through the whole scan first and then
23 turned on the CAD and then see if there's any extra
24 lesions that could be picked up.

25 DR. D'ORSI: I am still hung up on reader

1 performance does not show safety and effectiveness.

2 MS. MORRIS: Okay, I'll clarify. So let's
3 look at the mammography CADs that are on the market
4 now, and they've done one -- let's say they've done a
5 reader study and it had a mixture of
6 microcalcifications and masses, but it's overall
7 performance. But it's not -- you can't do a
8 statistical subgroup analysis to show a difference
9 between the two groups. But we know from the
10 literature that it does perform better for
11 microcalcifications. Is it important clinically
12 enough to show that demonstration in both groups?

13 DR. D'ORSI: Ah, that's much more clear
14 now. All right, let's try and answer that.
15 Dr. Glassman.

16 DR. GLASSMAN: I think if the study is
17 powered to show the combination. But when staff
18 looks at the data, it's like the lung example that we
19 had before, if they look at the data and find that
20 it's 80 percent microcalcifications, 20 percent
21 masses, and it missed 80 percent of the masses, even
22 though it may have passed as an aggregate, that that
23 would be a red flag and they would have to either
24 change it or it would be labeled only for
25 calcifications. On the other hand, if the masses

1 were lower but not egregiously lower, then it could
2 be labeled for the general use of CAD for breasts.
3 So it would have to look at the subgroup analysis,
4 even though it's not necessarily powered to prove
5 truth.

6 DR. D'ORSI: Let's go to Dr. Rosenberg and
7 then Dr. Dodd.

8 DR. ROSENBERG: And there is a downside if
9 it doesn't work for one or the other because the
10 radiologist will be spending time tracking down those
11 marks, and if they're not valuable, they probably
12 shouldn't be spending their time.

13 DR. D'ORSI: Dr. Dodd.

14 DR. DODD: Yeah, I just want to come back
15 to this. If it's not egregiously lower, then it
16 could be, I think you said, labeled for both. But,
17 again, that comes down to having enough power to -- I
18 mean, we need some statistical power to include that
19 in the labeling, if we're going to do that.

20 DR. D'ORSI: Yes, Dr. Bourland. Sorry.

21 DR. BOURLAND: So we had talked previously
22 about subgroup analysis and whether we need power or
23 not, and we sort of said, well, we probably don't to
24 some degree. So now we're saying, well, maybe we do.
25 And so the question is are there very important

1 groups of diseases that have a prevalence or a risk
2 associated with either the diagnosis, or especially
3 maybe their non-diagnosis, that they should be looked
4 at at the subgroup level? I don't know exactly, but
5 this might be a very good case for that.

6 DR. DODD: And I would certainly think if
7 it's in the label, it needs to be established, right?
8 Statistically.

9 MS. MORRIS: Yeah, if it was in the label.
10 Right now it's just a general. It's not making that
11 distinction. But because of what we've seen in the
12 literature, we're wondering if this is a unique case
13 where we have to take exception. So the previous
14 discussion, I think, is still valid, about subgroup
15 analysis doesn't need to be statistically powered.
16 But we're asking specifically for mammography. Is
17 this a unique case that we should take an exception
18 to?

19 DR. D'ORSI: And my opinion is yes. Those
20 two findings represent usually two different types of
21 disease. The calcification usually is representative
22 of in situ disease and the soft tissue or mass are
23 asymmetry, but the soft tissue part usually is
24 invasive. So they are aimed at two different
25 situations, and I think if they didn't power it for

1 that individually, I think they should do that. How
2 do others feel about that? Yes, Dr. Lin.

3 DR. LIN: I think an analogous situation
4 will be with CT colonography and flat and
5 pedunculated polyps. The flat polyps are more
6 difficult to detect. They're genetically different
7 and also have a different prognosis. So I think that
8 will be a very analogous situation. Right now, with
9 CT colonography, they don't really differentiate
10 between these two types of polyps. In other words,
11 they're not powered separately to detect each type of
12 polyp on its own.

13 DR. D'ORSI: Yes, Dr. Zhou.

14 DR. ZHOU: But if they're not in the label,
15 I don't know why we want to power to detect the
16 subgroup analysis because you could do analysis, but
17 I mean, what are you going to use the results for if
18 the label doesn't say they're -- a mammogram is good
19 for both masses and microcalcifications? If they do
20 say that, then you have to -- I mean, design the
21 study to have enough power to detect the subgroup
22 analysis.

23 DR. D'ORSI: So I don't know what I'm
24 hearing here, but yes, Dr. Dodd.

25 DR. DODD: Well, I guess I thought I heard

1 Dr. D'Orsi say that these two groups of types of
2 lesions were so different, but important, that you
3 would want to require those to have power in both of
4 those categories.

5 DR. D'ORSI: Yes, I think I'm going back to
6 what Bob Rosenberg said. They are two distinct
7 findings that very frequently, above 90 percent in
8 those two scenarios, relate to those findings. One's
9 in situ disease and the other non-in situ disease
10 pretty much of the time when you're dealing with
11 calcium and soft tissue. So if it's not -- if
12 there's a question whether it's going to effectively
13 demonstrate one over the other, then Bob is right.
14 Why waste your time even looking at that? Am I
15 getting that straight, Bob, or not?

16 DR. ROSENBERG: Yes, that's what I said.

17 DR. GLASSMAN: Can I ask a question?
18 Because I don't know the answer to this. How big a
19 difference is it to power the multi-reader study for
20 both together versus both separately? Because if
21 it's a huge difference, then I would suggest that we
22 go with the red-face test for the masses. That is,
23 if the data looks reasonable to staff when they do
24 the subgroup analysis, that's one thing. If it looks
25 unreasonable to staff, then they either have to do a

1 powered study for the unreasonable half or change the
2 labeling.

3 DR. D'ORSI: Can we have some discussion
4 statistical point? Dr. Dodd.

5 DR. DODD: Well, I guess if you're doing a
6 stratified analysis, it would potentially double the
7 size of the study. But I guess the other question I
8 would have is would this be a place where you would
9 do extensive standalone testing which would be much
10 more limited?

11 DR. D'ORSI: That's what I was going to
12 suggest. I'm glad you brought that up. You can
13 stress test the system under that scenario, as a
14 standalone, and you should get a better idea, then.
15 If it fails all subtle findings, then either you go
16 back to the drawing board or you try and do a
17 clinical study with that. But I think that's a good
18 point, that you can stress test a system with more
19 subtle calcium or more subtle soft tissue findings.
20 How do people feel about that? Is that valid?

21 DR. ZHOU: So the only way -- if you do
22 that, you cannot get the unbiased estimator of the
23 diagnostic performance.

24 DR. D'ORSI: You can't or you can?

25 DR. ZHOU: Cannot.

1 DR. D'ORSI: You cannot.

2 DR. ZHOU: Because you don't have --

3 DR. DODD: But you could, within the
4 particular categories, get an estimate of the
5 performance for microcalcifications or the masses,
6 right? And that would be what would be reported.
7 But presumably there's also a standalone study that
8 has a more representative distribution of the types.

9 DR. D'ORSI: The original study, right. So
10 if the sub-analysis, which may not be powered to
11 those questions, you could do just a standalone
12 stress test or get more subtle findings in there or
13 repeat just that as a standalone. How do people feel
14 about that? Is that going around in a circle or is
15 it okay?

16 DR. ZISKIN: I have a different point.

17 DR. D'ORSI: Yeah, I'm sorry, Dr. Ziskin.

18 DR. ZISKIN: I guess this is a question for
19 the people who are doing mammography all the time.
20 But it would seem to me that in the clinical practice
21 you need to have both microcalcification detection
22 and mass detection, and that one device that will
23 only do one of those two would not be useful at all.
24 So if it's going to be useful, I feel that it doesn't
25 make any sense. And so on that basis, I feel that

1 you would need to be able to do -- both things have
2 to be proven clinically in order to be a viable
3 device.

4 DR. BOURLAND: A question.

5 DR. D'ORSI: Yeah, Dr. Borland.

6 DR. BOURLAND: So what FDA is -- one of the
7 things, are you asking the following, and that is
8 that for labeling, for instance, that detection of
9 soft tissue masses and calcifications, that they're
10 both important enough that in a label it would be
11 structured such as this device is indicated for
12 calcifications? And then if it didn't do well on
13 masses, it would say, and not masses? I mean, is
14 this a consideration at this point?

15 MS. MORRIS: Yeah, I think we have a couple
16 of different options. The question is is this an
17 important clinical distinction, now that we see data
18 in literature about the performance of mammo CAD and
19 that it appears to have a higher performance with
20 microcalcifications? Is this clinically important?

21 And then if it is, then we have a couple of
22 options, that we would require labeling to show the
23 relative performance of each so that it's just truth
24 in labeling, that this is the MRMC study, this is
25 standalone performance study, and these are the

1 results between these two subgroups. And we're
2 picking these two subgroups because they're the
3 biggest subgroups. We're not talking about the
4 architectural distortion and things like that. So
5 that's one option. We could just say, well, whatever
6 you have, you put it in the labeling so that the user
7 knows what that relative performance is. Or we are
8 asking for, you know, statistically significant --
9 you know, powering the study so that there is some
10 robustness behind that disclosure in the labeling.
11 Or we change the indications for use.

12 So if it turns out that the performance is
13 really only in microcalcifications, then the IFU, the
14 indications for use, is just for that. So there's a
15 tiered approach, different opportunities to address
16 the --

17 DR. D'ORSI: The only thing I can give,
18 hopefully, to help is that (a) those two features are
19 important, and (b) they're visually very different.
20 So with that, I would think that the machine should
21 be effectively describing each independently. I
22 don't know what it means to okay a machine for both,
23 unless there's calcium in a mass, and that doesn't
24 happen that often.

25 MS. MORRIS: Well, there's an assumption

1 that the study is giving you a mixture of those and
2 that it's generalizable across that. So it's overall
3 performance and whether or not it's clinically
4 important information for you as a clinician to see
5 the relative performance between the two.

6 DR. D'ORSI: Yes.

7 MR. UZENOFF: Bob Uzenoff. Two comments.
8 One is I hear what Dr. Ziskin is saying, and not that
9 I would disagree, but I wonder if that's getting into
10 the market so much. In other words, would we want to
11 rule out the possibility that if someone came up with
12 a fantastic mass detecting and that's all it did and
13 the user knew that, would you want to insist that
14 only devices that could do both things be on the
15 market? So I think it takes away from the market a
16 possible future use if you were to do that.

17 And then if I understood correctly when
18 we're talking about the subgroup analysis, there
19 would be an MPMC study, subgroup analysis, if there
20 was a level of concern raised, that staff could ask
21 the manufacturer to do additional testing and that
22 could be standalone testing, for instance, a stress
23 test for whatever the area of concern is. Is that
24 what I heard?

25 DR. D'ORSI: Yeah. Can we go around and

1 just get -- what's on the table now is does a mammo
2 CAD manufacturer have to show effectiveness for both
3 calcium and mass, if that's the intended use of the
4 machine?

5 MR. UZENOFF: Bob Uzenoff. I'm not sure I
6 like the question. I think the device should do what
7 it's labeled.

8 DR. D'ORSI: What is it labeled?

9 MR. UZENOFF: If it's labeled to do both,
10 it should do both.

11 DR. D'ORSI: What do they usually -- I
12 honestly don't know how they label these. That's how
13 stupid I am. How do they label that? It's going to
14 find cancer? It's going to find findings? It's
15 going to find calciums? It's going to find soft
16 tissue? It's going to find -- I don't know what the
17 label is.

18 And that's true. It should conform to the
19 label. But if it says it's going to help detect
20 significant findings, that means calcium, too, unless
21 it says I'm only going to go for calcium or I'm only
22 going to go for a mass. So I'm unclear about that.
23 You're right, it should be specifically what the
24 label says, but I don't know what the labels say.
25 Does anybody?

1 DR. SMITH: This is Robert Smith. The
2 labeling of these devices does have claims that it
3 detects masses, microcalcifications, architectural
4 distortion, even focal asymmetry.

5 DR. D'ORSI: Okay. So let's go around.
6 Mr. Uzenoff.

7 MR. UZENOFF: So could you restate the
8 question?

9 DR. D'ORSI: The question is if the label
10 on the mammo CAD unit, as was described, states that
11 it is effective and safe for the detection of
12 calcifications, masses, architectural distortion, and
13 focal asymmetries, but I'll be nice and say soft
14 tissue versus calcification, should there be an
15 independent validation showing the safety and
16 effectiveness for each of the calcific and non-
17 calcific findings?

18 MR. UZENOFF: So yes. So the scientific
19 evidence should support the claims of the labeling.

20 DR. DUEHRING: I concur.

21 DR. CARRINO: Yes, I think this is where
22 the clinical context is very important. So you know,
23 I can imagine if somebody made a knee MRI CAD that
24 showed ACL tears but not meniscal -- I mean, it only
25 picked up one or two percent of meniscal tears, but

1 that's where I really have a problem, then it's not
2 too useful a system because I can see ACL tears a lot
3 better.

4 So the point is, is that the clinical
5 context should drive it. And so if we know as much
6 as we know about the clinical context, that there's
7 differences between calcifications and masses, it
8 should be labeled as such. I would not say that you
9 should label it good for calcifications but not good
10 for masses. You should just only, you know,
11 emphasize what it's good for.

12 DR. KIM: I would say yes, as well.

13 DR. BOURLAND: I think this is a tough one
14 because I think it's been done in general, and now
15 we're asking for better statistics on two groups in
16 particular, and I think that would be very important
17 since probably the majority of disease is in those
18 two categories. So a very important issue, and to
19 encourage industry to work in that way.

20 Whether this should take place, when? Now?
21 In a year? And what about for the future ones, which
22 is the Part (ii) here, which I think is what do we do
23 with the ones that aren't performing as well in a
24 subset? Which I think is the next question. It's
25 very important. So I haven't yes or no, but within

1 some bounds, I think it's a good way to go, direction
2 to head.

3 DR. LEITCH: Yes, I think the labeling
4 should reflect what you've demonstrated
5 scientifically to be the case.

6 DR. LIN: If the label says so, then
7 definitely yes.

8 DR. ABBEY: So I'm a little confused, and
9 I'm not a clinician. So if it's says good for
10 detecting cancer and they do an MRMC study that's all
11 microcalcifications, I don't think that's correct and
12 I don't think anybody would do that. So there's a
13 tradeoff between the balance or the spectrum used in
14 a validation that's not powered individually versus
15 studies that are powered for each individual
16 subgroup, and then, at what point do you stop
17 splitting the hairs?

18 So I kind of think something ought to be
19 done, but it either ought to be taken care of at the
20 stage of the case spectrum or the individual studies
21 for each -- you know, each specific claim. If they
22 say microcalcifications, obviously. But if they
23 say -- if something is said more generally, it
24 strikes me as then you have to sort of match the
25 balance of cases that you would expect to find

1 reasonably, clinically.

2 DR. D'ORSI: Dr. Rosenberg.

3 DR. ROSENBERG: I mean, I think the
4 labeling has to reflect the performance and -- yeah,
5 period.

6 DR. ZISKIN: Based upon Mr. Uzenoff's very
7 tactful way of saying I was wrong, you dissuade me
8 and I retract what I said, although I still wouldn't
9 buy the machine.

10 DR. MITTAL: I think the labeling should
11 reflect the intended use.

12 DR. PAYNE: And I concur with the same.

13 DR. DODD: I concur.

14 DR. GLASSMAN: I concur but, and the but is
15 that if the prior products that have been approved
16 have been approved with studies powered for the
17 general case and not for each individual subset, that
18 the follow-on product should have the same barrier to
19 pass with the staff looking and making sure that
20 there isn't a glaring hole in the data.

21 DR. TOURASSI: Yes, I agree with
22 Dr. Glassman to a large extent. First of all, I
23 believe subgroup analysis of standalone performance
24 is very important because this needs to be conveyed
25 to the user so that there is no misconception of what

1 to expect from the device. But beyond that point, to
2 have, to expect statistically powered studies
3 demonstrating significant improvement in
4 effectiveness, it's almost unrealistic for some of
5 the signs, such as architectural distortions. I
6 mean, the less prevalent ones. As long as there is
7 no statistical evidence of detrimental effect, I
8 think we can still go on with the device as is.

9 DR. JIANG: So I'd like to modify my
10 earlier comment a little bit, swayed by the clinical
11 perspective here. I think, you know, if masses is
12 that important, I think that should be stated a
13 priori, that we want the mass to be tested
14 significantly. You know, I think that's a valid
15 thing to do. But I don't think this should be
16 generalized to all of the subcategories. I mean,
17 that's just not a realistic thing to do.

18 In the case that we don't know a priori
19 which subclass is most important because we didn't
20 know how the computer would affect the radiologist,
21 maybe I would suggest to the FDA to consider in the
22 labeling to state that we don't know whether in that
23 subcategory the improvement is significant. You
24 know, but not saying it's not going to help, right,
25 because we don't know that either. So that would be

1 my modification.

2 DR. SWERDLOW: I'd like to also echo what
3 Dr. Glassman said. I agree. If there's any
4 particular finding that a CAD system is better at,
5 then industry is more than welcome to tout that as an
6 improvement over the existing. But I don't think we
7 can raise the bar on them when there's already
8 approved systems around.

9 DR. STEIER: Yes, but clinical correlation
10 is advised, with a repeat study in three months.

11 (Laughter.)

12 DR. STEIER: Yes, the labeling should
13 reflect the clinical study.

14 DR. D'ORSI: Mr. Uzenoff.

15 MR. UZENOFF: Yeah, Dr. D'Orsi, I think
16 Drs. Glassman and Tourassi articulated what I was
17 thinking but not saying, and what I said earlier, I
18 think, was that the evidence should support the
19 claim. But the point was brought that we've got good
20 systems on the market now. We have systems on the
21 market now.

22 If we're talking about doubling sample
23 sizes to power separately for statistical
24 significance on calcifications and separately on
25 masses, that's not what I had in mind. But I think

1 what we're doing now is powering for improvements in
2 cancer detection and then we're asking the staff
3 to -- there will be sub-analysis, and if there are
4 areas of concern in masses or calcs, that could be
5 dealt with with standalone testing after that.

6 DR. D'ORSI: Yeah.

7 MR. UZENOFF: That's what I --

8 DR. D'ORSI: Yeah.

9 MR. UZENOFF: -- intended to say.

10 DR. D'ORSI: I think that's what we're
11 getting at. Let me see if I can clarify that.

12 MS. MORRIS: Would you like me to summarize
13 what I heard?

14 DR. D'ORSI: That would be good --

15 MS. MORRIS: Okay.

16 DR. D'ORSI: -- since you're going to have
17 to write the regs.

18 MS. MORRIS: So I actually heard kind of a
19 mixture because maybe there was a misunderstanding of
20 the premise. So I'm going to set a premise of what I
21 would expect we are currently doing in terms of when
22 we get an MRMC study for a mammography CAD, that the
23 dataset is proportional to the characteristics in the
24 patient population.

25 So if masses and microcalcifications is 40

1 and 40 and the rest is the last 20 percent, then the
2 case mix would be like that, the enriched dataset
3 would be like that. And so then if you get overall
4 performance, then that's what that indication is
5 stating is that yes, it was able to do it across that
6 whole mix. And if that is the current standard and
7 it's acceptable, it'll go forward.

8 But we then just see -- look at those two
9 important subgroups. It may not be statistically
10 based, but you could build some inference from it.
11 You could do that or you could ask for additional
12 standalone performance to provide better information
13 on those two subgroups.

14 DR. D'ORSI: I think that's fair. What
15 does everybody think, that we really don't have to do
16 a reader study for each one of those, that we can get
17 that information by standalone and even stressing it?
18 Does everybody feel comfortable with that? Okay.
19 Okay, let's go to the next point.

20 MS. MORRIS: Okay. So I think we'll go to
21 the next major question, and that talks about the
22 clinical data. And we had a lot of discussion in the
23 presentations, as well as among Panel members, about
24 issues of reuse of data, what's the appropriate
25 control, the endpoint, and which endpoint should

1 power the study. And then there are some questions
2 about reader characteristics.

3 So this is covering Questions -- the reuse
4 of data is covered in 3(a). And basically, as the
5 result of the last Panel meeting, we heard two
6 different messages, that under reasonable conditions,
7 it might be appropriate to reuse data versus no, you
8 never reuse data. It always has to be a new dataset.
9 So we want a little bit more discussion. If you can
10 bracket if it's possible to reuse data, what would be
11 a reasonable approach? Or if you all agree that it
12 should always be a new dataset. So we'll start with
13 3(a). And I think, Robert, can you read it?

14 DR. OCHS: Yeah, I'll read 3 first.

15 MS. MORRIS: Okay.

16 DR. OCHS: So the broad number 3 question
17 is: Under Section 4 for the Clinical Performance
18 Assessment draft guidance, the Agency describes
19 considerations regarding the use of sample
20 enrichment, study endpoints, and reader
21 characteristics. Has the Agency provided sufficient
22 clarity of its expectations for what constitutes a
23 scientifically sound study? To assist in the
24 discussion, we have provided the following the
25 questions and examples.

Free State Reporting, Inc.
1378 Cape Saint Claire Road
Annapolis, MD 21409
(410) 974-0947

1 So Question 3(a) deals with data reuse. In
2 a 510(k) submission to support clearance of a CAD
3 device, are there circumstances where test data can
4 be reused in standalone performance or clinical
5 assessment? If so, what type of constraints do you
6 recommend on this reuse of data?

7 For example, in a CAD 510(k) device
8 submission, the manufacturer sequestered the test
9 dataset and only used it once to support clearance of
10 a CAD device. This CAD device is now the predicate
11 in a 510(k) submission of the same manufacturer's new
12 device. Can test data be reused to support the
13 clearance of the new CAD device? If so, what are the
14 constraints you recommend on this reuse?

15 DR. D'ORSI: Okay. Yes, Dr. Glassman.
16 They should turn there automatically.

17 DR. GLASSMAN: Yeah. If the cases were
18 truly sequestered and there was no teaching of the
19 company or the device based on those, it could be
20 reused. And I don't think it could be reused over
21 and over and over again, but at least once, I think,
22 would be safe.

23 DR. D'ORSI: Let's separate the standalone
24 from the clinical. On a standalone, can you use the
25 same dataset over and over? Standalone. No reader

1 involvement.

2 DR. GLASSMAN: I would say that it's the
3 same problem, not over and over and over again. I
4 think the burden of developing a standalone set of
5 cases is much less than the burden of developing a
6 clinical set of cases. But I think if the stuff has
7 been sequestered, a second time is probably not
8 unreasonable and then either a complete washout or
9 some kind of graded, you know, 25 percent every time
10 you use it and the 25 percent are in the new -- in
11 the next iteration, something like that.

12 DR. D'ORSI: Dr. Dodd.

13 DR. DODD: Yeah. So I mean, from a
14 statistically pure standpoint, one would say you
15 cannot reuse data, but I recognize the complications
16 with getting these datasets. So I'm open to the type
17 of proposal that Petrick was -- Nick Petrick was
18 proposing with a limited reuse of data with no
19 feedback about the specific lesion types that maybe
20 the algorithm failed on. I would also expect that
21 there are going to be cross-validated estimates of
22 the performance from the test and training set, which
23 is approximately an unbiased estimate of the overall
24 performance. Now, I'm not assuming that we're in a
25 situation where you have a gazillion features.

1 And so one could compare those estimates to
2 the estimates that were obtained from reusing the
3 data a single time. I think you should limit the
4 amount of time that you -- the number of times you
5 can reuse the data. I also question whether you
6 should incorporate some type of multiple comparison
7 adjustment, you know, with the cap of the number of
8 reuses you're allowing.

9 DR. D'ORSI: Dr. Leitch.

10 DR. LEITCH: It may be self-evident, but I
11 just would say that the dataset, if it were on film-
12 screen mammography originally, if it's a new device,
13 should be on digital, full-field digital mammography.
14 So in that circumstance, I don't think you should
15 reuse a dataset that would be film screen.

16 DR. DODD: And following along that, one
17 other thing I wanted to mention, I would assume that
18 these databases have some natural expiration date.
19 As the technologies change, you know, the databases
20 are no longer going to be relevant.

21 DR. D'ORSI: Dr. Jiang.

22 DR. JIANG: Yeah, the other comment I want
23 to add is this also depends on the size of the
24 database. If the database is very large, then to me
25 it's advantageous to reuse it. But if the database

1 is small, you don't want to reuse it too often.

2 DR. D'ORSI: Dr. Bourland.

3 DR. BOURLAND: It seems like if -- to wrap
4 some of these together, the image quality is
5 appropriate to where things are now. The data have
6 not been used as a learning dataset for the
7 algorithms involved and things like that. And then I
8 think the term you used, Dr. Glassman, is
9 sequestered, meaning not corrupted in any way or, I
10 don't know, somebody drew on it, so to speak. And
11 the distribution of disease is relevant.

12 Then, actually, there's a great beauty to
13 having that dataset because it was used previously,
14 and then to test other devices on it actually tells
15 you -- gives you a reference, albeit it's a floating
16 one, and I think that's your concern. Can you reuse
17 the thing? So there's a little bit of a danger to
18 using it, but there's a great beauty of it just to
19 test certain aspects of your equipment. But in that
20 sense, maybe then you have to have databases that
21 overlap so that you can move as things change, so to
22 speak.

23 DR. D'ORSI: Mr. Uzenoff.

24 MR. UZENOFF: I think Dr. Petrick showed
25 some exciting possibilities that I wasn't aware of

1 before today, about possible reuse. If we were
2 thinking prior to this, is a binary decision reused
3 or not reused? And I think he showed us some
4 techniques about dealing with it. I think
5 Dr. Glassman holds out even the possibility for
6 limited reuse maybe without having to deal with some
7 of these things. So I think that's a very important
8 area that we need to learn more about during this
9 comment period. But yes for reuse.

10 DR. PAYNE: Yeah, just one comment, and I
11 guess just to reiterate what I think I've heard from
12 some others, is I think the manufacturer, in
13 demonstrating to the FDA, needs to indicate how they
14 feel that the database is appropriate in light of
15 technological advances. I mean, matrix sizes
16 increased and so forth and so forth. So I think if
17 they address these issues and then say that the
18 database is appropriate, then I think that would be
19 helpful.

20 DR. D'ORSI: So is it -- oh, I'm sorry,
21 Dr. Carrino.

22 DR. CARRINO: So I guess the comment would
23 be on the FDA's side. It would be for them to look
24 that it's an appropriate database, meaning not too
25 small. And, you know, if somebody went through the

1 trouble to get a very large database with the right
2 spectrum and high image quality, then it would be
3 suitable. So it has to be, you know, the suitability
4 of it as well for reuse.

5 DR. D'ORSI: So just on the standalone now,
6 if I'm hearing correctly, is it fair to say that if
7 we have a database that is sequestered, that is not
8 used for training a new device, that has a good mix
9 of findings, that this can be used again as a
10 standalone?

11 DR. DODD: With a limited number of reuses,
12 not again and again and again.

13 DR. D'ORSI: And why is that? I'm not
14 clear on that. If it's not being used to train the
15 machine, the machine you're doing, and it's not
16 going -- and we're not talking about an interpreter
17 interface, wouldn't it actually be a plus to have --
18 to really get equality between two units?

19 DR. DODD: So what I want to see is an
20 unbiased estimate of the performance of this
21 algorithm.

22 DR. D'ORSI: Right, let's say it happens.
23 Oh, I see what you mean.

24 DR. DODD: And if I do it again and again
25 and again, I'm inadvertently potentially training my

1 algorithm to the independent validation set.

2 DR. D'ORSI: Okay.

3 DR. DODD: That's the concern.

4 DR. TOURASSI: Yeah, Dr. Dodd is correct
5 because if you test once and you watch where the
6 algorithm failed, that is knowledge you acquire to go
7 back and improve the algorithm to work better with
8 these signs.

9 DR. D'ORSI: No, no, I'm saying it's not
10 used to correct.

11 DR. TOURASSI: So it's not going to train,
12 but the knowledge you get from where the algorithm
13 fails in the test set --

14 DR. D'ORSI: Oh, okay, that's a good point.

15 DR. TOURASSI: -- is the bias you bring in
16 the development. So if you do it a few times, it's
17 okay, but you cannot keep doing it.

18 DR. D'ORSI: Okay, that's a good point.
19 Dr. Jiang.

20 DR. JIANG: I guess I agree with
21 Dr. D'Orsi. I think you may be clear, but I'm still
22 confused. I'm not sure why there is bias. So
23 there's two issues. We're not doing a statistical
24 test here. So I don't really understand why there's
25 a multiple comparison issue here. And the second

1 point is, assuming we have a very large database
2 that's pretty much representative, then if
3 inadvertently you get your CAD technique, standalone
4 technique, to learn that dataset, isn't that what we
5 want? And it depends on the database. If really we
6 can get to that stage, I think that's a good thing to
7 happen.

8 DR. DODD: So the question is how
9 generalizable are those results? If we train it to
10 that database, you know, if that's our complete
11 population, sure. But we want to generalize it to
12 the next set of patients. And so that's where we're
13 concerned that we've over-trained it to this
14 particular dataset and our estimates of the
15 predication accuracy, error rates, whatever, may be
16 over-fit to that dataset.

17 And I guess I was assuming we are doing
18 some kind of informal statistical testing. By that I
19 mean, you know, confidence intervals, right? So our
20 coverage probability -- I mean, our estimate is
21 biased and therefore our coverage probability is not
22 correct either.

23 DR. JIANG: Can I just quickly respond
24 before -- I think I understand what you're saying
25 now. I want to clarify that we're not really

1 training, explicitly training the CAD, but I do
2 understand your point because you're measuring the
3 performance on a single test, that there's a
4 potential bias there.

5 DR. D'ORSI: Yes, Dr. Ziskin.

6 DR. ZISKIN: If the database is
7 sufficiently large like, for example, that was
8 proposed earlier today as a possible large database
9 from which we could pull, and that when any test
10 comes up, you would actually take a sample from that,
11 not use the entire database for any study, but it's
12 the basis upon which you sample. And then each
13 sample would be different, although it's still the
14 same, you know, total database.

15 I think that helps a great deal. When
16 you're dealing with small sets, one of the things we
17 used to do is that type of procedure, is you take all
18 but one and then you do the study, and then you slip
19 a different one out and do that in a random manner to
20 try to get the best estimate in the future with a
21 relatively limited database.

22 DR. BOURLAND: Comment.

23 DR. D'ORSI: Yes, Dr. Bourland.

24 DR. BOURLAND: So the answer is we'd like
25 everybody to get an A and we want it to be the A that

1 covers everything, and we're not quite there yet
2 because we don't have that database. But maybe that
3 can be built. And then everyone has their own
4 databases, anyway. A lot of times when devices come
5 out to market, the first thing that's done is someone
6 takes it and uses it on their database. So at least
7 there are things that happen. Yes, that's outside of
8 FDA's sort of purview in terms of what happens once
9 things have come to market, but maybe having a few of
10 these reuses while things are being built for the
11 larger database allows things to go forward.

12 DR. D'ORSI: So I'm hearing that reuse
13 standalone in a controlled dataset is okay over some
14 small finite number. Is that fair? Is that enough?
15 Okay, let's go to the next point.

16 MS. MORRIS: You talked about standalone,
17 but for the reader study, would it be a different
18 answer? Yes or no?

19 DR. GLASSMAN: For the reader study you
20 either have to vary the cases or vary the readers,
21 and if you have a different cadre of readers, you
22 could use the same case dataset for a comparison, at
23 least for a limited number of times.

24 DR. D'ORSI: I mean, I like what
25 Dr. Petrick, I think, showed this morning with the

1 dropout of a block from here and adding a new one and
2 then moving that down after a certain number of
3 reuses. I think that really is nice. It doesn't put
4 the burden on you of replacing the whole dataset, but
5 it does give you the variability by taking a group
6 out and putting another group in. How do people feel
7 about that? Okay.

8 DR. ABBEY: For reader studies, I'm a
9 little -- I don't know about that. If you're going
10 to give the -- you're going to give the reader -- if
11 it's the same reader looking at the same case they
12 looked at before, without a whole lot of time, I
13 don't know about the consequences of memory effect
14 and things like that.

15 DR. D'ORSI: Let's say there is, you know,
16 four to six weeks in between, you know, so the memory
17 effect is out.

18 DR. ABBEY: And is this just the second
19 time you've done it or is the thirteenth time you've
20 done it or the --

21 DR. D'ORSI: I don't know enough of what
22 was presented this morning. Maybe you can quickly
23 reiterate that replacement, Dr. Petrick. It seemed
24 very nice.

25 DR. PETRICK: Can I pull up the slides

Free State Reporting, Inc.
1378 Cape Saint Claire Road
Annapolis, MD 21409
(410) 974-0947

1 here? Yeah, I can go through it. And I guess I was
2 not necessarily saying that was for -- do you know
3 which one it is, Robert?

4 DR. ZISKIN: It would depend a lot on
5 whether the reader was told whether he was correct or
6 not.

7 DR. PETRICK: So the layout that I proposed
8 was sort of -- oh, go ahead.

9 DR. DODD: I was just going to say, I'm
10 assuming if it's a 510(k) submission, that it's not
11 six weeks, it's two years, and it's not 13 times,
12 it's maybe once or twice. Yes, well, maybe it's
13 seven, excuse me.

14 DR. JIANG: Can I just quickly agree with
15 what Dr. Ziskin just said? Whether you can reuse
16 that dataset in a large part depends on whether the
17 reader was told the truth or not.

18 DR. PETRICK: So what I laid out is this
19 generic scenario. I wasn't actually differentiating
20 between -- particularly between what the readers
21 would be doing and standalone. And, in particular,
22 for standalone, this is, I think, you know, at least
23 in my opinion, a viable option, where for standalone
24 you're going to rotate through the dataset. Some
25 cases will fall off after some use. New cases will

1 be added at some percentage. And then there'll be an
2 random sampling from that dataset as you're going
3 along. So for standalone, to me that seems like a
4 fairly clear approach.

5 For when the clinician's involved, this
6 question about what -- in this particular case, if
7 you limit the information going back to the
8 developer, you have some control about how much
9 information is received. For the clinician, it's a
10 lot more complicated. What information did they
11 glean from the cases they saw before?

12 Now, Dr. Glassman pointed out this
13 situation. If you get new readers in a study, they
14 wouldn't have obviously seen the cases before. If
15 it's the same readers in the study, then I think it's
16 obviously a little bit more of a complicated
17 scenario, and I didn't really have a solution for
18 that, but it certainly is legitimate to think about
19 new cases.

20 Typically, in the reader studies, the
21 number of cases is going to be smaller, so
22 potentially companies will have a larger selection
23 base to go through to actually do different studies.
24 But that may not always be the case. Typically, you
25 have a larger database for standalone compared to

1 what you actually need for the reader study.

2 DR. D'ORSI: Okay. So I was a little
3 confused. I thought it was related to readers. So
4 is it fair to say the answer is yes, you need a new
5 dataset for readers in some way, a new set of cases
6 when you introduce an observer or not?

7 UNIDENTIFIED SPEAKER: Not necessarily.

8 DR. D'ORSI: Okay. If you have the same
9 reader and the same dataset and different readers and
10 the same dataset?

11 DR. DODD: And what kind of feedback have
12 they gotten from the first reader study? I mean,
13 we're assuming they don't know the truth?

14 DR. D'ORSI: Oh, are you talking to me?

15 DR. DODD: Yeah, I'm talking to you.

16 DR. D'ORSI: I'm looking over there. I'm
17 sorry. Go ahead. What did you say?

18 DR. DODD: Are we assuming the readers got
19 feedback? Do they know what the truth was in that?

20 DR. D'ORSI: Yes.

21 DR. DODD: They do?

22 DR. D'ORSI: Yeah.

23 DR. DODD: Okay. So that's a different
24 question then, right?

25 DR. D'ORSI: Right, right.

1 DR. DODD: If they don't know the truth and
2 seven years has passed --

3 DR. D'ORSI: Then it's okay. Right. Yeah,
4 even if they do know the truth, in seven years -- I
5 can't remember seven minutes.

6 MS. MORRIS: Can I offer -- and I don't
7 want to confound things.

8 DR. D'ORSI: No, no, please offer.

9 MS. MORRIS: But is it reasonable, like for
10 the reader study, that you have the same dataset but
11 you're changing the readers? Which is harder,
12 changing the data or changing the readers, or do you
13 have to do both?

14 DR. D'ORSI: They're about equal.

15 MS. MORRIS: They're about equal. Okay,
16 never mind.

17 DR. D'ORSI: Have you tried to get readers,
18 different readers?

19 DR. GLASSMAN: You just have to pay enough
20 money.

21 (Laughter.)

22 DR. D'ORSI: Mr. Uzenoff.

23 MR. UZENOFF: I think Dr. Glassman has a
24 good point and I think -- and Dr. Abbey threw out
25 some variables, you know, as did Dr. Dodd. What did

1 the reader learn about the study? And, you know, if
2 it's used twice and there's four weeks or eight weeks
3 in between it, I think it's okay. If the reader
4 hasn't seen it in a year or two years. So I think we
5 should be careful about being too proscriptive about
6 how to use it. Maybe that's part of the negotiation
7 with the Agency, to explain why the reuse scheme was
8 valid and try to define it. But I think it's okay to
9 reuse it with some conditions.

10 DR. D'ORSI: Is that enough?

11 MS. MORRIS: Yeah. Primarily we wanted to
12 know if there was a window of opportunity in which to
13 reuse data and whether we should explore what those
14 conditions should be. And we can take it from there
15 and see what kind of constraints we would put on
16 that.

17 DR. D'ORSI: Okay, great. Let's go to the
18 next one.

19 MS. MORRIS: Okay. So I want to jump to
20 the powering the study based on which endpoints,
21 which is 3(e). Robert, could you read 3(e)?

22 DR. OCHS: Okay. Okay. Manufacturers
23 typically report Receiver Operating Characteristic
24 curves, Area Under the Curve, Sensitivity and
25 Specificity for their clinical performance studies.

1 Should the studies be powered for all summary
2 endpoints? If not, which endpoints should be used to
3 size the study?

4 For example, a clinical performance
5 assessment for a breast CAD device could be powered
6 for AUC based on a radiologist's reported probability
7 that an image contains a malignancy, or it could be
8 powered for sensitivity and specificity based on a
9 cut point in the BI-RADS scale.

10 DR. D'ORSI: Dr. Dodd.

11 DR. DODD: I knew this was coming. Oh, do
12 I have to answer this?

13 DR. D'ORSI: Yes.

14 DR. DODD: This is a tough one, you know,
15 and I think, personally, I see advantages to
16 sensitivity and specificity as endpoints. I also see
17 the advantages of ROC because what we heard this
18 morning, which I agree with, when you're looking at
19 sensitivity and specificity, the relative impact of
20 the tradeoffs has to be considered with some kind of
21 utility analysis. And utility functions, in my
22 opinion, are always so relative to an individual that
23 it's hard to assign a common utility function, unless
24 you're talking from a policy level, and that's not
25 the framework I think the FDA has to work under.

1 So I would say that any of them are valid,
2 ROC, AUC, and then sensitivity and specificity
3 considered jointly. I don't think I would advocate
4 powering for all of them. I think that's probably
5 overkill. I would certainly analyze each one of
6 them, but I would specify one as a primary endpoint.
7 So I don't know. Is that enough of an answer?

8 MS. MORRIS: Well, we are asking industry
9 to report on all of them. So the question is
10 powering the study will depend upon which one we use
11 as the primary.

12 DR. DODD: So let me ask this question,
13 then. What am I powering for? Am I powering to
14 detect a difference between the predicate CAD in,
15 say, ROC and I want to prove that there's a
16 statistically significant improvement in ROC and AUC
17 and sensitivity and specificity? Or is it going to
18 be enough to say I have a significant improvement in
19 my area under the curve and my sensitivity and
20 specificity at this cut point is such and such and
21 here's a confidence bound for those things?

22 MS. MORRIS: Would it make a difference?
23 Maybe if we start with a brand new CAD. Is it more
24 important to the user to see the -- to power it to
25 sensitivity and specificity and then just have

1 interpretation of the other endpoints? Or is it
2 adequate to use some combination, particularly area
3 under the curve, like a partial area under the curve,
4 as a Dr. Petrick talked about?

5 DR. DODD: And, again, this is for the
6 clinical performance studies. We're not talking
7 about standalone. So I think that you could get by
8 with -- I see advantages to both, and I hesitate to
9 recommend one over the other. If they're in the
10 labeling, I think, at a minimum, you should -- you
11 could have a primary endpoint, such as ROC or AUC,
12 and then simply provide confidence intervals for
13 sensitivity and specificity in the label, and then
14 it's up to the user to understand that. But I don't
15 know if anybody has another comment.

16 DR. ABBEY: So I had one thing. Just
17 thinking back to this morning, I saw in a lot of
18 these enriched cases false positive rates of 45
19 percent or something like that and so -- you know,
20 which makes sense for an enriched case set. But how
21 do you interpret that unless you have some correction
22 for the enrichment or something along those lines?
23 I'm not sure that those terms are -- and we also know
24 there's going to be a change in sensitivity as well.

25 DR. DODD: So I guess I thought we were

Free State Reporting, Inc.
1378 Cape Saint Claire Road
Annapolis, MD 21409
(410) 974-0947

1 talking about the setting of an MRMC study which had
2 a representative sample of cases.

3 UNIDENTIFIED SPEAKER: Yeah.

4 DR. DODD: So it wouldn't be a biased
5 estimate of these parameters, which is separate --

6 MS. MORRIS: Right, but it's still
7 enriched, but it's --

8 DR. ABBEY: Yeah.

9 MS. MORRIS: -- proportional to --

10 DR. ABBEY: Prevalence is much higher
11 than -- right, prevalence won't be .05 percent or .5
12 percent or something like that.

13 DR. DODD: So unless you're saying that it
14 affects reader performance because of the knowledge
15 of prevalence --

16 DR. ABBEY: How do you interpret the false
17 positive you get from an MRMC study that's got 50
18 percent malignant cases versus -- and 50 percent
19 benign? What do you with that when you go back to --

20 DR. DODD: I think it's still an open
21 debate about whether the knowledge of the prevalence
22 and the enrichment affects reader performance. From
23 a statistical perspective, I'm conditioning on
24 disease state. So if it doesn't affect reader
25 performance, then my analysis is unbiased. But the

1 extent to which a radiologist's knowledge of the
2 prevalence and the enrichment affects their
3 performance, that could potentially bias the result.

4 DR. ROSENBERG: Yeah, I had another
5 comment, and I'm curious what Carl and Lenny, when he
6 gets back, would say. When we read the screens with
7 or without the CAD, it's binary. It's a callback or
8 it's not a callback. So sensitivity and specificity
9 seems to work the best. For a CADx device, it might
10 make sense to use a full scale, an area under the
11 curve, but I always worry about that we make these
12 binary decisions and then you create a curve over it.

13 DR. D'ORSI: Yeah, that's what I was going
14 to point out. The sensitivity/specificity is closest
15 to the clinical use of this device, and you're right,
16 you'll probably get different pieces of information.
17 But I think, in my own opinion, the most important
18 information that I would use clinically was this
19 binary sensitivity/specificity cut.

20 DR. DODD: Okay. So I hadn't appreciated
21 that fact. I was assuming there was a BI-RADS scale
22 on top of this.

23 DR. D'ORSI: No, not this.

24 DR. DODD: So if that's the case, then
25 sensitivity and specificity would make the most

1 sense.

2 DR. ROSENBERG: People can read them with,
3 you know, a scale of one to five for concern for
4 malignancy, and I think they've done that. But I
5 think when we read the mammograms clinically, it's
6 really binary.

7 DR. D'ORSI: Yeah, Dr. Jiang.

8 DR. JIANG: I'd like to add some comment to
9 that.

10 DR. D'ORSI: I noticed.

11 DR. JIANG: If I understand this correctly,
12 I think ROC offers you a lot more power than just
13 measuring sensitivity and specificity. You know, we
14 looked at data from the consortium data, you know,
15 over two million mammograms, and we were trying to
16 look at a difference in terms of cancer detection
17 rate. That's sort of a similar surrogate to
18 sensitivity. And what we found is it's extremely
19 difficult to do that.

20 So I think my question would be whether
21 that's practical. So, you know, I like ROC analysis,
22 but I recognize ROC analysis has this, you know,
23 question: How does that link to clinical practice?
24 Right? Because I think that -- but the question is,
25 if we actually go for measuring sensitivity, is that

1 feasible? Do we need a very large study? Can a
2 single manufacturer do that?

3 DR. D'ORSI: The problem -- yeah, that's a
4 good problem. You have three of the four data points
5 you need for sensitivity. The one that's going to be
6 hard is if both you and the machine miss a finding.
7 How are you going to know that? If you're
8 calculating sensitivity with and without the machine
9 and the machine misses something and you and the
10 machine miss something, you're not going to know that
11 unless you either look in a year, when she comes back
12 to see if there's a finding, and then assign a false
13 negative or true negative statement to that. So it's
14 strictly not -- you're right, it's strictly not a
15 sensitivity. Do you know what I'm saying?

16 DR. JIANG: I think you already have truth,
17 though. When you start the experiment, you already
18 know. Isn't that the case?

19 DR. D'ORSI: That's true. Yeah, I'm
20 thinking clinically here all of a sudden, I'm sorry.
21 Yeah, I'm sorry.

22 DR. DODD: Can I just get some
23 clarification, then, from the FDA, since this
24 proposal -- what I've heard is that the clinical use
25 is going to be detection or not detection. Is the

1 alternative -- I mean, when we're talking about ROC
2 analysis, is one proposal to impose some kind of a
3 BI-RADS curve or some kind of a rater curve on top of
4 what is going to be done clinically so that we can
5 get an ROC analysis? It seems to me that's more a
6 shift away from clinical practice than I had
7 previously appreciated.

8 DR. GWISE: Hello, I'm Tom Gwise. I think
9 the intent is to do all of the analyses, the ROC
10 curve, look maybe at probability of malignancy or a
11 BI-RADS scale. But the question is which endpoints
12 to actually size the study for. And the tradeoff
13 between sensitivity and specificity, in my opinion,
14 is what's really important here. If we're powering
15 the study for the full ROC curve, how much influence
16 is that high false positive rate area playing? So
17 does that help?

18 DR. DODD: Yeah, thank you. I don't know
19 if Nick --

20 DR. PETRICK: So I just wanted to make a
21 couple clarifications. Typically we're going to talk
22 about -- so Craig was asking about what do you do
23 with a particular operating point that you find? And
24 I would just say, typically we're going to do a
25 comparison study. So we're going to compare unaided

1 to aided reading or Device 1 to Device 2.

2 The absolute numbers, while they hopefully
3 would have some meaning, if there's a stress test or
4 a significant enrichment in the study, they may not
5 actually meet clinical practice. That's what I tried
6 to point out this morning. So what the information
7 is, is that comparison, how do these two compare to
8 each other?

9 With a sensitivity/specificity endpoint
10 with a CAD, we'll typically see this tradeoff in
11 sensitivity for specificity. And then, obviously one
12 of the -- I guess the approach that would advocate --
13 that I'm advocating is the ROC is important. We look
14 at variabilities and thresholds. The readers have
15 that. That's a real effect. And if we're ignoring
16 that effect or we're trying to average across that
17 big effect, that increases the size of the study, and
18 does it really provide more information? It clearly
19 is more clinically what clinicians are used to doing,
20 but again, we're talking about studies that are
21 smaller and controlled, and is that really the
22 appropriate use? And that's really the question, I
23 think, for the Panel.

24 DR. DODD: So I guess would follow up on
25 that, though. There is one situation in which the

1 ROC analysis does -- an AUC comparison does give a
2 misleading interpretation, and that is when the
3 curves cross, right? So I was wondering, from the
4 presentations earlier this morning, if you did -- if
5 you were concerned about that, if you could specify
6 a priori some false positive threshold so it would
7 kick over to a partial AUC analysis under that. But
8 that would have to be done a priori.

9 DR. SMITH: If I could just provide a
10 clinical comment on that. I think what Dr. Gwise was
11 saying as far as powering the studies, it doesn't
12 make sense to me clinically, and I agree with
13 Dr. Rosenberg's comment that, you know, certainly for
14 screening mammography, it's 0, 1, or 2. You could
15 force someone in a trial to use the full
16 BI-RADS scale, but clinically it's really a 0, 1,
17 or 2, and 1 or 2 has the same clinical action, so
18 it's really a binary task. And if you used ROC,
19 you'd have to use the clinically relevant portion of
20 the curve, which is relatively narrow.

21 The recommendations are to have the recall
22 rate between 5 and 10 percent. So you'd have to use
23 a relatively narrow portion of the curve, and I'm not
24 sure the number of -- the sample size would really
25 differ that much if you used just partial AUC. So

1 you may want to consider that. And as a clinician,
2 it's much easier for us to understand the sensitivity
3 and specificity.

4 DR. PETRICK: Can I just add one more?
5 This is Dr. Petrick again. I just wanted to comment
6 on Dr. Dodd's comment. What we, I guess, would
7 advocate is looking at pilot study data. We have
8 seen some curves that cross. We haven't typically
9 seen a whole of ROC curves that have crossed, at
10 least the data that I've seen coming in to the
11 Agency. But when that would happen, then obviously
12 you'd have to pick at some other measure, some
13 partial area or some other way of looking at the
14 area. And I guess what we would advocate in all of
15 these experiments is to run pilot studies, both to
16 understand what endpoint might be appropriate for the
17 data as well as trying to train the readers and make
18 sure they understand how to actually participate in
19 the study. And this is a huge problem with the
20 studies. They don't know how to either rate the
21 individual cases or don't know how to use the scales,
22 and this causes all kinds of complications that
23 really probably is unnecessary with better training.

24 DR. D'ORSI: So taking all of this into
25 consideration, what is the recommendation between ROC

1 -- as far as powering your study between an ROC and a
2 sensitivity/specificity study? Is that the question
3 you want to know or answer?

4 MS. MORRIS: Yeah, but I think that you
5 gave us some options so it doesn't have to be a
6 binary scale or an absolute answer of one versus the
7 other. It sounds like we could use some options,
8 that if we have some pilot data that gives us some
9 predictability of whether or not the curves are going
10 to cross, then a priori the study would have -- for
11 sensitivity --

12 DR. D'ORSI: Right.

13 MS. MORRIS: -- and specificity. So we
14 could put a few conditions and give industry options,
15 and then it's up to them to convince us it's the
16 right way to go. Unless the Panel feels that it
17 should clearly be sensitivity/specificity, or if you
18 could define a partial area under the curve that you
19 feel is clinically meaningful, then we can do that.

20 DR. D'ORSI: Lori.

21 DR. DODD: So, again, I just want to
22 emphasize, I do think sensitivity and specificity are
23 relevant quantities. The concern I have comes in
24 with the relative tradeoffs and assigning the utility
25 function. So if you're on a different ROC curve,

1 then I feel more comfortable. But if I have an
2 improvement in sensitivity and a detriment in
3 specificity, then I have -- the interpretation of
4 that becomes a little more problematic.

5 So I see the advantages and disadvantages
6 to both, but I don't know, particularly at the level
7 the FDA is approving things, you know. I mean, I
8 think if we were talking for a policy level, the
9 discussion would be entirely different. And also
10 taking in light the least burdensome approach.

11 DR. D'ORSI: Do you have enough to deal
12 with or do you need a little more discussion? Let me
13 ask a question, then. Maybe this'll help. How would
14 you train a reader to do an ROC analysis on a
15 recall/no recall setting? What instructions would
16 you give the reader? Here's a set of cases and rate
17 them 1 through 5, 1 being X and 5 being Y, and then
18 2, 3, 4 in between. What gray scale would you give
19 them for an ROC curve?

20 DR. DODD: Can we turn that over to -- that
21 falls out of my expertise.

22 DR. PETRICK: So I can give you a couple
23 scenarios. So one question would be, are these -- if
24 there was something about the lesions you're going to
25 call, I'm going to recall, or I'm not going to do

1 anything with, is there some differentiation between
2 those? Can you break that group up into smaller
3 pieces or some borderline? You know, you're sort of
4 sure they're normal but not 100 percent sure. Or
5 you're sort of sure they're abnormal but you're not
6 100 percent sure. And there's another group that's
7 clearly abnormal. If we can get some more
8 differentiation in those groups, that allows us to
9 build the ROC curve.

10 DR. D'ORSI: Okay, but what are you
11 going -- what kind of features are you going to give
12 them? Are you only talking about malignancy features
13 or are you talking a finding? In other words, are
14 you asking them what's your confidence that there is
15 a finding that requires recall versus what is your
16 feeling that there is a malignancy in here that needs
17 a recall?

18 DR. PETRICK: So it all depends on the
19 question. If you're talking about colon, you're
20 talking about polyps, obviously.

21 DR. D'ORSI: Um-hum, right.

22 DR. PETRICK: You wouldn't necessarily be
23 talking about malignancy. There may be a confidence
24 score or a comparison between cases, of which case is
25 more likely to be a polyp or not. If clearly it's

1 suppose to be a cancer or not, it may be what's more
2 likely to be a cancer or not, if that's the
3 particular endpoint of the particular CAD device.

4 So it's hard to say there's a definitive
5 differentiation, but the idea is not -- is to take
6 that binary decision that the clinician is making and
7 try to get into and probe whether there's a little
8 bit of grayness to that decision. Is it clearly
9 every case that's left to right, or are there some
10 incremental decisions that are made within that while
11 the clinician's making the process? And can we glean
12 that information out to, again, the big advantages?

13 This threshold variability is huge between
14 different readers. And if that's incorporated into
15 the statistic, it obviously increases the sample
16 size, but it also leaves us with these tradeoffs in
17 what is the right utility for trading some
18 specificity for increased sensitivity or vice versa.
19 And so it's really to try to glean some more
20 information out of it. Again, it's not the complete
21 clinical task. But more than likely, every clinical
22 task involves some grayness in decision. So I'm
23 thinking about what's the likelihood of this being a
24 lesion or not, and it's not always completely 100
25 percent one way or another.

1 DR. D'ORSI: Okay, that makes a little
2 clearer. If you're going to ask them of their
3 confidence that something is present that needs
4 further evaluation, that makes sense to me, if you're
5 going to ask that question. Bob.

6 DR. ROSENBERG: My experience is
7 mammography, so outside of mammography, I wouldn't
8 comment on that, so in terms of colon in particular.
9 But if the readers are using the device the way it's
10 supposed to be used, then the device then triggers
11 something new or it doesn't. And it's not like you
12 saw it, you dismissed it, the device identified it,
13 and now you're going to be recalling it because
14 that's not the way I think we're supposed to be using
15 those, at least the mammography devices. So that's
16 why I still see at least the mammography as binary.

17 DR. GLASSMAN: No. Carl?

18 DR. D'ORSI: Yes.

19 DR. GLASSMAN: A mammography reading
20 doesn't -- even though the outcome of screening is
21 binary, if you ask the question, you have identified
22 something, what is the likelihood it will end up
23 being a BI-RADS 4 biopsy versus what is the
24 likelihood that it will end up being normal? And
25 what is the likelihood it will end up being a BI-

1 RADS 3? Then you could develop an ROC from that kind
2 of likelihood of being malignant, in effect, scale.

3 DR. D'ORSI: I wouldn't use the malignancy.
4 I would use action because otherwise we wouldn't need
5 diagnostic workup. We could just say --

6 DR. GLASSMAN: Yeah.

7 DR. D'ORSI: -- all right, that's fine.

8 DR. GLASSMAN: Yeah, with the BI-RADS.

9 DR. D'ORSI: Yeah. Dr. Jiang.

10 DR. JIANG: I want to state my opinion
11 about this. One thing that's not appropriate as a
12 reading scale is the BI-RADS scale because the BI-
13 RADS are not necessarily a rating scale. I'm stating
14 this as an opinion here.

15 DR. D'ORSI: That's an excellent point, and
16 that's a lot of problems and a lot of studies that
17 are out and published, they're using BI-RADS as an
18 ROC scaler. And while it's ordinal, if ordered
19 correctly, it's not equal. So you get cockeyed
20 curves when you plot them. Yes.

21 DR. DODD: But I want to come back and
22 emphasize, though, that I think you should also --
23 because of these fine differences between how you're
24 going to estimate your ROC curve and how you're going
25 to ultimately use it in clinical practice, it seems

1 to me that in your reader study, you need to also
2 estimate sensitivity and specificity with the
3 intended use of the device and at a minimum provide a
4 confidence interval.

5 DR. D'ORSI: So it sounds like, clinically,
6 that's taking -- although the ROC is good to have in
7 a clinical manner, since that's the way it's used,
8 that that would be preferred, the preferred method.
9 Am I hearing this right?

10 DR. STEIER: Yes. And for a pulmonary CT
11 or a lung CT, specificity/sensitivity may be
12 confidence intervals, but I don't know what I would
13 do with the ROC.

14 DR. D'ORSI: Right. Is this enough info
15 now?

16 MS. MORRIS: Yes, I believe it is.

17 DR. D'ORSI: Good. Okay, let's go on.

18 MS. MORRIS: Okay, the next one I want us
19 to talk about is the control group, and that's in
20 2(a). So essentially what we're trying to determine
21 is what -- you know, depending upon what control
22 group you use will change the difficulty of the study
23 and the magnitude of the study. And so it's a matter
24 of, you know, what is clinically meaningful and
25 whether or not there would be a difference in

1 choosing the control arm. So, Robert, would you go
2 ahead?

3 DR. OCHS: Please discuss what you would
4 consider to be a valid control arm for such studies.
5 And this refers to the clinical performance guidance
6 and 510(k) and PMA studies.

7 So what would be the expected clinically
8 meaningful outcome to demonstrate that a new or
9 modified CAD device is substantially equivalent to a
10 legally marketed predicate CAD device?

11 And then Part (ii) is what should be the
12 expected outcome to demonstrate a reasonable
13 assurance of safety and effectiveness for a CAD
14 device subject to a PMA or a PMA supplement?

15 MS. MORRIS: So whether or not there should
16 be a distinction between those two groups and then we
17 have two examples. So (iii) is proposing that it
18 should be the control group. Because it's a 510(k),
19 it should be a predicate device that you would
20 compare. And whereas traditionally under PMA, it is
21 using unaided read as the control group. Would you
22 like us to read these two examples?

23 DR. D'ORSI: Anybody want them read? Okay,
24 let's open it for discussion, then. Len.

25 DR. GLASSMAN: It sounds reasonable that

Free State Reporting, Inc.
1378 Cape Saint Claire Road
Annapolis, MD 21409
(410) 974-0947

1 for (a)(i), that if the predicate is available for
2 use, that that would be the preferred control arm.
3 However, if the predicate is a different
4 manufacturer's unit or a unit that is no longer in
5 existence and there is no upgraded form that could be
6 used, then a reader -- an unaided reader study, I
7 think, would be acceptable because you have no other
8 choice. And whereas for (ii), it would be an unaided
9 reader study because predicate takes no part in the
10 PMA process.

11 DR. D'ORSI: I get a sense of how difficult
12 it would be to get a comparison predicate -- a
13 comparison to a predicate versus a standalone trial
14 if you had to deal with another manufacturer,
15 problems with proprietary information, datasets, et
16 cetera. Off the top of my head, it almost would seem
17 easier to just start and get a standalone study,
18 rather than to try and get all of these variables.
19 But I don't know. Can somebody answer or comment on
20 that? Yeah, Mr. Uzenoff.

21 MR. UZENOFF: Bob Uzenoff. One comment.
22 For Class III, I think we would want to allow, in a
23 case where -- I'm not answering your question, but in
24 a case where a predicate is easy, take a manufacturer
25 for a supplement for a Class III situation, should be

1 able to test against their own device as the
2 predicate rather than going back to an absolute
3 level.

4 MS. MORRIS: Yeah, as long as we don't
5 think that it's such a different device, you know,
6 that it's a modification and we need that clinical
7 performance, then. There are cases under PMA that
8 it's true, the device has to stand on its own, but
9 there are supplements that you're just referencing
10 from the original device.

11 DR. D'ORSI: Any other comments?

12 DR. DODD: I'm struggling here, and I'm
13 also getting tired, but equivalence is always a
14 tricky thing, and there's a concern that you have a
15 gradual drift downwards, right? You show equivalence
16 to some threshold and then you have another. So that
17 now becomes your predicate, and you bring in your
18 next CAD and you evaluate it to that, and then,
19 slowly, your overall performance is degraded.

20 So maybe I'm not thinking clearly here, but
21 it seems to me, I mean, this is a difficult question
22 and there are advantages to comparing it just to the
23 unaided read, right? And particularly if the
24 populations -- the samples you're comparing things
25 with are changing. So I don't have a clear answer,

1 but maybe somebody can help out.

2 DR. TOURASSI: You have to understand, the
3 predicate device is always the same, the first one
4 that went through the PMA. So even if the next one
5 comes in and it's some were slightly inferior but
6 within the acceptable statistical limit, it doesn't
7 become the predicate for the next one. Is that true
8 or not?

9 MS. MORRIS: Let me clarify. When we
10 usually talk about the predicate device, we're
11 talking about a 510(k) for a Class II device, and it
12 could be a different manufacturer's CAD. So a new
13 manufacturer wants to come in with their CAD device,
14 and they have to do an MRMC study, and they need to
15 compare it to a predicate. Well, it's another
16 manufacturer's device again.

17 Would we require them to do that and
18 conduct a study with that other device, or would we
19 allow them to use the unaided read as the control
20 arm? For a PMA, we've always traditionally allowed
21 them to come in and use the control arm as an unaided
22 read because it's just assessing the device by itself
23 and not comparing to a predicate.

24 DR. DODD: But once you approve something
25 under the 510(k), it could become a predicate for a

1 future CAD, right?

2 MS. MORRIS: Yes.

3 DR. DODD: Okay.

4 MS. MORRIS: You can have a predicate that
5 is your own device and you've modified it, or you
6 could have a predicate that is another manufacturer's
7 CAD. That's not clear?

8 DR. D'ORSI: No, it's clear.

9 DR. SMITH: Excuse me. This is Robert
10 Smith. Maybe I can help clarify that for you. I had
11 a slide where I showed you could have a direct
12 comparison for the 510(k) or you could do an indirect
13 comparison. But ultimately you have to compare
14 yourself to the predicate device. You have to
15 demonstrate that you're at least as safe and
16 effective to the predicate. You could do it directly
17 by a direct comparison or you can do your own unaided
18 read and then somehow make an argument that you're
19 substantially equivalent to the predicate device, but
20 you have to compare yourself to the predicate device.

21 DR. D'ORSI: That's a good point.

22 DR. ABBEY: But what's the predicate?

23 DR. TOURASSI: Yeah.

24 DR. ABBEY: Once you're substantially
25 equivalent, you are now eligible to be a predicate;

1 is that correct?

2 DR. TOURASSI: For the next one? Is that
3 how it goes?

4 DR. SMITH: Any legally marketed device
5 that's cleared through 510(k) can be used as a
6 predicate.

7 DR. ABBEY: So yes.

8 DR. D'ORSI: Yeah, Mr. Uzenoff.

9 MR. UZENOFF: Bob Uzenoff. So for a Class
10 II device, I think I want to put out, what would be
11 allowed would be a standalone test against the
12 predicate, using the same case mix.

13 DR. D'ORSI: Dr. Jiang.

14 DR. JIANG: Yes. So I guess I'm still
15 struggling with the same line of questions. So let's
16 go back to the predicate. The predicate label would
17 say something like the CAD device improves the reader
18 performance in cancer detection or something. So to
19 me, if you say that the second device is equivalent
20 to the first device, you would test on the same
21 question, whether that device improves the reader
22 performance in that same question. They were not
23 necessarily comparing to that device. I think this
24 is a unique situation. It's different from other
25 systems. When MR first came out, you can't compare

1 it to something before that. CAD is different. CAD
2 is an add-on. So the reader alone is always there.
3 You could always compare to that. But I don't think
4 you have that option with other devices. So maybe
5 this is a unique situation.

6 DR. D'ORSI: Can I get back to the point
7 that you must compare the Class II to a predicate
8 device and is that a difficulty if you do a
9 standalone study?

10 MS. MORRIS: Our decision making is to
11 determine substantial equivalence. So we can do that
12 in many ways. So if we didn't have direct comparison
13 data with device A and device B and we're dealing
14 with apples and apples, there are other means if we
15 know with certain confidence that some performance
16 data, whether it be clinical or bench, meets an
17 adequate performance compared to a predicate. It's
18 just what amount of certainty do you want.

19 So in other medical devices, we may just
20 ask them to do performance testing that meets some
21 kind of criteria and they don't do a side-by-side
22 comparison of it. But we know that it has this
23 acceptable performance, and that is enough to
24 determine substantial equivalence because we knew
25 what the performance was of the other device. And so

1 it's just a subtlety of whether or not in these
2 devices you need to have that side-by-side comparison
3 of the data.

4 DR. DODD: And I guess where I'm hung up is
5 that it's not clear to me which is a better
6 comparison, if I'm comparing to some predicate that
7 has less than -- you know, now I have, let's say,
8 sensitivity and specificity of 80 percent and 80
9 percent, whereas with -- you know, because I've
10 drifted down. So I'm comparing my new one to this
11 one with lower -- if I actually did an analysis
12 comparing it to the unaided read, you know, the
13 results would --

14 MS. MORRIS: I understand. Yeah, we
15 struggle with this all the time. If they decide to
16 choose and do a direct comparison with the predicate,
17 they will take the -- well, I would assume they would
18 take the predicate that has the lowest performance
19 because it would be easier for them to show that
20 they're equivalent. And whether or not that would be
21 done with unaided read, you can make similar
22 arguments because you're not comparing it to
23 something. You're just saying that it's, you know,
24 better than an unaided read.

25 DR. DODD: So let me ask this. How

Free State Reporting, Inc.
1378 Cape Saint Claire Road
Annapolis, MD 21409
(410) 974-0947

1 burdensome is it to require the unaided read
2 information in these MPMC studies?

3 MS. MORRIS: Actually I believe, but I'd
4 like perhaps Nick to comment on this, that the
5 unaided read study would be less burdensome.

6 DR. DODD: Or also the alternative, I
7 guess, is embedding it, as there would be three
8 reads.

9 DR. PETRICK: I guess if we're talking
10 about -- now we're talking about doing some sort of
11 clinical study, and we're assuming now that, for
12 whatever reason, we need to do these clinical studies
13 and we're not talking about standalone performance.

14 The complication is as Janine said. If you
15 wanted to go to the lowest common denominator, that
16 may have been a CAD approved 10 years ago. That
17 device is no longer on the market. If you're a
18 different company, that may not be a viable predicate
19 for you to ever get access to to use. So that's one
20 complication for doing those particular studies.

21 As far as doing the comparison, if you have
22 access to the particular device and you're doing a
23 direct comparison, that may be a non-inferiority
24 study, that may have a smaller sample size than
25 actually doing the unaided to the aided read to show

1 superiority. However, if you have no access to that
2 predicate device, then obviously being able to just
3 do it on your own dataset with your own cases to show
4 that you're improving reader performance, but not
5 necessarily doing a direct comparison to how much
6 your performance compares -- your improvement
7 compares to that predicate device improvement may be
8 actually less burdensome to do. Does that answer the
9 question?

10 DR. DODD: I think so.

11 DR. PETRICK: Okay.

12 DR. GWISE: Hello, this is Tom Gwise again.
13 I'd like to add something to what Nick just said.
14 When we're talking about substantial equivalence, we
15 still have to remember we need to show clinical
16 utility, and that's some increase in performance over
17 the reader alone. So we have to consider that.

18 And if we're going to compare a new CAD
19 device or a 510(k) CAD device to a predicate using
20 the label information, we have to remember that the
21 case mix is going to make that -- if they both had a
22 sensitivity of 90 percent, say, they're not really
23 comparable because of the case mix. So I just want
24 to be sure that's clear.

25 DR. DODD: Okay. So let me make sure I'm

1 understanding, though. But if you are able to
2 compare -- are you advocating comparing to a
3 predicate with an unaided read so that you're sure
4 you have the same case mix?

5 DR. GWISE: I would advocate having the
6 three modalities in the study, the unaided read --

7 DR. DODD: Right.

8 DR. GWISE: -- the predicate, and the new
9 device --

10 DR. DODD: Right.

11 DR. GWISE: -- so that you could make a
12 clear comparison and show clinical utility.

13 DR. DODD: Okay, thank you.

14 DR. SMITH: This is Robert Smith again.
15 Yeah, I completely agree with Tom. When we're
16 talking about a head-to-head comparison of the new
17 device to a predicate, the unaided read is going to
18 be done on both devices, and you're also going to
19 look at the aided read on both devices. And
20 obviously, if the new device showed no improvement
21 with the unaided read, that would raise some serious
22 questions.

23 In addition, I just don't want you to get
24 misled. Even if you're doing just an unaided read
25 for your new device, nothing stops you from going

1 back and making an argument that you're substantially
2 equivalent to a device that was cleared 25 years ago
3 and is no longer marketed. You can still try to make
4 that argument. So it's not going to make any
5 difference whether you did the unaided read or you
6 did the head-to-head comparison, you could still
7 compare yourself to any legally marketed predicate
8 device.

9 DR. DODD: And that's done just by looking
10 at your sensitivity and specificity values and
11 providing some assurance that the case mix was
12 similar?

13 DR. SMITH: Well, I'll give you my personal
14 opinion here. I think it would be extremely
15 difficult because for the reasons Tom said. If you
16 had different case mixes, different scoring
17 methodologies, all of those differences, the
18 manufacturer's going to have to somehow account for
19 those. It'd be very difficult to do that. Just if
20 you're going by what's on the labeling and you're
21 stuck with the case mix and methodology of the
22 manufacturer of the predicate device, it'd be
23 extremely difficult to do that. But we can't rule
24 out that possibility. You know, if you can make a
25 valid scientific argument, we'd have to consider it,

1 but I think it would be extremely, extremely
2 difficult.

3 DR. STEIER: So you have to be better than
4 the unaided read, theoretically, or you should be
5 better than the unaided read, but you don't have to
6 be better than the predicate? You have to be
7 equivalent or --

8 DR. SMITH: Well, I guess I'd ask you if it
9 would make any sense. If you're not better than the
10 unaided read, I'm not sure what clinical utility you
11 have, but you don't have to be better than your
12 predicate. You have to be at least as safe and
13 effective as the predicate. But if you did the head-
14 to-head, you're certainly going to have to have some
15 improvement over an unaided and then you're still
16 going to have to be at least as safe and effective as
17 the predicate.

18 DR. STEIER: Right. I was trying to
19 distinguish between how you have to compare it to the
20 unaided read versus the predicate.

21 DR. SMITH: Well, I guess I'm saying you
22 don't necessarily -- you're saying compare your
23 unaided read on the new device to the unaided read on
24 the predicate?

25 DR. STEIER: No, comparing your device to

Free State Reporting, Inc.
1378 Cape Saint Claire Road
Annapolis, MD 21409
(410) 974-0947

1 unaided read and comparing your device to the
2 predicate.

3 DR. SMITH: I guess, to me, I'm suggesting
4 a two-step process. You're going to have show some
5 improvement --

6 DR. D'ORSI: Yeah.

7 DR. SMITH: -- over the unaided read for
8 your own device and then you're going to have to
9 still make a comparison with the predicate.

10 DR. STEIER: Right.

11 DR. D'ORSI: Let's just lay out what the
12 control arms are and take a vote on them, okay? One
13 control arm would be -- or one method would be what
14 Dr. Gwise recommended, that there would be an unaided
15 read and a comparison between the aided for each
16 device; is that correct? So basically 3(iii), is
17 that correct?

18 DR. GWISE: Well, you would have the three
19 modalities tested on the same data. That's what I'm
20 saying.

21 DR. D'ORSI: Okay. And another use of a
22 control arm would be just the unaided read. It would
23 be basically an unaided with an aided read and a
24 comparison to a predicate. Is that another option?
25 Are there any more options? Yeah.

1 DR. JIANG: Could I just ask a quick
2 clarification? In the labeling, does it state how
3 much improvement there is, or does it just say it
4 improves whatever?

5 DR. SMITH: That's highly variable. It
6 depends on precisely what labeling the manufacturers
7 request. I mean, there are some mammography CAD
8 devices out there that have on their labeling that
9 they can cause cancer to be detected in a certain
10 percent of cases a certain, you know, number of
11 months earlier. There's all sorts of -- the
12 manufacturer can try to claim anything they want in
13 their labeling, but they have to provide the valid
14 scientific evidence to back it up.

15 Typically, if you don't compare yourself to
16 a predicate device, you're probably not going to have
17 in your labeling that you're better than, you know,
18 somebody else's device. You would think if you're
19 using your own prior device as a predicate, I'm not
20 sure why anybody would buy it if it doesn't have some
21 improvement, but they have to describe what that
22 claimed improvement is.

23 DR. JIANG: So am I understanding
24 correctly, that you have both kind of labels,
25 qualitative, stating quality of the improvement, and

1 the other kind is stating in some quantitative way
2 quantifying the improvement?

3 DR. SMITH: It's very common to quantify,
4 certainly, things like standalone performance, or if
5 there's been some specific change in sensitivity or
6 specificity, that would be specifically quantified.
7 It wouldn't necessarily be compared to another device
8 because manufacturers have not -- you know, strictly
9 for the CAD mammography devices, they've not been
10 head-to-head comparisons. So there's no real way to
11 label the -- put on the label that you're better than
12 somebody else's device.

13 DR. D'ORSI: Janine, can I -- can you just
14 reiterate the control arms and let's just vote on
15 them, all right? We strive for perfection but we
16 frequently don't get it. So let's see. What are the
17 possibilities we just discussed as control arms, if
18 you remember?

19 MS. MORRIS: Okay. To the best of my
20 recollection, when we're talking about a 510(k)
21 device, because of the problem with the case mix, you
22 know, there's a device, a predicate device out there
23 that was studied under a certain case mix and now,
24 you know, either a manufacturer with a modification
25 to their device or a new manufacturer is coming in,

1 and the question is, is whether or not it's good
2 enough for them to compare it to unaided read.

3 And as long as the basic performance
4 from -- on that individual CAD device is better, then
5 that should be enough to -- enough information to
6 make a substantial equivalence decision. But if you
7 believe that the differences in the case mix would
8 affect your ability to make that comparison, then you
9 would need to have the unaided read and device A and
10 device B comparison so that you could see that you've
11 not only increased performance but it's, on a head-
12 to-head comparison, as safe and as effective as the
13 chosen predicate device.

14 DR. D'ORSI: Okay, let's really -- those
15 are the choices, and that second choice is what
16 Dr. Gwise -- all right. So there's your choices,
17 okay? So let's start with Mr. Uzenoff. You have two
18 choices. Bob? Hello? Did you pass out or -- I
19 know, we're all in this boat.

20 MR. UZENOFF: No, I didn't. I wasn't
21 expected to be asked about --

22 DR. D'ORSI: Well, it's not a vote. It's
23 just an opinion.

24 MR. UZENOFF: An opinion.

25 DR. D'ORSI: There's no voting here. I

1 just want to get a sense of what people favor of
2 those two control arms. Can you quickly repeat it,
3 Janine?

4 DR. CARRINO: I think --

5 DR. D'ORSI: No?

6 DR. CARRINO: -- my understanding of the
7 fundamental question is, in a 510(k) process, do you
8 need a clinical multi-reader study, or can you use
9 standalone studies? Is that what you were advocating
10 before?

11 MS. MORRIS: No, actually this is on the
12 premise that we do need reader studies.

13 DR. D'ORSI: Yeah, clinical, yeah.

14 MS. MORRIS: And we're trying to choose the
15 control group --

16 MR. CARRINO: How to do it.

17 MS. MORRIS: -- for the reader study.

18 DR. CARRINO: Okay.

19 MS. MORRIS: We already have gotten past
20 standalone performance.

21 DR. CARRINO: Okay. So that's easier.

22 MR. UZENOFF: Could you repeat them one
23 more time for me?

24 MS. MORRIS: Okay, the question is, is that
25 if you have to provide a clinical performance reader

1 study to submit in a 510(k) to show substantial
2 equivalence, is it adequate to just have the control
3 arm to be unaided read or would you need to have a
4 head-to-head comparison with a predicate device
5 because the case mix throws in too much uncertainty
6 about how to understand that performance that is, you
7 know, is the result?

8 DR. D'ORSI: All right. So those are the
9 choices that we've gone around.

10 MR. UZENOFF: And you're talking about the
11 unaided read comparing to one case mix and comparing
12 to -- are we talking about absolute performance or
13 performance to some previous test for the predicate
14 device?

15 DR. D'ORSI: The unaided read and the aided
16 read compared to the predicate device's performance
17 characteristics, sensitivity/specificity.

18 MR. UZENOFF: Yeah. So where I'm confused
19 is I don't understand, if we're doing aided and
20 unaided read, why do we need a predicate?

21 DR. D'ORSI: Well, because it may be
22 easier.

23 MR. UZENOFF: So I don't -- yeah. But I
24 don't understand the efficiency or the tradeoffs
25 between those two.

1 DR. SMITH: This is Robert Smith again.
2 Maybe I can help clarify. I gave an example in my
3 slides, where you had one device. Let's say it had a
4 sensitivity of 80 percent. Say that was the metric.
5 And you had easy cases and inexperienced readers.
6 And then you had another device that had the same
7 exact sensitivity, where you had experienced readers
8 and difficult cases. So you had the same result.
9 And so I said that that's not a valid comparison.
10 Even though the outcome is exactly the same, whether
11 it was the area under the curve or the sensitivity,
12 because the cases were different, the scoring
13 methodology may have been different, you know, how
14 would we compare that? Because they're two different
15 devices, two different -- you know, again, everything
16 is so different, even though the outcome is exactly
17 the same. Could we conclude that they're
18 substantially equivalent because the sensitivity or
19 the area under the curve is the same under the
20 scenario when all those factors have confounded the
21 comparison?

22 DR. DODD: I'm sorry, I thought we had
23 required a representative sample with difficult cases
24 included. So now I'm confused. Because if we have
25 this representative sample, I don't see that

1 situation as being a concern.

2 DR. SMITH: Well, it depends on what sample
3 was used for the predicate device. We had not
4 necessarily any control over that. It may have been
5 cleared with a dataset that had, you know, again,
6 depending on what the device and the organ is, easy
7 cases, difficult cases. It may not have been
8 representative.

9 DR. DODD: Why are we comparing it to the
10 predicate? I mean, I'm just stuck. Because if I now
11 have a representative sample and I get a sensitivity
12 and specificity of something that seems acceptable,
13 that would be something a radiologist would like, why
14 should I be -- why I should have to compare it to a
15 predicate to market that device, if I have an
16 adequate sampling of the kinds of cases and things
17 I'm interested in?

18 DR. SMITH: Well, the standard under the
19 510(k) process in the regulation is you have to be at
20 least as safe and effective as the predicate. You
21 can certainly construct an argument along the lines
22 you're saying, but you have to be at least as safe
23 and effective as the predicate device.

24 DR. DODD: So can a company then come in
25 under a PMA? I mean, is it -- no.

Free State Reporting, Inc.
1378 Cape Saint Claire Road
Annapolis, MD 21409
(410) 974-0947

1 DR. D'ORSI: Okay.

2 DR. TOURASSI: But given the options, the
3 only option they have is the three-arm, most
4 burdensome approach, it almost make sense that they
5 will come back with a PMA.

6 DR. D'ORSI: All right, let's say this. If
7 we can get a standalone with an adequate
8 representation of cases and readers and we have some
9 method that the FDA will approve as a comparison, how
10 does that sound?

11 DR. DODD: We're talking about a reader
12 study, right, not standalone?

13 DR. D'ORSI: Correct, a reader study.

14 DR. DODD: Okay.

15 DR. D'ORSI: Is that valid? We'll leave
16 the comparison part up to the FDA, how to compare.
17 Don't look at me. You write the rules. Yes.

18 DR. BOURLAND: I'll try something here. I
19 don't know if it'll work or not, but -- and I'm not
20 sure which question I'm answering any longer, but --

21 (Laughter.)

22 DR. D'ORSI: That makes two of us.

23 DR. BOURLAND: Yeah, yeah, that's exactly
24 what I was going try to do. So I think, you know, if
25 a predicate is available, it's the thing to do. It's

1 what the regulations say and it's what ought to be
2 done, and you know, it has to be evaluated at that
3 time. If things have changed so much since that
4 time, the thing's not even available, obviously
5 that's not going to work. You know, garbage in,
6 garbage out type of thing. So then the unaided read
7 is the backup. And I think that the applicant can
8 justify that. I mean, it's got to be addressed some
9 way. You can't just stop and say, well, I'm sorry,
10 you don't have the predicate.

11 DR. D'ORSI: Yeah. Go ahead, John.

12 DR. CARRINO: I would also support that,
13 that if the predicate device is not available to use,
14 then I would question the suitability of that as a
15 predicate device. So you have to test equivalency to
16 something that's available, and then we can go
17 through the paradigm that Dr. Bourland mentioned.

18 DR. D'ORSI: Three months ago when we
19 started this we said that we understand that the
20 predicate, if it's from the same manufacturer, it's
21 the easiest thing. That's a way. But we were coming
22 with another proviso, if case mix was a problem, if
23 reader was a problem, and this is the example we came
24 up with, a standalone with an adequate case set that
25 could, under FDA manipulation -- that's the wrong

1 word -- under FDA statute, could be compared to a
2 predicate. That's what we're getting at. So how
3 does everybody feel about that? You're not going to
4 say anything now. Okay.

5 DR. GLASSMAN: I love it.

6 DR. D'ORSI: Janine, is that --

7 MS. MORRIS: Yeah. Just to put clarity on
8 it, I am hearing from the Panel that we have options
9 here, that it's not a mandatory requirement that it
10 has to be compared to the predicate; that with
11 certain restrictions on the case mix that we have
12 representation in the dataset, that an unaided read
13 and showing adequate performance is an option in
14 which we can make a substantial equivalence decision.

15 DR. D'ORSI: Yeah, that sounds great.
16 Okay, let's go to the next one. What is the next
17 one?

18 MS. MORRIS: Reader characteristics. This
19 is basically asking, you know, when we're doing an
20 MRMC study, what are the number of readers and what
21 should they look like, what should the composition
22 be? And that is --

23 DR. D'ORSI: I would say human, first.

24 MS. MORRIS: -- 3(b).

25 DR. OCHS: So the guidance calls for the

Free State Reporting, Inc.
1378 Cape Saint Claire Road
Annapolis, MD 21409
(410) 974-0947

1 trial readers to be representative of the intended
2 population of clinical users. Can you provide
3 examples of sets of readers that are representative
4 of clinical users? Should there be a minimum number
5 of readers? And if there are important subgroups of
6 readers, should the number of readers in each of the
7 subgroups be proportional to the number of readers in
8 the population of clinical users?

9 DR. D'ORSI: Let me throw something out,
10 and then you can think about it. This gets to be
11 very burdensome, and I don't know if it's needed for
12 what you want to do with these devices. This is a
13 great question to ask as a post-market -- not a post-
14 market thing but as a grant application or something.
15 But I don't think this level of splitting is
16 necessary. So with that, why don't we open the
17 discussion. Anybody.

18 DR. CARRINO: Yeah. For some of the ROC --
19 I mean, for these multi-reader studies, the numbers,
20 two is too few, and I think people are trying to go
21 toward four to six readers, and I think you're
22 targeting general radiologists, those are
23 appropriate, and you don't need the subgroup under
24 specialists, if you want a simple --

25 DR. D'ORSI: At this point, yes. Len.

1 DR. GLASSMAN: Yeah, I think that to try to
2 recreate the proper percentage of academic
3 radiologists and private practice radiologists and
4 hospital radiologists from small communities will
5 drive everybody crazy and be really burdensome, and I
6 think board-certified or board-eligible radiologists
7 with a heartbeat.

8 (Laughter.)

9 DR. GLASSMAN: I don't know how many, but
10 that would be my criteria.

11 UNIDENTIFIED SPEAKER: Or a pacemaker.

12 DR. GLASSMAN: Well, they'd still have a
13 heartbeat.

14 DR. ROSENBERG: We would assume MQSA
15 certified.

16 DR. D'ORSI: Yes, yes, yes. Did you have
17 something?

18 DR. PAYNE: No, I was just qualifying. I
19 mean, you've got mammography versus colon and et
20 cetera. So appropriate to the modality, yeah.

21 DR. D'ORSI: Right. Is that enough
22 information?

23 MS. MORRIS: I believe so, as long as
24 there's no --

25 DR. D'ORSI: Does everybody -- let's go

1 down. How do you feel about what we just said? Bob.

2 DR. CARRINO: Great.

3 DR. D'ORSI: Dr. Jiang.

4 DR. JIANG: Me too.

5 DR. TOURASSI: I agree.

6 DR. D'ORSI: Dr. Tourassi.

7 DR. GLASSMAN: I agree.

8 DR. D'ORSI: Dr. Glassman. Dr. Dodd.

9 DR. DODD: I have to agree.

10 DR. PAYNE: Yeah.

11 DR. D'ORSI: Dr. Payne.

12 DR. ZISKIN: Yes.

13 DR. D'ORSI: Bob.

14 DR. ROSENBERG: Yes.

15 DR. D'ORSI: Craig.

16 DR. ABBEY: Yes.

17 DR. STEIER: Yes.

18 DR. D'ORSI: Daniel.

19 DR. BOURLAND: Yeah.

20 MR. UZENOFF: Yes.

21 DR. D'ORSI: Okay. All right, what's the
22 next point?

23 DR. ZISKIN: We're making progress.

24 MS. MORRIS: We are finished with pre-
25 market requirements.

1 DR. D'ORSI: All right.

2 MS. MORRIS: So if you can endure, we'll go
3 to post-market considerations.

4 DR. D'ORSI: Yeah, let's do it. Okay, what
5 are we covering under there?

6 MS. MORRIS: All right, this is covering
7 Questions 6, 7 and 9. We'll start with 6. Robert,
8 you can read it.

9 DR. OCHS: So historically, PMA
10 applications for mammography CAD devices include
11 retrospective studies with enriched data but did not
12 include data from prospective clinical trials due to
13 the significant burden of adequately powering a
14 prospective study. Published literature of clinical
15 studies evaluating CAD mammography in a post-market
16 setting have not presented a consensus of findings or
17 have limitations that minimize generalizability.
18 Please comment on the following:

19 a. Although a retrospective study with
20 enriched data may be adequate to demonstrate a
21 reasonable assurance of safety and effectiveness,
22 should the question of device performance under
23 actual conditions of use, that is, post-market, be
24 answered by a post-approval study?

25 DR. D'ORSI: Yeah, Len, Len.

Free State Reporting, Inc.
1378 Cape Saint Claire Road
Annapolis, MD 21409
(410) 974-0947

1 DR. GLASSMAN: Yeah, I think the answer to
2 that is no, unless there is suspicion on the part of
3 the FDA that there is a problem either identified by
4 complaints from users or published peer-reviewed
5 studies.

6 DR. D'ORSI: I would agree with that.
7 Let's go down the --

8 DR. CARRINO: I agree.

9 DR. D'ORSI: Georgia, sorry.

10 DR. TOURASSI: Yes, agree.

11 DR. D'ORSI: Dr. Dodd.

12 DR. DODD: I wasn't listening.

13 DR. D'ORSI: That's all right.

14 DR. DODD: I was getting some sugar to
15 carry on --

16 DR. GLASSMAN: Unless there's a problem
17 identified.

18 DR. DODD: Wait, could you restate that?

19 DR. GLASSMAN: Okay. Post-market analysis
20 studies would only be necessary, would not be
21 routinely necessary, would only be necessary if the
22 FDA received complaints from users or there were
23 peer-reviewed articles which showed a problem.

24 DR. DODD: I need to think about that.

25 DR. D'ORSI: All right.

1 DR. ZISKIN: Yes.

2 DR. D'ORSI: Bob.

3 DR. ROSENBERG: I think, yes, it would be a
4 difficult process to require, otherwise.

5 DR. D'ORSI: Craig.

6 DR. ABBEY: Yes.

7 DR. CARRINO: Yeah, I agree.

8 DR. D'ORSI: Bob.

9 MR. UZENOFF: I agree.

10 DR. D'ORSI: Lori. I'm sorry.

11 DR. DODD: There's no abstaining?

12 UNIDENTIFIED SPEAKER: Yes.

13 DR. DODD: Yes, I can abstain?

14 UNIDENTIFIED SPEAKER: Yes, there's no
15 abstaining.

16 DR. DODD: You got me. Yes.

17 DR. D'ORSI: Okay, let's go on to the next
18 one. It's getting tight. The next point.

19 MS. MORRIS: I think, in the interest of
20 what we can possibly accomplish, I think the Agency
21 is willing to end here, and if it's necessary to go
22 on with those other questions, we'll come up with a
23 homework assignment.

24 DR. D'ORSI: Is that valid? Are you okay
25 with that?

1 MS. MORRIS: I think so. I checked with
2 our post-market staff and they were --

3 DR. D'ORSI: It's okay?

4 MS. MORRIS: Yeah.

5 DR. D'ORSI: All right, thank you. The
6 meeting's adjourned. You did phenomenally.

7 (Whereupon, at 6:30 p.m., the meeting was
8 concluded.)

9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25

C E R T I F I C A T E

This is to certify that the attached proceedings
in the matter of:

RADIOLOGICAL DEVICES PANEL

November 18, 2009

Gaithersburg, Maryland

were held as herein appears, and that this is the
original transcription thereof for the files of the
Food and Drug Administration, Center for Devices and
Radiological Health, Medical Devices Advisory
Committee.

TODD BROWN,

Official Reporter

Free State Reporting, Inc.
1378 Cape Saint Claire Road
Annapolis, MD 21409
(410) 974-0947