UNITED STATES OF AMERICA

DEPARTMENT OF HEALTH AND HUMAN SERVICES

FOOD AND DRUG ADMINISTRATION

CENTER FOR DEVICES AND RADIOLOGICAL HEALTH

RADIOLOGICAL DEVICES PANEL

TUESDAY, MARCH 4, 2008

The meeting came to order at 8:00 a.m. in the Grand Ballroom of the Hilton Washington DC North, 620 Perry Parkway, Gaithersburg, MD. Leonard M. Glassman, MD, Acting Chairman, presiding.

PRESENT:

LEONARD M. GLASSMAN, MD   ACTING CHAIRPERSON
NANCY WERSTO   EXEC. SECRETARY
JOHN D. BOURLAND, PHD   MEMBER
CARL J. D'ORSI, MD   MEMBER
BHARAT B. MITTAL, MD   MEMBER
MARVIN C. ZISKIN, MD   MEMBER
CRAIG K. ABBEY, PHD   TEMP. VOTING MBR.
DONALD A. BERRY, PHD   TEMP. VOTING MBR.
JOHN A. CARRINO, MD/MPH   TEMP. VOTING MBR.
LORI E. DODD, PHD   TEMP. VOTING MBR.
BRIAN S. GARRA, MD   TEMP. VOTING MBR.
DAVID KIM, MD   TEMP. VOTING MBR.
MARILYN LEITCH, MD   TEMP. VOTING MBR.
OTTO LIN, MD   TEMP. VOTING MBR.
ROBERT D. ROSENBERG, MD   TEMP. VOTING MBR.
BERKMAN SAHINER, PHD   TEMP. VOTING MBR.
KENNETH J. STEIER,DO/MPH/MHA TEMP VOTING MBR
DANIEL R. SWERDLOW, MD   TEMP. VOTING MBR.
GEORGIA D. TOURASSI, PHD   TEMP. VOTING MBR.
A. CHRISTINE WATT, MB/CHB TEMP VOTING MBR.
ROY K.H. WONG, MD   TEMP. VOTING MBR.
NANCY FINKEN, MFA   CONSUMER REP.
DAVID SPINDELL, MD   INDUSTRY REP.
NANCY BROGDON   FDA

NEAL R. GROSS

COURT REPORTERS AND TRANSCRIBERS
1323 RHODE ISLAND AVE., N.W.
(202) 234-4433   WASHINGTON, D.C. 20005-3701   www.nealrgross.com

TABLE OF CONTENTS

TABLE OF CONTENTS

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

2   8:05 a.m.

3   EXEC. SEC. WERSTO: Good morning. We

4 will now begin the open session. Thank you.

5   CHAIRMAN GLASSMAN: Okay. There we

6 go. Okay. I would like to call this meeting

7 of the Radiological Devices Panel to order. I

8 am   Dr.   Leonard   Glassman,   the   Acting

9 Chairperson of this Panel. I am a Diagnostic

10 Radiologist   in   private   practice   in   the

11 Washington, D.C. area. I am also the American

12 College of Radiology Breast Imaging Scientist

13 at the Armed Forces Institute of Pathology.

14 And I'm a Clinical Professor of Radiology at

15 both GW and Georgetown.

16   If you haven't already done so,

17 please, sign in at the attendance sheets that

18 are on the table by the doors. The agenda for

19 this meeting is also available outside the

20 door if you need to get one.

21   If you are presenting at the open

22 public hearing sessions today and have not

1 previously presented an electronic copy of

2 your presentation to the FDA, please arrange

3 to do so with Sunder Rajan. Sunder, can you

4 stand up? So give him your presentations,

5 please. And please, silence all cell phones.

6 I would like to announce the

7 remaining tentatively scheduled meetings of

8 this panel for 2008. They are August 12th and

9 November 4th. Please, remember that these are

10 tentative dates. You may monitor the panel

11 website for updated information.

12 I note for the record that the

13 voting members present constitute a quorum

14 required by 21 CFR Part 14, and I hope none of

15 you ask me what that means.

16 Ms. Wersto, the Executive Secretary

17 for the Radiological Devices Panel, will make

18 some introductory remarks in just one second.

19 For the panel, I would like to let

20 you know that your microphone has two buttons

21 on it. The one on the right, the larger

22 button, you push, the red light goes on and

1 then you can speak. We can only have four of

2 these microphones off -- I mean on, at one

3 time. So please, remember to turn off the red

4 light, push the button again, so that someone

5 else has a chance to speak.

6   Also, the button on the left does

7 not work for you, but it does for me, and it

8 turns off your light if I need to.

9   Okay. Nancy?

10   EXEC. SEC. WERSTO: Good morning,

11 everyone. Before I turn the meeting over to

12 Dr. Glassman, I am required to read the

13 Conflict of Interest statement into the

14 record.

15   FDA Conflict of Interest Disclosure

16 Statement, particular matters of general

17 applicability. Radiological Devices Panel of

18 the Medical Devices Advisory Committee, March

19 4 and 5, 2008.

20   The Food and Drug Administration,

21 FDA, is convening today's meeting of the

22 Radiological Devices Panel of the Medical

1   Devices Advisory Committee under the authority

2   of the Federal Advisory Committee Act of 1972.

3   With the exception of the industry

4   representative, all members and consultants of

5   the panel are special Government employees,

6   SGEs, or regular federal employees from other

7   agencies and are subject to Federal Conflict

8   of Interest Laws and Regulations.

9   The following information on the

10   status of this panel's compliance with Federal

11   Ethics and Conflict of Interest Laws covered

12   by, but not limited to, those found at 18 USC

13   Section 208 and Section 712 of the Federal

14   Food, Drug and Cosmetic Act are being provided

15   to participants in today's meeting and to the

16   public.

17   FDA has determined that members and

18   consultants of this panel are in compliance

19   with Federal Ethics and Conflict of Interest

20   Laws. Under 18 USC Section 208, Congress has

21   authorized FDA to grant waivers to special

22   Government employees who have financial

1   conflicts when it is determined that the

2   Agency's need for particular individuals'

3   services outweighs his or her potential

4   financial conflict of interest.

5   Under Section 712 of the Food, Drug

6   and Cosmetic Act, Congress has authorized FDA

7   to grant waivers to special government

8   employees and regular government employees

9   with potential financial conflicts when

10 necessary to afford the Committee essential

11 expertise.

12   Related to the discussion of

13 today's meeting, members and consultants of

14 this panel who are special government

15 employees have been screened for potential

16 financial conflicts of interest of their own

17 as well as those imputed to them, including

18 those of their spouses or minor children, and

19 for purposes of 18 USC Section 208, their

20 employers.

21   These interests may include

22 investments, consulting, expert witness

1 testimony, contracts, grants, CRADAs,

2 teaching, speaking, writing, patents and

3 royalties, and primary employment.

4 The Agency involves a general

5 discussion of -- the agenda involves a general

6 discussion of Computer-Aided Detection and

7 diagnosis, CAD devices, for radiological

8 images, such as mammograms, chest x-rays and

9 computed tomography of the lungs or colon.

10 The general discussion will focus

11 on the general methodologies for CAD,

12 including how CAD devices are used in clinical

13 decision making, how the devices are tested

14 and the information needed to properly assess

15 their safety and effectiveness.

16 The general discussion will be

17 followed by specific discussions related to

18 mammography CAD devices, colon CAD devices and

19 lung CAD devices. These discussions will

20 include how the different types of CAD devices

21 are used and the literature published

22 regarding these devices with focus on testing

1 issues related to the different devices.

2 This is a particular matters
3 meeting, during which general issues will be
4 discussed.

5 Based on the meeting and all
6 financial interests reported by the panel
7 members and consultants, conflict of interest
8 waivers have been issued in accordance with 18
9 USC Section 208(b)(3) and Section 712 of the
10 Food, Drug and Cosmetic Act to Dr. John
11 Carrino.

12 Dr. Carrino's waivers address
13 personal consulting arrangements with a firm
14 at issue. He receives an annual fee of less
15 than $10,001 for these arrangements, which are
16 unrelated to today's agenda. The waivers
17 follow Dr. Carrino to -- the waivers allow Dr.
18 Carrino to participate fully in today's
19 deliberations.

20 FDA's reasons for issuing the
21 waivers are described in the waiver documents,
22 which are posted on FDA's website at:

1 www.fda.gov/ohrms/dockets/default.htm. Copies

2 of the waivers may also be obtained by

3 submitting a written request to the Agency's

4 Freedom of Information Office, Room 6-30 of

5 the Parklawn Building.

6 David Spindell, MD is serving as

7 the industry representative, acting on behalf

8 of all related industry and is employed by

9 Abbott Laboratories Medical Products Group.

10 We would like to remind members and

11 consultants that if discussions involve any

12 other products or firms not already on the

13 agenda for which an FDA participant has a

14 personal or imputed financial interest, the

15 participants need to exclude themselves from

16 such involvement and their exclusion will be

17 noted for the record.

18 FDA encourages all other

19 participants to advise the panel if any

20 financial relationships that they may have

21 with any firms at issue. Thank you.

22 Now, for a few general

NEAL R. GROSS

COURT REPORTERS AND TRANSCRIBERS
1323 RHODE ISLAND AVE., N.W.
(202) 234-4433   WASHINGTON, D.C. 20005-3701   www.nealrgross.com

1 announcements. Transcripts of today's meeting

2 will be available from Neal Gross and Company

3 by calling (202) 234-4433. Information on

4 purchasing videos of today's meeting can be

5 found at the table outside of the meeting

6 room. Presenters to the panel who have not

7 already done so, should provide FDA with a

8 hard copy and an electronic copy of their

9 remarks.

10 I would like to remind everyone

11 that members of the public and the press are

12 not permitted around the panel area, that is

13 beyond the speaker's podium. The press

14 contact for today's meeting is Peper Long. I

15 don't know if she is here. Okay, there she

16 is.

17 I request that reporters wait to

18 speak to FDA officials until after the panel

19 meeting. Thank you.

20 CHAIRMAN GLASSMAN: Thank you, Ms.

21 Wersto. Good morning, everyone. At this

22 meeting, the panel will be making

1 recommendations to the Food and Drug
2 Administration on general issues pertaining to
3 Computer-Aided Detection devices and on
4 specific issues pertaining to mammography
5 CADs, colon CADs and lung CADs.

6   Before we begin this meeting, I
7 would like to ask our distinguished panel
8 members, who have generously given their time
9 to help the FDA in the matters being discussed
10 today, and other FDA staff seated at this
11 table, to introduce yourselves. Please, state
12 your name, your area of expertise, your
13 position, your institution and your status on
14 the panel as a voting member, deputized voting
15 member, consumer representative or industry
16 representative.

17   And I have already introduced
18 myself, Leonard Glassman, as a diagnostic
19 radiologist. I'm an expert in medial
20 ultrasound and in breast imaging and I have
21 given you my credentials briefly. So why
22 don't we then go around the table. Why don't

1 we start with Ms. Brogdon down there?

2 MS. BROGDON: Good morning, I'm

3 Nancy Brogdon. I'm not a member of the panel.

4 I'm the Director of FDA's Division of

5 Reproductive, Abdominal, and Radiological

6 Devices.

7 MS. FINKEN: I am Nancy Finken. I

8 am a non-voting Consumer Advocate, a member of

9 the Board of Directors of the Virginia Breast

10 Cancer Foundation and Why Me, a national

11 breast cancer support group. I represent the

12 consumer side of this panel.

13 DR. TOURASSI: I'm Georgia

14 Tourassi, Associate Professor of Radiology at

15 Duke University Medical Center. My primary

16 area of expertise is CAD development and image

17 retrieval in mammography. And I'm a temporary

18 voting member.

19 DR. BERRY: I'm Donald Berry. I'm

20 a Biostatistician, Chair of Biostatistics and

21 Head of the Division of Quantitative Sciences

22 at M.D. Anderson Cancer Center. I'm a voting

1 member. I'm not sure whether I'm a deputy or

2 not.

3 DR. SWERDLOW: I am Dan Swerdlow.

4 I'm an Assistant Professor of Radiology at

5 Georgetown University in the Section of

6 Abdominal Imaging. My interests are robotics

7 and intervention applications and, recently,

8 virtual colonoscopy.

9 DR. WATT: I am Christine Watt.

10 I'm a Diagnostic Radiologist, Chief of Breast

11 Imaging in the Detroit area of two breast

12 centers. I also work for three separate

13 health systems in the Detroit area. My

14 expertise is mammography.

15 DR. GARRA: I'm Brian Garra. I am

16 the Director of Ultrasound and a Professor of

17 Radiology at the University of Vermont and

18 Fletcher Allen Healthcare. My interests are

19 diagnostic ultrasound and digital signal

20 processing. And I'm a temporary member.

21 DR. ABBEY: Hello, my name is Craig

22 Abbey. I'm also a temporary member. I'm a

1 researcher in the Department of Psychology at

2 University of California, Santa Barbara, and

3 also in Biomedical Engineering at UC Davis.

4   CHAIRMAN   GLASSMAN: Temporary

5 member?

6   DR.   ABBEY: Temporary   member.

7 Thank you.

8   DR. WONG: My name is Dr. Roy Wong.

9 I'm the Chief of Gastroenterology at Walter

10 Reed Army Medical Center and a Professor of

11 Medicine at the Uniformed Services University

12 of the Health Sciences. My major interest is

13 actually in the esophagus, but also with

14 virtual colonoscopy and CTC. I'm a non-voting

15 member, I guess.

16   DR. ZISKIN: My name is Marvin

17 Ziskin.   I'm a Professor of Radiology and

18 Medical Physics at Temple University Medical

19 School and I'm the director of the Center for

20 Biomedical   Physics.   My   expertise   is   in

21 physics   and   safety   of   ultrasound   and

22 electromagnetic fields.

NEAL R. GROSS

COURT REPORTERS AND TRANSCRIBERS
1323 RHODE ISLAND AVE., N.W.
(202) 234-4433   WASHINGTON, D.C. 20005-3701   www.nealrgross.com

1   DR. MITTAL: Good morning. My name

2   is Bharat Mittal. I'm Professor and Chairman

3   of   Radiation   Oncology   at   Northwestern

4   University in Chicago. My area of expertise

5   is -- relates to radiation oncology. And I

6   believe I'm the voting member.

7   DR. STEIER: I believe -- I'm Ken

8   Steier, and I'm the Dean of Academic Affairs

9   and Clinical Professor of Medicine in the

10   Pulmonary and Critical Care Division and

11   Patient Safety Officer at the NASA University

12   Medical Center in Long Island, New York. And

13   I'm a temporary voting member.

14   DR. BOURLAND: I'm Dan Bourland,

15   Associate Professor in Radiation Physics in

16   the Department of Radiation Oncology at Wake

17   Forest University. My interests are imaging,

18   the uses of digital imaging, and radiation

19   oncology; and I am a voting member.

20   DR. LIN: My name is Otto Lin. I'm

21   a Gastroenterologist at Virginia Mason Medical

22   Center and also a Clinical Associate Professor

NEAL R. GROSS

COURT REPORTERS AND TRANSCRIBERS
1323 RHODE ISLAND AVE., N.W.
(202) 234-4433   WASHINGTON, D.C. 20005-3701   www.nealrgross.com

1 of Medicine at the University at Washington

2 School of Medicine. I'm a temporary voting

3 member. My area of research interest is colon

4 cancer screening, including screening

5 colonoscopy and virtual colonoscopy.

6   DR. D'ORSI: I'm Carl D'Orsi. I'm

7 a Professor of Radiology and Hematology and

8 Oncology at Emory University and Director of

9 the Breast Imaging there. My interests are in

10 technology assessment for breast imaging and I

11 think I'm a voting member.

12   DR. DODD: I'm Lori Dodd. I'm a

13 Mathematical Statistician at the Biometrics

14 Research Branch at the National Cancer

15 Institute. And I have an interest in clinical

16 trials and imaging.

17   DR. ROSENBERG: I'm Robert

18 Rosenberg, Professor at University of New

19 Mexico, Chief of Diagnostic Imaging. Areas of

20 expertise are mammography and community

21 mammography outcomes.

22   DR. CARRINO: Hi, I'm John Carrino.

1 I'm an Associate Professor of Radiology in

2 Orthopedic Surgery at Johns Hopkins

3 University. My area of expertise is PACs,

4 Picture Archive and Communication systems, and

5 I'm a temporary voting member.

6 DR. SAHINER: I am Berkman Sahiner.

7 I'm also a temporary voting member. I'm an

8 Associate Professor of Radiology at the

9 University of Michigan. And my areas of

10 interest are medical imaging in general and

11 CAD, in particular.

12 DR. LEITCH: I'm Marilyn Leitch.

13 I'm a Surgical Oncologist and Professor of

14 Surgery at UT Southwestern Medical Center in

15 Dallas. I have a special interest in breast

16 cancer and am the Medical Director for the

17 Center of Breast Care there. My interests in

18 this are primarily related to the clinical

19 applications and how they impact patient care.

20 DR. KIM: I'm David Kim. I'm an

21 Assistant Professor of Radiology at the

22 University of Wisconsin. My area of research

1 is CT colonography, and I am a temporary

2 voting member.

3 　　　　　DR. SPINDELL: Good morning, I'm

4 David Spindell. I'm the Division Vice

5 President of Medical Affairs for Abbott

6 Laboratories, and I'm the Industry

7 Representative.

8 　　　　　CHAIRMAN GLASSMAN: Okay. Thank

9 you all very much. Ms. Brogdon would like to

10 make some remarks to thank those members whose

11 terms have ended on January 31, 2008.

12 　　　　　MS. BROGDON: Thank you, Dr.

13 Glassman. We have four panel members who have

14 recently rotated off this panel. As Dr.

15 Glassman said, their four year terms ended on

16 January 31st, so they are not able to be

17 present at this meeting.

18 　　　　　But I want to acknowledge their

19 contributions and thank them in absentia for

20 their service on the panel. They will each

21 receive a plaque and a letter of commendation

22 and thanks from the Commissioner of Food and

1   Drugs.

2   First is Dr. Elizabeth Krupinski.

3   Dr. Krupinski was a voting member for three

4   years and then the Panel Chair for one year.

5   Dr. Andrew Zhou was a voting member,

6   statistician on the panel. Mr. Barry Burns

7   was the Consumer Representative, and Ms.

8   Deborah Moore was the Industry Rep. And I

9   would like to thank them all for their public

10   service. Thank you.

11   CHAIRMAN GLASSMAN: Thank you, Ms.

12   Brogdon. No one has any questions, at this

13   point. We will now proceed with the FDA

14   presentations providing a general overview of

15   CAD devices and highlighting current issues

16   related to mammography CADs.

17   The first presenter is Joyce Whang,

18   Deputy Division Director of RARD. Thank you.

19   Joyce?

20   DR. WHANG: Good morning. Thank

21   you for joining us for this meeting of the FDA

22   Radiological Devices Panel. We have a lot of

NEAL R. GROSS

COURT REPORTERS AND TRANSCRIBERS
1323 RHODE ISLAND AVE., N.W.
(202) 234-4433   WASHINGTON, D.C. 20005-3701   www.nealrgross.com

1 ground to cover over the next two days, and

2 we're delighted to have such an esteemed panel

3 of experts to help us.

4   As we have all just heard, we have

5 many different disciplines represented here

6 today, and we are looking forward to a

7 vigorous and productive discussion.

8   The topic of this meeting will be

9 Computer-Aided Detection and Diagnostic

10 Devices, that is CAD devices as used for

11 radiological images. The emphasis will very

12 much be on detection devices, although as Dr.

13 Petrick will discuss today, there is some

14 overlap between detection and diagnosis.

15   We will be looking for input from

16 the panel as to the types of data that should

17 be provided to support the approval and

18 clearance of CAD devices. This meeting is

19 part of the process for developing guidance

20 document regarding CAD devices. We will be

21 looking closely at the recommendations of the

22 panel members as part of our efforts to

NEAL R. GROSS

COURT REPORTERS AND TRANSCRIBERS
1323 RHODE ISLAND AVE., N.W.
(202) 234-4433   WASHINGTON, D.C. 20005-3701   www.nealrgross.com

1 develop a draft version of this guidance

2 document.

3   The draft will be put out for

4 public comment, and then revisions will be

5 made as appropriate. This entire process of

6 developing a guidance document typically takes

7 two to three years.

8   I would like to remind the Panel

9 that FDA is required to be least burdensome in

10 our expectations of companies. We are not

11 looking for the perfect study design, but what

12 is needed for demonstrating safety and

13 effectiveness.

14   We're going to start out today with

15 general background about CAD. Dr. Robert

16 Smith will provide an introduction to the

17 clinical use of CAD devices and then some

18 comments about the regulatory background.

19 Then Nick Petrick will discuss the general

20 methodologies that are used for assessing CAD

21 devices, and Tom Gwise will speak to specific

22 statistical issues in relation to these

NEAL R. GROSS

COURT REPORTERS AND TRANSCRIBERS
1323 RHODE ISLAND AVE., N.W.
(202) 234-4433   WASHINGTON, D.C. 20005-3701   www.nealrgross.com

1   methods.

2   This general background is largely

3   common to various types of CAD devices, and it

4   provides the framework for our four major

5   areas of discussion. For each of these areas,

6   we will provide some background specific to

7   the clinical applications. Then we will have

8   an open public hearing in which we will hear

9   comments from the public. Then we will ask

10   the Panel to address a series of discussion

11   questions.

12   We will cover two of these sessions

13   today. Robert Smith will provide the clinical

14   background for the mammography CAD devices,

15   and then Frank Samuelson will discuss colon

16   CAD devices.

17   Tomorrow, we will do the other two

18   topics. Sophie Paquerault will provide the

19   background for lung CAD devices, and Stacey

20   Bilek will discuss future issues related to

21   CAD devices. For this last session on future

22   issues, we ask the Panel members to step back

1   from the specifics of mammography CAD, colon

2   CAD, and lung CAD that they will be discussing

3   in the other sessions, and to think about the

4   methodologies we have been discussing in these

5   sessions and how they apply to other areas.

6   This is the time to focus on

7   computer-aided diagnostics, on CAD for other

8   clinical applications for other modalities,

9   such as breast MRI, for CADs of the future,

10   and how these discussions fit in with other

11   computer-based technologies that might not be

12   CAD.

13   There are some subjects that will

14   be raised in this morning's background

15   presentations on general CAD methods and

16   additional statistical issues that we will be

17   asking you to address during this final

18   discussion period tomorrow afternoon. And Dr.

19   Bilek will remind you of what these topics are

20   before we get to those questions.

21   With that, I'll turn things over to

22   Dr. Smith to provide the clinical and

1   regulatory background for CAD devices.

2   DR. SMITH: Thank you, Joyce. My

3   name is Robert Smith.   I'm a Medical Officer

4   and Diagnostic Radiologist at CDRH.   And I'm

5   just going to give a very brief overview of

6   the clinical use of CAD devices and a brief

7   overview of general regulatory background of

8   medical devices.

9   Where does CAD fit into clinical

10   practice?   CAD is intended to aid or assist

11   radiologists   with   the   interpretation   of

12   imaging tests.   It is, therefore, important to

13   understand how radiologists interpret images

14   in order to identify where or how CAD might be

15   helpful.

16   There are four basic steps of image

17   interpretation: detection,   description,

18   diagnosis, that is analysis, and reporting.

19   Detection is the identification of a finding

20   that might be abnormal and that might require

21   further   scrutiny.   If   not   an   abnormal

22   structure, the next step is to describe the

1   features of the finding.

2   Description of the features does

3   require detection of individual features or

4   sub-features. A description directly drives

5   the diagnosis, and that is why particular

6   descriptive features have been developed over

7   time.

8   Diagnosis is the analysis of

9   imaging findings. And the diagnosis may be

10   definitive. For example, a definitely benign

11   or a definitely malignant finding where a

12   particular entity may be diagnosed. But a

13   diagnosis may also include a recommendation

14   for a clinical action. The clinical actions

15   may include additional concurrent imaging,

16   additional follow-up imaging in the future,

17   biopsy or surgery.

18   Reporting is the communication to

19   one or more referring clinicians and sometimes

20   the patient themselves.

21   CAD devices can be designed to

22   assist with detection, description, diagnosis

1 and/or reporting. Whatever the design,

2 however, the primary purpose of a CAD device

3 is to reduce errors associated with the

4 interpretation of imaging tests. It is,

5 therefore, important to know what are the

6 different types of errors that radiologists

7 can make when interpreting imaging tests.

8 There are two basic types of

9 errors: perceptual errors, that is errors of

10 detection, cognitive errors, that is errors of

11 analysis. There are also technical and

12 administrative errors, which may be associated

13 with or cause perceptual or cognitive errors.

14 A perceptual error is when a

15 radiologist fails to perceive an abnormality

16 at the time of interpretation and that

17 abnormality is in some way judged to be

18 evident that is actually revealed by the test,

19 and this is usually done in retrospect at a

20 later time. A perceptual error also includes

21 the failure to perceive important features of

22 an abnormality that is otherwise detected.

NEAL R. GROSS

COURT REPORTERS AND TRANSCRIBERS
1323 RHODE ISLAND AVE., N.W.
(202) 234-4433   WASHINGTON, D.C. 20005-3701   www.nealrgross.com

1   A cognitive error, or error of

2 analysis, is when a radiologist does perceive

3 an abnormality, but misinterprets the nature

4 or significance of the abnormality due to

5 incomplete knowledge or faulty analysis or

6 judgment.

7   Technical errors occur when the

8 quality of the examination is diminished due

9 to some technical factor, such as under- or

10 over-exposure of a film, choice of the wrong

11 reconstruction algorithm on CT, or poor choice

12 of imaging parameters on an MRI exam that

13 results in diminished signal/noise ratio.

14   Examples of administrative errors

15 would be a failure to obtain adequate patient

16 history or a failure to compare with prior

17 tests, imaging the wrong body part or even the

18 wrong patient, losing images whether it be

19 film or digital, mislabeling right versus

20 left, or the failure even to correct an error

21 in a dictated report.

22   Technical and administrative errors

1 are important because they can either cause or

2 contribute to perceptual and cognitive errors.

3 I'll now very briefly just go

4 through some general issues that relate to the

5 regulation of medical devices. The

6 classification of medical devices is based on

7 risk. Class I devices are low-risk devices,

8 such as stethoscopes. Class II devices have a

9 moderate-risk. And this would include most

10 imaging devices, such as CT, MRI, and

11 ultrasound scanners.

12 And Class III devices are high-risk

13 devices. For example, pace makers, but other

14 devices may also be Class III because they are

15 used in conjunction with a Class III device,

16 because they have a large potential effect on

17 the public health, or because the scientific

18 principles of the device are not well-known.

19 There are two basic ways that

20 medical devices reach the market in the United

21 States through either the 510(k) process or

22 the PMA process. Most Class II devices are

NEAL R. GROSS

COURT REPORTERS AND TRANSCRIBERS
1323 RHODE ISLAND AVE., N.W.
(202) 234-4433   WASHINGTON, D.C. 20005-3701   www.nealrgross.com

cleared for marketing via what's called a 510(k), and this is just named after the numbered section of the statutory law. And in a 510(k) application, the manufacturer is required to demonstrate substantial equivalence to another legally U.S. marketed device. And that other device is referred to as a predicate device.

A device is substantially equivalent if compared to a predicate device, it has the same intended use as the predicate and the same technological characteristics as the predicate, or the device has the same intended use as the predicate and it has different technological characteristics.

For example, a different material, different design, a different energy source, a different software algorithm. And when it has different technological characteristics, it is also required that the information submitted to FDA does not raise different, that is new, types of questions of safety and

NEAL R. GROSS

COURT REPORTERS AND TRANSCRIBERS
1323 RHODE ISLAND AVE., N.W.
(202) 234-4433   WASHINGTON, D.C. 20005-3701   www.nealrgross.com

1 effectiveness, and that the information

2 demonstrates that the device is as safe and

3 effective as the predicate device.

4 The PMA process, Pre-Market

5 Approval application, this applies to most

6 Class III devices, that's how they typically

7 enter the market. And unlike a 510(k)

8 application, a PMA is not typically a

9 comparison to other legally marketed devices,

10 but must provide information that stands on

11 its own to demonstrate the safety and

12 effectiveness of the device for its intended

13 use.

14 I just want to go through some of

15 the statutory definitions, as these are very

16 important. FDA defines safety as follows:

17 There is a reasonable assurance that a device

18 is safe when it can be determined, based upon

19 valid scientific evidence, that the probable

20 benefits to health outweigh any probable risk.

21 The valid scientific evidence used

22 to determine the safety of a device must

1 adequately demonstrate the absence of

2 unreasonable risk of illness or injury

3 associated with the use of the device for its

4 intended uses and conditions of use.

5 Among the types of evidence that

6 may be required to determine that there is

7 reasonable assurance that a device is safe,

8 are investigations using laboratory animals,

9 investigations involving human subjects, and

10 non-clinical investigations that may include

11 in vitro studies.

12 The FDA defines effectiveness as

13 follows: there is reasonable assurance that a

14 device is effective when it can be determined,

15 based upon valid scientific evidence, that in

16 a significant portion of the target

17 population, the use of the device will provide

18 clinically significant results.

19 One last topic that I just want to

20 touch on briefly is the least burdensome

21 concept. A central purpose of the Food and

22 Drug Administration Modernization Act of 1997

NEAL R. GROSS

COURT REPORTERS AND TRANSCRIBERS
1323 RHODE ISLAND AVE., N.W.
(202) 234-4433 WASHINGTON, D.C. 20005-3701 www.nealrgross.com

1   is to ensure the timely availability of safe

2   and effective new products that will benefit

3   the public. Congress' goal was to streamline

4   the regulatory process to improve patient

5   access to breakthrough technologies.

6   While Congress wanted to reduce

7   unnecessary burdens associated with the

8   premarket clearance and approval processes,

9   Congress did not lower the statutory criteria

10   for demonstrating substantial equivalence or

11   reasonable assurance of safety and

12   effectiveness.

13   FDA has defined the term least

14   burdensome as follows: a successful means of

15   addressing a pre-market issue that involves

16   the most appropriate investment of time,

17   effort, and resources on the part of industry

18   and FDA. This concept applies to all devices

19   and device components of combination products

20   regulated by FDA.

21   When conscientiously applied, FDA

22   believes the least burdensome concept will

NEAL R. GROSS

COURT REPORTERS AND TRANSCRIBERS
1323 RHODE ISLAND AVE., N.W.
(202) 234-4433   WASHINGTON, D.C. 20005-3701   www.nealrgross.com

1 help to expedite the availability of new

2 device technologies without compromising

3 scientific integrity in the decision-making

4 process or FDA's ability to protect the public

5 health. The least burdensome concept should

6 be integrated into all pre-market activities.

7   Under the least burdensome

8 approach, FDA applies the following basic

9 principles, and this will be my last slide:

10 The basis for all regulatory decisions will be

11 found in sound science in the spirit and

12 letter of the law. Information unrelated to

13 the regulatory decision should not be a part

14 of the decision making process.

15   Alternative approaches to

16 regulatory issues should be considered to

17 optimize the time, effort, and resources

18 involved in resolving the issue consistent

19 with the law and regulations. And finally,

20 all reasonable measures should be used to

21 reduce review times, and render regulatory

22 decisions within the statutory time frames.

1 Thank you.

2 DR. PETRICK: Hello. My name is

3 Nick Petrick. I'm the Deputy Director of the

4 Division of Imaging and Applied Math from the

5 Office of Science and Engineering Labs at the

6 FDA. I'm going to talk about general CAD

7 methods, what is a CAD, and some of the ways

8 that they have been evaluated. So the outline

9 will be to discuss, first, what is a CAD, to

10 give a basic introduction to some of the basic

11 components of CAD algorithms, to discuss a

12 little bit about the clinical implementations

13 of CAD, and then to discuss the evaluation of

14 the CAD algorithms.

15 So first, what is a CAD? We can

16 talk about two different basic components to

17 CAD. What we call a CADe, which is a

18 Computer-Aided Detection device. This is also

19 termed a C-A-D or a CAD. It is designed to

20 identify findings or regions on the image that

21 may be abnormal. And if you can just look at

22 the images on the right hand side, you can see

NEAL R. GROSS

COURT REPORTERS AND TRANSCRIBERS
1323 RHODE ISLAND AVE., N.W.
(202) 234-4433   WASHINGTON, D.C. 20005-3701   www.nealrgross.com

1 different types of marks. These are prompting

2 devices. They put marks on the image to

3 identify potential locations.

4   The second component is what's

5 termed a CADx or Computer-Aided Diagnosis.

6 This is also termed C-A-D in a number of

7 cases, so it's hard sometimes to differentiate

8 between a detection and a diagnostic device.

9 It is designed to process specific findings or

10 regions to characterize the findings.

11   And it could be things like the

12 likelihood of malignancy, so in this case,

13 this is a CADx device, but getting scores to

14 different regions on the image. It could be a

15 recommendation for clinical action. In this

16 example, maybe these are BI-RAD

17 categorizations to those particular lesions to

18 help the physician. But it is designed to

19 help the physician determine what he or she is

20 looking at.

21   CAD can be divided into a number of

22 different disciplines. It incorporates

1 engineering disciplines like image processing,

2 artificial intelligence, physics, the physics

3 of the imaging system, the medicine and

4 biology that goes into patient management, and

5 certainly statistics and mathematics.

6 What I'm going to do is show some

7 basic blocks to Computer-Aided Detection

8 algorithm. I have ordered them sort of going

9 through a streamline process. These are

10 blocks that could appear in any algorithm,

11 things like image processing, segmentation,

12 features, classification, but the sequence in

13 the block details would differ between CAD

14 algorithms.

15 A lot of CAD algorithms have

16 multiple branches. They certainly don't

17 necessarily go all in this order, but this

18 will give you at least a basic idea of how CAD

19 algorithms -- the different blocks in the CAD

20 algorithms.

21 So, the first component is data

22 acquisition. Digital data can come from

1 either digitized film, this happens in screen

2 film mammography and chest x-ray, or from

3 direct digital devices, things like full field

4 digital mammography or CT or many other

5 disciplines. Thank you.

6 Here I'm just going to show you an

7 example. I'll go through this example with

8 the mammography case. You probably can't see

9 the mass, but anyway, there's one there.

10 Another component is image processing. This

11 is where the image is enhanced or processed to

12 facilitate analysis. And you can see here

13 this is a filtered image trying to identify

14 potential structures within the breast.

15 Segmentation is identifying

16 boundaries or regions within the image. This

17 could be, like in this example, showing lesion

18 candidates where there is a lot of different

19 areas to identify in the image or it could be

20 organ segmentation. When breast is done, a

21 lot of times the breast boundary is identified

22 or lung field is identified to make sure that

1 the CAD is operating in the right region.

2 Another important element is

3 features. These are -- features characterize

4 regions or pixels within the image, things

5 like shape or texture or curvature within the

6 image. And just as an example here, I show

7 one of the regions identified in this

8 particular algorithm, and you can see that

9 area could be a feature or perimeter, so the

10 number of pixels that define the area of that

11 region or the number of pixels on the

12 perimeter would be features associated with

13 that.

14 An important component to features

15 is typically feature selection. This is a

16 process of selecting informative features. So

17 CAD algorithms in general sometimes start with

18 a large number of feature candidates, and then

19 they use a feature selection algorithm somehow

20 to reduce that to a useful number of features.

21 Classification is also critical to

22 most CAD algorithms. Classification is where

NEAL R. GROSS

COURT REPORTERS AND TRANSCRIBERS
1323 RHODE ISLAND AVE., N.W.
(202) 234-4433   WASHINGTON, D.C. 20005-3701   www.nealrgross.com

1 features are input to a learning algorithm to

2 come up with a single output score. And I

3 just show a simple example here, where you

4 have end features going in to some sort of

5 trained machine, and comes out with an object

6 score for that particular object.

7   There are many different types of

8 classifiers, as simple as a threshold as a

9 classifier, so there are algorithms that use

10 multiple thresholds as a starting point for

11 classification. Certainly, things like linear

12 discriminate analysis or linear classifiers,

13 are used as well as neural networks, which are

14 non-linear types of classifiers, but there are

15 many other types of classifiers as well that

16 different CAD algorithms may use.

17   It's important to keep in mind that

18 the training and test paradigm is critical to

19 the development of classification. So the

20 process of -- the data that's used to train

21 the algorithm or how it's -- the process of

22 training that algorithm will differentiate

NEAL R. GROSS

COURT REPORTERS AND TRANSCRIBERS
1323 RHODE ISLAND AVE., N.W.
(202) 234-4433   WASHINGTON, D.C. 20005-3701   www.nealrgross.com

1   different classifiers.

2   Once classifiers have been defined,

3   typically a threshold is applied to that

4   classification score, and you can just see in

5   the example here, that we started out with a

6   large number of regions. Once a threshold is

7   place on that classifier output, it reduces

8   the number of regions.

9   And then, the final output for a

10   Computer-Aided Detection algorithm is the

11   actual prompts. And here, I just show simple

12   Xs marking the centroid of different locations

13   within that image.

14   For our Computer-Aided Diagnosis

15   algorithm, that -- it is used to -- for

16   characterization of findings, so we'll

17   differentiate detection from this

18   characterization of findings. Many of the

19   steps are actually the same. And, in fact,

20   they could all be the same in these

21   algorithms. They have image processing,

22   feature selection, and classification. They

NEAL R. GROSS

COURT REPORTERS AND TRANSCRIBERS
1323 RHODE ISLAND AVE., N.W.
(202) 234-4433   WASHINGTON, D.C. 20005-3701   www.nealrgross.com

1 certainly include segmentation and other

2 steps. Again, the sequencing and block

3 details would differ between the different CAD

4 algorithms.

5   So here is just a simple example.

6 Again, the physician may identify these

7 regions or a different computer algorithm may

8 identify these different regions. What

9 differentiates these is, in this case, putting

10 some sort of score on those individual regions

11 that the physician would then use to

12 characterize those findings.

13   Again, to get back to the issue of

14 training CAD algorithms, training is the

15 process of systematically improving

16 performance for a set of data known as the

17 training dataset. It could be something like

18 maximizing sensitivity or maximizing the area

19 under the ROC curve. There are many other

20 types of maximization that occur here, but

21 there is some sort of process of maximizing

22 performance in the training process.

1   Training can be performed by the

2   computer, by some sort of regression or

3   optimization techniques.   A statistical

4   classifier typically incorporates some sort of

5   automatic training.   It also can be done by

6   humans. So the process of tweaking parameters

7   or combining parameters, making determinations

8   on what should go together, is also a form of

9   training that is used -- the developer's

10 knowledge to do that.

11   So, it could be either by computer

12 or by human. And again, in this -- and what

13 we are going to talk about here, is algorithms

14 that are fixed after training. So it's not --

15 they are not adapting to every new piece of

16 data that comes in, but there is a fixed

17 training set. The algorithm is then developed

18 and fixed at that point. They may change with

19 new revisions to that algorithm, but as it

20 goes through the FDA process, it's a fixed

21 algorithm. It does not change.

22   The training process incorporates

NEAL R. GROSS

COURT REPORTERS AND TRANSCRIBERS
1323 RHODE ISLAND AVE., N.W.
(202) 234-4433   WASHINGTON, D.C. 20005-3701   www.nealrgross.com

1   something that we call the learning curve.

2   Everyone with kids or people that work in

3   sports or anything else, really understands

4   the process of the learning curve. This is a

5   training, it's a dynamic process. When you

6   first start out, your performance may be

7   lower. Just like in CAD algorithms, with few

8   cases, the performance may be lower.

9   As more and more cases, here on the

10   X axis I'm just showing, the number of

11   patients per class going up, the performance

12   will increase up to some sort of asymptotic

13   performance. So increasing training data,

14   both increases performance to some asymptotic

15   level and decreases variability, so that the

16   error associated with that training process is

17   going down. And we just show the error bars

18   here.

19   So now to shift modes a little bit

20   into the clinical use of CAD. First, I'll

21   talk about CAD reading paradigms, and then

22   I'll just show you on the edge of the slides

1 the questions that these relate to for the

2 panel. And M refers to mammography, C to

3 colon, L to lung and G, which will show up a

4 little bit later, to the general section. So

5 these questions will hopefully be a way to

6 look back at these if you're interested.

7 So the CAD reading paradigms, I'll

8 talk about first, the first reader mode. This

9 is where the physician reviews only regions or

10 findings marked by the CAD device. So,

11 unmarked regions are not necessarily evaluated

12 by the physician. And just to let you know,

13 that no radiological CAD devices have been

14 approved or cleared for this particular mode,

15 but it's certainly a possible mode for a

16 device.

17 What is called the -- termed the

18 second reader CAD paradigm, this is where the

19 physician first conducts the complete

20 interpretation without the CAD, this is called

21 the so-called unaided read, then re-conducts

22 an interpretation with the CAD device, that's

NEAL R. GROSS

COURT REPORTERS AND TRANSCRIBERS
1323 RHODE ISLAND AVE., N.W.
(202) 234-4433   WASHINGTON, D.C. 20005-3701   www.nealrgross.com

1   the aided read.

2   This has also been termed second

3   detector or sequential reader type of CAD

4   reading paradigm. And examples of this are

5   mammography CAD devices and some of the lung

6   CAD devices that have been approved or cleared

7   by the Agency.

8   A final reading that I'll discuss

9   is what is termed concurrent read mode. This

10 is where physicians perform a complete

11 interpretation in the presence of the CAD

12 marks. So the CAD marks are available at any

13 time. It's up to the physician to determine

14 how they should use those marks. So, examples

15 of this are some of the colon CAD devices that

16 can potentially be used in this concurrent

17 read mode paradigm.

18   Other factors that influence

19 clinical use, one of the important ones, is

20 the physical characteristics of the marks.

21 And I have just given a reference at the

22 bottom here of some research paper that

1 investigates this. So physicians may respond

2 differently to different types of CAD marks.

3 So, if I show here segmented, here I show the

4 bottom slide segmented regions as opposed to

5 marking the centroid, the physician may

6 interact with those different types of marks

7 differently.

8   Another important characteristic is

9 the CAD stand alone performance. So things

10 like the number of CAD marks may influence

11 clinicians. The clinician's knowledge of the

12 sensitivity or the false positive rate may

13 also affect their confidence or their

14 attention in the CAD that they give to the CAD

15 marks.

16   And in general, the introduction of

17 CAD changes the clinical interpretation in

18 some way. One of the easy -- one of the most

19 apparent ways is changing potentially in the

20 reading time. For a second reader type of

21 CAD, this may result in an increase in review

22 time. For something like a concurrent CAD, it

1 may either maintain or potentially even

2 decrease the review time.

3   So now, I'll shift modes again into

4 discussing the evaluation of CAD algorithms,

5 and I'll break this up into two sections, non-

6 clinical first, and then clinical evaluation.

7   So non-clinical evaluation for my

8 definition, at least, will incorporate devices

9 in algorithm description and stability

10 analysis for the algorithms.

11   Algorithm description, just the

12 reason that this is important, is because

13 different CAD devices contain different

14 processing. So, even though they may contain

15 exactly the same steps and even maybe the same

16 types of classification, when you change the

17 ordering of those steps or you change some

18 basic parameters of those steps, they produce

19 different outputs.

20   So it's easier to access or compare

21 different devices if they are not considered

22 black boxes. We have had CAD devices that

NEAL R. GROSS

COURT REPORTERS AND TRANSCRIBERS
1323 RHODE ISLAND AVE., N.W.
(202) 234-4433   WASHINGTON, D.C. 20005-3701   www.nealrgross.com

1 have come into the Agency as black boxes, but

2 it becomes very difficult to compare the

3 technologies in those without having a

4 description of the information.

5 To understand the CAD device, at

6 least this type of information is necessary.

7 Certainly, there could be other information as

8 well. But information based on the patient

9 targeted for the device, if a device is

10 targeted to a different patient population,

11 they may or may not be comparable.

12 The device usage, so what is the

13 reading mode? Again, a change from a

14 concurrent to a second reader mode would imply

15 a different use for that particular device.

16 Differences in the image processing or

17 segmentation steps, different features or

18 different classifiers, and also very

19 importantly ,is the training and the training

20 data that is used to develop the algorithm.

21 Another component to non-clinical

22 testing is stability analysis. What we define

NEAL R. GROSS

COURT REPORTERS AND TRANSCRIBERS
1323 RHODE ISLAND AVE., N.W.
(202) 234-4433   WASHINGTON, D.C. 20005-3701   www.nealrgross.com

as stable algorithm is an algorithm that produces similar performance with changes in the algorithm, features, training, or training databases. And just to keep note that stability increases as either the number of training cases increases. If we use more data, we get a more stable algorithm.

The number of initial features decreases, so as the algorithm becomes less complex, so as complexity of the algorithm decreases, stability generally increases for the algorithm.

And to give you a little bit better feel for what stability means and why we're interested in stability analysis, one of the reasons we are interested is to indicate if the performance is due to a fortuitous combination of training and test data. It is possible that for whatever reason, the training data happens to match very well with the test data making it appear that the algorithm is very stable. But when it would

1   be used in general, that may not be the case.

2   It's also important to consider

3   stability analysis because algorithm updates

4   produce evolving performance. An algorithm

5   would come in at time 1, it's likely there

6   will be a revised version of that algorithm a

7   year from now or some time down the road. And

8   again, the performance of that would evolve.

9   Knowing something about stability, the

10   original algorithm may help to evaluate that

11   algorithm down the road.

12   What I'm showing here in yellow is

13   test error bars and in green are training

14   error bars. So the training error bars are

15   going to be associated with algorithm

16   stability, or the combination of training and

17   test error bars will be associated with

18   algorithm stability.

19   If we have very narrow error bars

20   here, then as this algorithm evolves over

21   time, it's likely that we will keep our

22   performance in some well-maintained region.

1 If, on the other hand, we have very wide error

2 bars, because we had very few trainers, it's

3 possible that that performance could degrade

4 significantly over time. It's also possible

5 the algorithm could actually get better over

6 time as well.

7 So stability tells us something

8 about how that initial algorithm will perform

9 with variations to the algorithm.

10 I'll now switch modes again into

11 clinical testing. To talk about clinical

12 testing, we have to first talk about the

13 hierarchical model of efficacy. And I'm not

14 going to go through all of these stages, but

15 it starts out with the basic elements of

16 technical efficacy, so physical and bench

17 testing of the device and goes all the way

18 through societal effect. So how well does

19 this device influence society overall?

20 In general, imaging technology

21 sponsors have generally focused on Levels 1

22 and 2 when they are trying to bring a product

NEAL R. GROSS

COURT REPORTERS AND TRANSCRIBERS
1323 RHODE ISLAND AVE., N.W.
(202) 234-4433   WASHINGTON, D.C. 20005-3701   www.nealrgross.com

1   to market through the FDA. Certainly,

2   sponsors in the FDA are not constrained to

3   these particular levels.

4   It is important to keep in mind

5   though, that as we move down this chain from

6   Levels 2, 3, 4 and 5, we would need to show

7   efficacy at all the prior levels in order to

8   maintain efficacy at the higher levels.

9   There are two different classes of

10   testing or clinical testing that I'll discuss.

11   The first, is standalone performance testing.

12   This is the performance of the device by

13   itself. And it's really a measure of the

14   intrinsic functionality of the device.

15   We also can talk about reader

16   performance testing. This is the performance

17   of the physician using the actual device, and

18   it's really the impact of the device on a

19   physician's performance.

20   So I'll start out talking about

21   standalone performance testing. And in

22   particular, I'll go through some of the basic

NEAL R. GROSS

COURT REPORTERS AND TRANSCRIBERS
1323 RHODE ISLAND AVE., N.W.
(202) 234-4433   WASHINGTON, D.C. 20005-3701   www.nealrgross.com

blocks that are necessary in order to do this type of evaluation. So the first thing I'll talk about is acquiring the test data set. The test data set are clinical images used to determine the safety and effectiveness of the CAD.

So it's different from the set used to develop or train the algorithm. And a lot of times, it's different from the set used to provide some initial validation for the CAD algorithm. It is representative of the target population and the target disease condition. And it usually includes a clinically relevant spectrum of patients, imaging hardware, and imaging protocols that may be appropriate for that particular CAD.

The test data can be acquired through something called a field test accrual. This would really be collected during real time clinical interpretation, or it can be acquired during enrichment accrual. And there are two different types of potential

1 enrichment. Enrichment for low-prevalence of

2 the disease, or the enrichment would be for

3 with disease cases at a higher proportion than

4 in population.

5 Certainly, this is a common

6 approach used whenever there is a very low

7 prevalence of disease in the population.

8 There can also be enrichment for stress

9 testing. This would be enrichment with cases

10 containing challenging findings. And

11 typically, stress testing usually includes

12 some sort of comparison of modality. So you

13 would be interested in stress testing

14 something when you have a technology already

15 on the market or a common technology already

16 used, something like comparing with and

17 without CAD as being a comparison.

18 One of the issues that you will be

19 asked to focus on today is the reuse of test

20 data. In the ideal testing paradigm for a

21 device, there would be the development of the

22 CAD algorithm. There would then be some sort

1   of collection of test cases, all completely

2   independent of development. The CAD would be

3   applied.   And the reported standalone   or

4   reader performance results would be reported.

5   What happens with CAD devices,

6   especially in new versions of the software, or

7   sponsors may want to compare performance on

8   revised algorithms with the same or expanded

9   versions of test data set? What problem that

10   comes up is that developers may have gained

11   knowledge or learned something by knowing the

12   performance of the original CAD on that test

13   data set.

14   So, even by the fact that they just

15   know the end performance, that's a form of

16   learning that goes on. For large data sets

17   and with minimal feedback, the knowledge gain

18   may be actually quite small. So what you will

19   be asked to discuss is there may be a

20   possibility -- it may be possible to reuse

21   test data under appropriate constraints to

22   streamline assessment.

NEAL R. GROSS

COURT REPORTERS AND TRANSCRIBERS
1323 RHODE ISLAND AVE., N.W.
(202) 234-4433   WASHINGTON, D.C. 20005-3701   www.nealrgross.com

1   And what you will be asked to focus

2   on are what may be appropriate constraints to

3   balance   both   data   integrity   and   data

4   collection process and accrual process. So we

5   have   just   discussed   the   test   database

6   collection.

7   Another   important   component   is

8   establishing ground truth, and what the rules

9   and methods are used to do in doing that

10   process. Ground truth includes two different

11   -- well, I have broken this up into two

12   different measures. One is whether or not

13   disease is present at the patient level. And

14   then, the location and/or the extent of that

15   disease at the lesion level.

16   So there are a couple of different

17   types of ground truth. There are certainly

18   more than two types of ground truths. If

19   there are cancerous lesions, ground truth may

20   be   based   on   something   like   biopsy   and

21   pathology. And then, for normal patients,

22   follow-up imaging.

1   For non-cancerous lesions,

2   something that might happen with lung nodules,

3   truth may be based on something on an expert

4   panel that reviews all available clinical

5   information. And certainly, there may be many

6   other types of truthing that happen.

7   So what's important to keep in

8   mind, is that ground truthing by expert panels

9   is almost always required to determine lesion

10 location. So in breasts, even though there

11 might be a biopsy and pathology that says this

12 patient has cancer, the actual location of

13 that lesion is going to still be up to some

14 clinician to identify where it is and

15 potentially the extent of that location.

16   It's also possible, again, that for

17 something like lung nodules or other types of

18 disease, that the experts may be used actually

19 to determine the presence of an abnormality as

20 well. It's important to keep in mind that

21 experts are susceptible to reader variability.

22   And in order to measure this type of truth

NEAL R. GROSS

COURT REPORTERS AND TRANSCRIBERS
1323 RHODE ISLAND AVE., N.W.
(202) 234-4433   WASHINGTON, D.C. 20005-3701   www.nealrgross.com

1 variability, then the use of multiple readers

2 would be one mechanism for evaluating that

3 truth variability.

4   So here, I just show a simple

5 example. At the patient level in this

6 particular case, pathology verified that the

7 cancer was seen in the left breast. At the

8 lesion level, the radiologist needs to

9 identify what actually is the region of that

10 particular lesion.

11   So the first case is pathology.

12 The second case, the clinician still has to

13 identify the location. And at the lesion

14 level, it also may be required for the

15 radiologist to segment the actual lesion to

16 actually look at the extent of the lesion. So

17 this is, you know, one potential segmentation.

18   So again, ground truthing, the

19 variations in that, variations in the readers

20 that perform that, were called variations in

21 the truth definition.

22   Another critical component for

1 assessment of CAD devices is establishing

2 scoring rules and methods. And scoring rules,

3 or what I'm defining, are used to determine

4 whether CAD marks the true lesion. So we have

5 the ground truth where the location is

6 identified. This is to determine whether the

7 actual CAD mark is on the true lesion.

8 One way of doing this may be the

9 overlap between the CAD and the truth. So we

10 will start out with our truth, determined by

11 some clinician. In this case, here is an

12 example of a particular CAD mark for that

13 truth. The question is is that CAD mark

14 marking the actual lesion? Is that a true

15 positive or false positive or false negative?

16 Well, there can be different rules

17 used. Rules become very critical to how this

18 is defined. In this case, we're looking at

19 the inner section of the CAD mark and the

20 truth divided by the area of the truth. And

21 you can see here, that we get a measure of

22 .39. If we change that rule to be the

NEAL R. GROSS

COURT REPORTERS AND TRANSCRIBERS
1323 RHODE ISLAND AVE., N.W.
(202) 234-4433   WASHINGTON, D.C. 20005-3701   www.nealrgross.com

1   intersection of the CAD and the truth divided

2   by the area of the CAD, you can see that we

3   get a measure of .94.

4   Based on different thresholds for

5   these, these could either be considered true

6   positives or potentially false negatives for

7   the device. Again, the truthing rule becomes

8   very critical to determine whether these are

9   actually called true positives or something

10   else.

11   Another common measure is the

12   distance between the CAD and the truth

13   centroid. Here in yellow I just show the

14   truth centroid, that blue X which is hard to

15   see is the CAD centroid. And the distance

16   measure is used. Again, a threshold on that

17   distance measure would determine whether that

18   is considered a true positive or not.

19   It's also possible to do scoring by

20   the physician. Where the physician -- again,

21   some physician or a group of physicians is

22   responsible for looking at the CAD mark,

NEAL R. GROSS

COURT REPORTERS AND TRANSCRIBERS
1323 RHODE ISLAND AVE., N.W.
(202) 234-4433   WASHINGTON, D.C. 20005-3701   www.nealrgross.com

1 looking at the truth and determining whether

2 they are marking the same location. Again,

3 all these components become critical in

4 evaluating the CAD performance.

5 Stand alone performance endpoints

6 include lesion-based sensitivity and number of

7 false positives per image. So here, I just

8 show a sensitivity false positive per image

9 care. These are typical performance measures.

10 That would be a binary-type of performance

11 measure.

12 We can also talk about something

13 called the Free Response Receiver Operating

14 Characteristic or FROC curve. This is a plot

15 of sensitivity versus the number of false

16 positives per image. And I'm not going to go

17 through the details here, but you can see that

18 it gives you all possible thresholds or all

19 possible performances for a particular CAD

20 algorithm. And later in this presentation,

21 I'll go into a little more detail about

22 exactly what that curve means.

NEAL R. GROSS

COURT REPORTERS AND TRANSCRIBERS
1323 RHODE ISLAND AVE., N.W.
(202) 234-4433   WASHINGTON, D.C. 20005-3701   www.nealrgross.com

1   We will switch again now into

2 reader performance testing. And I'll show you

3 a little bit more complicated flow chart.

4 Again, questions will be asked about both

5 standalone performance and reader performance

6 testing.

7   And just to go through this, we

8 have already talked about acquiring the test

9 database.   The problems associated with

10 standalone performance are similar for the

11 test database collection as well.

12   Again, there is a need to establish

13 truth, ground truth for these devices. Again,

14 the process would be similar between reader

15 performance and stand alone performance.

16 Likewise, the scoring rules and methods. So,

17 all these same basic issues apply. The only

18 box that is remaining here that I didn't talk

19 about, is identifying study readers.

20   Obviously, when you are talking

21 about reader performance, there are readers

22 associated with this process.   Readers are

NEAL R. GROSS

COURT REPORTERS AND TRANSCRIBERS
1323 RHODE ISLAND AVE., N.W.
(202) 234-4433   WASHINGTON, D.C. 20005-3701   www.nealrgross.com

1 generally selected to be representative of the

2 intended users, so they should be

3 representative of the clinicians who will use

4 the device and also representative of the

5 proper clinician experience level.

6   So, if for whatever reason this

7 device is supposed to be used by experts in

8 some way, however that is defined, then that

9 should be represented in the reader

10 population.

11   Reader performance testing depends

12 on the proper understanding and use of the CAD

13 device by the clinician, and proper

14 understanding and implementation of the study

15 protocols. This is an area that is sometimes

16 overlooked in reader performance testing.

17 Training becomes a very critical element. So

18 really, training of readers is a key to

19 achieving both of these endpoints.

20   A lot of times clinicians haven't

21 had experience with CAD, and they may not be

22 using it appropriately in the study.

1 Likewise, they may not understand how to

2 actually conduct the study protocol. And

3 again, training can go a long way to helping

4 this be achieved.

5 In designing reader studies, I'm

6 going to talk about common endpoints and then

7 some common CAD study designs that have been

8 used. So, we'll go into study endpoints. We

9 can talk about different levels of study

10 endpoints. We can talk about patient level of

11 analysis. These are things like sensitivity/

12 specificity or what we'll call ROC analysis,

13 which I'll talk about in a minute, or we can

14 talk about location-specific analysis. Things

15 like location-specific ROC or the Free

16 Response Receiver Operating Characteristic

17 curve.

18 So when would we be interested --

19 when would someone be interested in patient

20 level endpoints? Well, potentially, assessing

21 a Computer-Aided Diagnosis device, here, I

22 just show an example. This is an identified

67

region, say by the clinician. The CAD would give some sort of information. In this case, I give an example of the CAD giving a probability of malignancy score for this particular lesion.

The clinician would then use that information to come up with their own sort of assessment of this, the lesion, and determine some new probability of malignancy score for the physician.

Patient endpoints are also used to assess CAD when not accounting for a location. So here, we have one lesion per patient. We can look at the patient as the overall endpoint and do the assessment there. Again, patient-based analysis does not account for location of the lesion. So you get credit for being right, even though you marked the wrong location or you assess the wrong location potentially in a detection device.

There are different endpoints. We can talk about the binary decisions or a

1 single   threshold.   Again,

2 sensitivity/specificity operating points being

3 a single binary threshold, or we can talk

4 about rankings or ratings, which is a range of

5 thresholds.   And again, this would incorporate

6 what   is   called   the   receiver   operating

7 characteristic curve.

8   So I'll just go through and show

9 you examples of both of these.   Here we talk

10 about a CAD device, you know, potential CAD

11 device where this is a performance of the

12 reader alone.   I'm going to show you a plot on

13 the Y-axis of sensitivity versus what is

14 called the false positive fraction or 1.0-

15 specificity.

16   What happens a lot of time with CAD

17 devices in things like second readers is that

18 we have a higher sensitivity but a lower

19 specificity.   So we move along this axis.   The

20 question   becomes   is   that   tradeoff   in

21 specificity worth the increase in sensitivity?

22 There   are   certainly   many   other   possible

1 endpoints. Here I show a couple of prevalent

2 dependent endpoints, sensitivity/specificity

3 aren't typically prevalence related, but

4 positive predictive value and negative

5 predictive value would be.

6 Also, things like additional work-

7 up for additional cancers, again, would be

8 prevalence dependent. There are certainly

9 many other endpoints that could come and/or

10 have been used in the assessment of devices.

11 So those were single binary threshold/single

12 operating points.

13 We can also look at overall

14 distribution. Here I just show a distribution

15 in yellow of non-disease cases and that

16 distribution of disease cases. And on the X-

17 axis, I'm just showing the computer scores or

18 the CAD outputs. Keep in mind, that this

19 example is really showing you Gaussian or

20 normal distributions, but ROC does not depend

21 on normal distribution. So these

22 distributions don't have to look as neat as

1 this example shows.

2   Again, we have talked about one

3 threshold on this curve, we get one operating

4 point. If, on the other hand, we look at all

5 possible thresholds, we can plot the entire

6 curve.   And this determines the Receiver

7 Operating Characteristic curve.   So this is

8 all possible combinations of sensitivity and

9 false positive fraction.

10   If we go back to our example where

11 we are comparing these two different with and

12 without CAD, and determine whether that is a

13 good -- that with CAD read is better in some

14 sense than the reading alone, we can talk

15 about looking at the ROC curve.   So ROC can

16 facilitate this comparison.

17   Again, ROC requires ordering the

18 cases from least to most suspicious. Ratings

19 are often used in these studies, but it's not

20 the only way to facilitate this ranking of the

21 cases.   So just keep in mind, that while

22 ratings are common in ROC studies, they may

1 not be the only way to achieve that ranking,

2 not the only way to get ROC performance.

3   So, just in this toy example, you

4 can see that we ended up with a higher ROC

5 curve for the reader performance. So based on

6 the ROC assessment, at least, you might claim

7 that that CAD is actually helping readers.

8   What   are   some   of   the   common

9 performance metrics for ROC? They include the

10 area under the ROC curve.   Here, I've just

11 shown examples, so we're just going to look at

12 the total area.   Note, that since this is a

13 false positive fraction, it goes from 0 to 1

14 on the X-axis as well as 0 to 1 on the Y-axis.

15   So   this   is   the   average   true

16 positive fraction across all possible false

17 positive fractions.   We can also talk about

18 partial areas. We could take a region of this

19 space, and integrate that, and get some sort

20 of overall performance metric.

21   The challenge here is to link the

22 AUC measure to clinical relevance.   And just

NEAL R. GROSS

COURT REPORTERS AND TRANSCRIBERS
1323 RHODE ISLAND AVE., N.W.
(202) 234-4433   WASHINGTON, D.C. 20005-3701   www.nealrgross.com

1 as an example of what the area under the curve

2 means in a clinical perspective, if you had

3 two patients that came to your clinic, one had

4 disease, one did not have disease, the area

5 under the ROC curve is a measure of the

6 probability that you could differentiate those

7 two.   You could tell which patient was a

8 diseased one, which one was not the diseased

9 one.

10 We can also talk about location-

11 specific endpoints. These may be appropriate

12 when assessing Computer-Aided Detection

13 device. Again, when location is important or

14 when there is multiple prompts in the same

15 image. And here, we just show an example of a

16 lung, an x-ray lung CAD that has multiple

17 marks.

18 What is important to keep in mind

19 is that now the truthing rule is a critical

20 component. Changing the truthing rule could

21 change the performance, could change the

22 comparisons. So we have gotten to a more

NEAL R. GROSS

COURT REPORTERS AND TRANSCRIBERS
1323 RHODE ISLAND AVE., N.W.
(202) 234-4433   WASHINGTON, D.C. 20005-3701   www.nealrgross.com

1 specific location, but we have now

2 incorporated more complexity.

3   So again, location-specific ROC is

4 ROC analysis that requires the correct

5 location of the lesion. So it's basically the

6 same exact analysis as before, but you only

7 get credit if you actually mark the right

8 location. This is typically done with one

9 score location per patient. And again, the

10 location must be on the lesion. So this is

11 one way of doing location-specific analysis.

12   We can also talk about locations-

13 based operating points. These can incorporate

14 the sensitivity and the number of false

15 positive operating points. Here again, I'm

16 showing this plot of sensitivity versus the

17 number of false positives per image. In this

18 case, this isn't false positive fraction, so

19 that this X-axis can actually go, basically,

20 to infinity. You could have large, large

21 numbers of false positives.

22   You can compare with or without

1 CAD. Again, you often end up with a higher

2 sensitivity with more false positives. So we

3 have potentially the same possible conflict

4 that we had with ROC in the patient-based

5 assessment. It's difficult to compare just

6 based on two operating points. Certainly,

7 there are many other possible endpoints that

8 could be used here as well.

9 What's termed the Free Response

10 Receiver Operating Characteristic curve is

11 really all possible combinations of

12 sensitivity and false positives. So all

13 possible thresholds. Again, we can plot out

14 an entire curve.

15 The performance metric could be the

16 area under the FROC curve. Again, because

17 this curve can go into infinity, we need to

18 now select a region for that. And so in this

19 case, I selected something that was less than

20 three false positives per image.

21 What would be the appropriate

22 endpoint would again be based on what the

1 clinical -- what makes clinical sense, if that

2 were used. There are other types of measures;

3 something called the area under the

4 alternative FROC curve or AFROC curve. Again,

5 I'm not going to go into the details of that,

6 but that's another measure that has been used.

7 Again, the challenges are to link

8 these measures to clinical relevance and also

9 in some of these cases, to develop statistical

10 methodology. I'll show you some of this a

11 little bit later of what has been developed,

12 but certainly there are still some challenges

13 remaining there.

14 So I'll switch again into

15 evaluating reader study designs that have been

16 used. First, I'll just introduce prospective

17 and retrospective, and then I'll talk about a

18 couple of the designs that have been used for

19 CAD devices. Again, you will have questions

20 from your panel for the panel to respond to

21 these.

22 I will go just briefly through

1 prospective and retrospective. The next

2 speaker will go into a little more detail on

3 some of the issues related to these study

4 designs.

5 A prospective reader study is when

6 CAD performance is measured as part of the

7 actual clinical practice. Again, field

8 testing of the CAD device.

9 Retrospective reader study would be

10 when cases are collected prior to image

11 interpretation. Again, typically, some sort

12 of enrichment or stress testing database is

13 used, but it wouldn't have to be that way.

14 The cases are read off-line by one or more

15 readers under specific reading conditions.

16 And examples of when this has been

17 used are some mammography CAD devices and lung

18 nodule CAD devices fused as actual reading

19 design. There are two reading designs. I'm

20 going to turn to the Warren- Burhenne Study

21 Design. This has been used in mammography CAD

22 devices. This is really two separate studies.

1 It's a retrospective study of CAD sensitivity

2 to detect abnormalities missed in clinical

3 practice, and it estimates relative reduction

4 in false negative rate with the CAD.

5   A second independent study was

6 commonly a prospective study of the work-up

7 rate of readers with or without CAD in

8 clinical practice. Again, the difference in

9 the work-up rate is attributed to the use of

10 the CAD device.

11   There are some -- there is a -- the

12 fundamental limitation is that the reduction

13 in the false negative rate and the increase in

14 the work-up rate are being evaluated in

15 different studies. So they are not being

16 evaluated in the same study. This leads to

17 the study design that could be difficult to

18 interpret statistically.

19   And again, the design goal here is

20 to estimate the potential effect on the false

21 negative rate, not necessarily the actual

22 effect that happens in clinical practice.

1   Another commonly used study design

2   for CADs is what's called the Multi-Reader

3   Multi-Case study design or the MRMC study

4   design.   This is a study where a set of

5   readers interpret a set of patient images in

6   each of two competing reading modalities. And

7   in the general design, this doesn't have to be

8   two.   It could be three, four or five,

9   whatever you want.

10   But for a CAD device, it would

11   typically be something like with or without

12   CAD.   And this could be either a prospective

13   or a retrospective study.   It's not limited.

14   So this is a very general framework for doing

15   clinical studies.

16   What is often termed the fully

17   cross design is when you have all readers read

18   all cases in both modalities or whatever the

19   competing modalities are.   This is the most

20   statistically powerful design for a given

21   number of cases.   But there are certainly

22   hybrid designs that may become very important,

NEAL R. GROSS

COURT REPORTERS AND TRANSCRIBERS
1323 RHODE ISLAND AVE., N.W.
(202) 234-4433   WASHINGTON, D.C. 20005-3701   www.nealrgross.com

1   especially in prospective settings, where

2   maybe the physician can only read their own

3   cases. So there are many hybrid designs and

4   these are also invaluable. So statistically,

5   these can be validated and evaluated.

6   The importance of the MRMC study

7   design is it's generalized, it's both new

8   readers and new cases. So both cases and

9   readers are random effects. The advantages

10 include greater statistical power for a given

11 number of cases.

12   It's also sometimes confusing that

13 the MRMC study design can only be used with

14 ROC area. This is not the case. The study

15 design is general. You could use any type of

16 endpoint. They could be location-specific

17 endpoints, they could be ROC endpoints. They

18 could be sensitivity/specificity endpoints.

19 And the real advantage is the MRMC studies are

20 generally statistically interpretable.

21   So here is -- again, we can talk

22 about patient level MRMC analysis. These

1 would include sensitivity or ROC endpoints.

2 And I'm just going to give you a number of the

3 different methodologies and tools that have

4 been developed to analyze these study designs.

5 We can also talk about location-

6 based MRMC analysis. This is where the

7 analysis counts for correct localization of

8 the lesions. And again, there are a number of

9 different methods that have been developed and

10 statistically, they are either being evaluated

11 or have been evaluated, validated to do these

12 location-specific analysis for particular

13 endpoints.

14 So I'm going to end there, and I'm

15 going to go off for Tom to come and talk about

16 further statistical issues associated with CAD

17 algorithm development.

18 DR. GWISE: Good morning. My name

19 is Thomas Gwise, and I'm going to talk to you

20 about some statistical aspects in CAD

21 evaluations. Here is an outline of my

22 presentation. I'll review some statistical

1 concepts that apply to diagnostic studies.

2 Then we will look at some reader study

3 designs, compare prospective and retrospective

4 study designs, look at two specific

5 retrospective reader study design examples,

6 and go over some complications in these

7 retrospective study designs. And in the end,

8 I'll discuss standalone studies versus reader

9 studies.

10 First of all, two dimensions are

11 always considered when evaluating the

12 diagnostic test performance. How well can the

13 test detect disease cases, and how well can

14 the test correctly identify the non-disease

15 cases? These are sensitivity/specificity,

16 respectively.

17 One important thing to note,

18 sensitivity/specificity are not comparable if

19 they are not estimated in the same study.

20 Does the test add value? When we are thinking

21 of a diagnostic test, we would like to know if

22 it adds any value above what could be done

1 without the test. For example, is the

2 diagnostic test for bone mineral density

3 better than just using a person's age in

4 diagnosing osteoporosis?

5 Another example could be does the

6 use of a CAD device improve diagnostic

7 performance of readers? A couple examples of

8 improvement could be improvement in

9 sensitivity and specificity, and improvement

10 in ROC/AUC area under the curve or, for

11 example, in a concurrent mode CAD use, maybe a

12 decrease in reading time with the same

13 performance and sensitivity and specificity.

14 The vast majority of CAD

15 submissions to date have been for those

16 labeled as second reader aids to physicians.

17 As such, it's expected that using the device

18 in accordance with the label will improve

19 performance of the physician.

20 I'll now discuss some reader

21 studies. First, because the study conducted

22 matches the intended use, it's generally

1 believed that a good way to test for a change

2 of performance is to do a multi-center

3 prospective randomized clinical trial.

4 For example, randomized patients to

5 the respective experimental conditions, that

6 being unassisted image reading and CAD-

7 assisted image reading, manage the patients

8 according to the evaluations as in routine

9 clinical practice, follow-up patients to

10 determine the true disease state, and analyze

11 the results to make the comparison between the

12 two devices or experimental conditions.

13 Prospective studies. The study

14 conducted matches the indications for use,

15 that is the routine clinical practice where

16 reader decisions affect patient management.

17 And we review -- and we consider that as a

18 positive aspect of these designs.

19 Some drawbacks to prospective

20 randomized trials, for the intended use -- for

21 intended use populations where disease

22 prevalence is low, a prospective study, as

1 described, would require large amounts of time

2 and result in large enrollments to obtain

3 enough data to make the comparison. Another

4 drawback is the risk to participants, if

5 patient management will depend on readings in

6 the study. And in that case, an

7 investigational device exemption may be

8 required.

9 Now, I'll discuss some

10 retrospective reader studies. Retrospective

11 reader studies are -- reader evaluations are

12 made off-line on a retrospective dataset of

13 images on which disease status with patients

14 has been established according to the ground

15 truthing rules. Multi-Reader Multi-Case

16 designs are often used, and the sample of

17 images is often enriched with disease cases.

18 Some positive aspects of

19 retrospective reader studying are: They are

20 not a significant risk because reader results

21 are not used in patient management. They are

22 very efficient, relatively small sample size

1 can result in precise estimates of

2 sensitivity/specificity in ROC curves.

3 Some drawbacks to retrospective

4 reader studies are the reading behavior in the

5 study may not be the same as in routine

6 clinical practice because the readers know

7 their readings do not matter to the patients,

8 and readers may detect enrichment that could

9 affect their reading behavior.

10 Enrichment will also result in some

11 biases. And I'll get to those more in a

12 moment. Also, a small number of readers may

13 not generalize to the population.

14 Now, I'll talk about two specific

15 retrospective reader study designs. The first

16 one is, I'm calling it, the sequential reading

17 study design. This design is for CAD acting

18 as a second reader. Every reader in the study

19 reads every image in the image set under the

20 two reading conditions, unassisted and CAD-

21 assisted.

22 So in this design, a given reader,

1 first, reads an image unassisted and records a

2 without-CAD score. Immediately, the reader

3 re-reads the image with the CAD-assistance and

4 records a with-CAD score. When the reading

5 phase is complete, each reader will have

6 recorded a with-CAD and a without-CAD score

7 for each image.

8 Follow-up already having been

9 completed, the data can be analyzed to make

10 the comparison. So this animation illustrates

11 this study design. This is the unassisted

12 reading and the score taken, followed

13 immediately by the CAD-assisted reading and

14 the CAD-assisted score being taken. The thing

15 to notice here is that the first reading is a

16 component of both scores.

17 Some positive aspects of this study

18 design are the intra-reader variability is

19 minimized by having the readings under the two

20 modalities as close in time as possible.

21 Minimizing this variability, increases the

22 precision of the estimates, which will in turn

1 increase the statistical power to detect the

2 difference between the modalities.

3 Some drawbacks to this design, for

4 each reader-image pairing the without-CAD read

5 is part of the with-CAD read. A study using

6 this design is open to reader behavior changes

7 because the reader knows modalities are

8 combined. For example, a reader may

9 subconsciously want to compete with the CAD

10 device and, therefore, be more aggressive than

11 usual in searching for lesions during the

12 unaided portion of the two part reading.

13 Another example, a reader could

14 become less vigilant trusting the CAD to

15 prevent false negatives.

16 To minimize test interpretation

17 bias, that's the bias I just described, the

18 previous design could be altered to include a

19 time period between the unaided and CAD-

20 assisted reading sessions. This is often

21 called a washout period. And this illustrates

22 -- this animation just illustrates how this

1 might work. I'm using May 1st as a reference

2 here. And we have the unassisted read and the

3 score followed by our washout period of,

4 approximately, four weeks here. And then the

5 first part of the second reader CAD modality

6 followed by the CAD-assist and the scoring.

7   Now, for randomly selected images,

8 we could switch the order of these reading

9 sessions, and there are other possible designs

10 that are plausible using this washout period.

11   Now, comparing these two types of

12 designs, the independent reading session

13 design requires two sessions, thus it is more

14 time consuming than the sequential design.

15 For the same sample size, the independent

16 design is less powerful in detecting

17 differences between image reading modalities

18 than the sequential reading design.

19   The sequential reading design is

20 subject to possible test interpretation bias

21 where the independent reading design attempts

22 to control for it.

1   Now, I'll talk about some

2   complications, some other complications, in

3   retrospective reader studies. We will discuss

4   enrichment-related bias, choice of controls,

5   reader variability issues, and how disease

6   localization is addressed.

7   First, I will talk a little bit

8   about the image sample. When collecting the

9   study sample, the investigator intentionally

10   or otherwise, could systematically affect the

11   case mix of images to favor one modality over

12   the other. This is known as selection bias.

13   Using a mammography study for

14   example, an investigator may tend to be less

15   aggressive in recruiting younger patients

16   having denser breasts than in recruiting older

17   patients. If the CAD performs poorly on

18   images of dense breasts, the study results

19   will be biased in favor of better CAD

20   performance.

21   A combination of administrative

22   controls and prospectively collecting cases

1   from several centers may limit this effect.

2   Enrichment is the process of supplementing the

3   image sample with disease positive images.

4   Performance estimates obtained with enriched

5   study samples will likely be different than

6   the performance in the intended use

7   population.

8   Difference in performance between

9   modalities may be qualitatively generalizable

10   to the intended use population if the spectrum

11   of disease is properly represented. Different

12   case mixes of lesion types will likely result

13   in different performance estimates. This is

14   known as spectrum effect.

15   For example, in mammography, a CAD

16   may have more difficulty detecting some masses

17   than microcalcifications. A small sampling

18   which the proportion of microcalcifications to

19   masses is large will give a higher performance

20   estimate than a sample in which that

21   proportion is smaller.

22   Consider a sample of images

1 enriched with the large proportion of disease

2 positive cases easily detected by readers and

3 CADs. Performance estimates for both

4 modalities will likely be high. This could

5 possibly make it difficult to detect a

6 difference in performance between the two

7 modalities.

8 A stress test is a study in which a

9 sample of images is enriched with a large

10 proportion of positive cases considered to be

11 difficult to detect by readers and CADs. The

12 goal is to show that the device can add value

13 in cases that are difficult for readers.

14 Performance results obtained from the study on

15 enriched samples cannot easily be generalized

16 across studies.

17 So more on enrichment. Readers in

18 the study environment will likely become aware

19 of the enrichment and could change their

20 reading behavior in response. Investigators

21 attempt to mitigate this context by estimating

22 relative performance. Study results depend in

1   part on distribution of clinical variables.

2   For example in mammography, breast

3   density, size of cancer, proportion and types

4   of masses and microcalcifications, studies

5   vary in clinical variable distribution; and

6   therefore, limiting the comparability of the

7   results that are put in the labeling of the

8   devices.

9   A standardized analysis attempts to

10   adjust to a standard population that

11   represents the target population. For

12   example, direct standardization is a common

13   method used in epidemiology to compare disease

14   rates among populations with different

15   distributions of categorical variables such as

16   age groups, sex and race, et cetera.

17   In direct standardizations for each

18   categorical variable, we weight the categories

19   according to their distribution in the

20   standard population. Performance is estimated

21   as an average of category-specific

22   performance.

1   Direct standardization has been

2 successfully used to support device

3 applications. Standardization can reduce

4 estimator bias, but it can also increase the

5 estimated variability.

6   Now, I'll spend a few minutes

7 talking about comparative modalities or

8 controls. And I'll just mention here that

9 most CADs are labeled as second readers, and

10 this can complicate the choice of controls.

11 In a study, it's desirable to control for all

12 possible confounders and isolate the desired

13 effect.

14   In the case of CAD reader studies,

15 the effective interest is the change in

16 performance due to using the CAD. CAD-

17 assisted reading performance is often compared

18 to performance of an unassisted single reader.

19 Now, consider CAD labels that require the

20 reader to perform a second read of the image

21 supplemented by CAD information.

22   There may be an increase in reader

1 performance, if after reading the image alone,

2 the readers were instructed to simply look

3 again, and the extra time may add some

4 benefit. Comparing CAD-assisted reads to

5 unassisted reads does not control for this

6 possibility.

7 Studies have compared CAD-assisted

8 reading performance to performance of double

9 reading where a second reader reads the images

10 and reviews the findings of the first reader.

11 Now, I'll spend a few minutes

12 talking about reader variability. The data

13 I'm about to show you is from a study by Beam

14 et al, in which 108 U.S. mammographers reading

15 a common set of 79 mammograms provided a

16 rating of suspicion of disease using the BI-

17 RAD scores where 5 is the highest level of

18 suspicion of cancer.

19 And here we see this data. The

20 data I'm showing you is the sensitivity and

21 specificity, each point represents the

22 sensitivity and specificity of one of the 108

NEAL R. GROSS

COURT REPORTERS AND TRANSCRIBERS
1323 RHODE ISLAND AVE., N.W.
(202) 234-4433   WASHINGTON, D.C. 20005-3701   www.nealrgross.com

1 readers. And we can see that the performance

2 of the readers varies from very well to not so

3 well.

4   Companies have submitted studies

5 with from 5 to 20 readers. Reader samples

6 should represent the intended use population

7 of readers. A small number of readers may not

8 be representative of the reader population.

9   Now, I'll just make a few comments

10 about the location of disease and per patient

11 analysis, as Dr. Petrick discussed earlier.

12 Per patient sensitivity/specificity will not

13 correctly represent the location of the

14 disease. And this image is an example of

15 this. The disease in this patient is actually

16 located over here, but the device calls the

17 disease over here.

18   This is typically represented as a

19 true positive finding in the analysis. It

20 could also be considered a false positive or a

21 false negative. And this also adds to the

22 complications to the statistical analysis of

1   the study.

2   Now, considering all of the
3   possible complications and more that I have
4   just discussed, we must consider the effect
5   size and the context of the study design
6   quality. A small statistically significant
7   difference in performance across the two
8   modalities could possibly be explained by
9   study biases.

10   Later on in the question session,
11   we will ask you to discuss effect size in this
12   context.

13   This is just a recap of comparison
14   of retrospective and prospective reader
15   studies. Retrospective reader studies are
16   smaller and less time consuming than
17   prospective studies. Prospective reader
18   studies are conducted as the device is
19   intended to be used in the intended use
20   population.

21   Reader behavior in retrospective
22   studies may not represent clinical practice

1   because the readers know that their readings

2   will not affect patient management.

3   A few words on standalone studies.

4   Standalone studies, as described earlier by

5   Dr. Petrick, can be useful in comparing a CAD

6   device to a previous version of itself or

7   investigating the performance of the device

8   without the reader. Standalone studies suffer

9   the same complications as reader studies with

10   respect to sample enrichment, that is, the

11   study results are not generalizable across the

12   studies.

13   Companies have proposed reusing

14   test data in evaluating updated version of

15   their CADs. We could consider each upgrade

16   iteration to be training. Testing on training

17   data will likely provide unreliable results.

18   We could consider this teaching to the test or

19   in other words fitting to the noise. It's

20   difficult to quantify this bias.

21   And finally, comparing reader

22   studies to standalone studies. Reader studies

1 investigate reader-device interaction, and

2 standalone studies investigate only device

3 performance. Thank you for your attention.

4 CHAIRMAN GLASSMAN: I want to thank

5 all of our three speakers, Dr. Smith, Petrick

6 and Gwise, for their presentations. Does the

7 panel have any questions for the FDA, at this

8 time? Please, let me know. No, we don't.

9 Okay. Great.

10 If no one has any questions, we

11 will now take a 15 minute break and resume at

12 9:50. Just two quick remarks. One, members

13 of the public, who are not -- who are out

14 there and members of the panel who are here,

15 there is an invisible line between us. We

16 request that members of the public not invade

17 the panel space, not because either you or I

18 have some communicable disease, but because

19 there are documents back here that are private

20 and you are not supposed to look at.

21 We also have a limited number of

22 danishes on our side. So, also for members of

1   the panel, let me remind you that outside of

2   the formal meeting time, we are not to discuss

3   what we have just heard. So you can all

4   congratulate me on my new granddaughter, but

5   that's about it. Thank you. Let's break for

6   15 minutes.

7   (Whereupon,   the   above-entitled

8   matter went off the record at 9:37 a.m. and

9   resumed at 9:53 a.m.)

Formatted: Different first page

10   CHAIRMAN   GLASSMAN: If   everybody

11   will take their seats, I would like to get

12   started. We will now proceed with the FDA

13   presentation   highlighting   current   issues

14   related to mammography CAD.   Dr. Robert Smith

15   is going to speak. Robert?

16   DR.   SMITH: Thank   you,   Dr.

17   Glassman. The primary purpose of mammography

18   CAD   devices   is   to   reduce   errors   when

19   interpreting screening mammograms.   Screening

20   mammograms are performed to identify patients

21   with breast cancer.   Breast cancer is the most

22   commonly   diagnosed   cancer   and   the   second

NEAL R. GROSS

COURT REPORTERS AND TRANSCRIBERS
1323 RHODE ISLAND AVE., N.W.
(202) 234-4433   WASHINGTON, D.C. 20005-3701   www.nealrgross.com

1   leading cause of cancer-related death among

2   women in the United States.

3   Based on the latest FDA statistics,

4   there are over 13,000 accredited mammography

5   machines in the United States that perform

6   approximately 36 million annual mammography

7   procedures, and approximately 80 percent of

8   all mammography examinations are performed for

9   screening.

10   The clinical mammographic and

11   pathologic characteristics of patients who

12   undergo screening mammography in the United

13   States are well-known from large published

14   clinical trials and from publicly available

15   databases.

16   The largest publicly available

17   database is the Breast Cancer Surveillance

18   Consortium.   This database   contains

19   information on 6 million mammograms for more

20   than 2 million women and contains over 74,000

21   breast cancers.

22   It is important to look at the