

FOOD AND DRUG ADMINISTRATION  
CENTER FOR DEVICES AND RADIOLOGICAL HEALTH  
RADIOLOGICAL DEVICES ADVISORY PANEL  
MEETING

TUESDAY,  
FEBRUARY 3, 2004

The Panel met at 9:00 a.m. in Salons B-D of the Gaithersburg Marriott Washingtonian Center, 9751 Washingtonian Boulevard, Gaithersburg, Maryland, Geoffrey S. Ibbott, Ph.D., Acting Chairman, presiding.

PRESENT:

GEOFFREY S. IBBOTT, Ph.D., Acting Chairman  
BRENT BLUMENSTEIN, Ph.D., Temporary Voting Member  
CHARLES B. BURNS, M.S., P.H., Non-Voting Consumer Rep.  
EMILY F. CONANT, M.D., Voting Member  
THOMAS FERGUSON, M.D., Temporary Voting Member  
ELIZABETH KRUPINSKI, Ph.D., Temporary Voting Member  
MINESH P. MEHTA, M.D., via teleconference, Chairman  
DEBORAH J. MOORE, Non-Voting Industry Representative  
STEPHEN SOLOMON, M.D., Temporary Voting Member  
DAVID STARK, M.D., Temporary Voting Member  
PRABHAKAR TRIPURANENI, M.D., Voting Member  
ROBERT DOYLE, Executive Secretary

FDA REPRESENTATIVES:

NANCY BROGDON  
NICHOLAS PETRICK, Ph.D.  
ROBERT A. PHILLIPS, Ph.D.  
WILLIAM SACKS, Ph.D., M.D.  
ROBERT F. WAGNER, Ph.D.

SPONSOR REPRESENTATIVES:

RONALD CASTELLINO, M.D.  
PABLO DELGADO, M.D.  
HEBER MacMAHON, M.D.  
DAVE MILLER  
KATHY O'SHAUGHNESSY, Ph.D.

**NEAL R. GROSS**

COURT REPORTERS AND TRANSCRIBERS

1323 RHODE ISLAND AVE., N.W.

WASHINGTON, D.C. 20005-3701

(202) 234-4433

[www.nealrgross.com](http://www.nealrgross.com)

## A-G-E-N-D-A

**Open Session**

Call to order and the Panel Introduction, Dr. Geoffrey Ibbott, Ph.D., Acting Chairman .....	4
FDA Introductory Remarks, Robert J. Doyle, Executive Secretary .....	7
Update on FDA Radiology Activities, Robert A. Phillips, Ph.D. ....	13

**Open Public Hearing**

Open Public Hearing; interested persons may present data, information, or views, orally or in writing, on issues pending before the committee .....	14
---	----

**Open Committee Discussion**

Charge to the Panel, Dr. Geoffrey Ibbott, Ph.D. ...	16
Overview of Contemporary ROC Methods, Robert F. Wagner, Ph.D. ....	17

## Presentations on P030012 by Sponsor

Introduction, Kathy O'Shaughnessy, Ph.D. ....	95
Current Clinical Practice, Heber MacMahon, M.D. ....	97
Device Description and Clinical Trial	
Introduction, Ronald Castellino, M.D.	103
Clinical Study, Dave Miller .....	115
User Experience, Pablo Delgado, M.D. ....	143
Summary, Kathy O'Shaughnessy, Ph.D. ....	148

Lunch

**NEAL R. GROSS**

COURT REPORTERS AND TRANSCRIBERS

1323 RHODE ISLAND AVE., N.W.

WASHINGTON, D.C. 20005-3701

(202) 234-4433

www.nealrgross.com

Presentations on P030012 by FDA

PMA Overview, Robert Phillips, Ph.D. .... 174  
 Clinical Background, William Sacks,  
 Ph.D., M.D. .... 175  
 Clinical Results, Nicholas Petrick, Ph.D. ... 179  
 PMA Review Summary, William Sacks,  
 Ph.D., M.D. .... 202

Reports by Panel Lead Reviewers

David Stark, M.D. .... 212  
 Brent Blumenstein, Ph.D. .... 225

Presentation of FDA Questions ..... 232

Break

Panel Discussion ..... 234

**Open Public Hearing**

Open Public Hearing: interested persons may  
 present data, information, or views, orally or  
 in writing, on issues pending before the  
 committee ..... 309

**Open Committee Deliberations**

Panel Recommendation(s) and vote ..... 311

Adjourn

**NEAL R. GROSS**

COURT REPORTERS AND TRANSCRIBERS

1323 RHODE ISLAND AVE., N.W.

WASHINGTON, D.C. 20005-3701

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

2 9:06 a.m.

3 DR. IBBOTT: I would like to call this  
4 meeting of the Radiological Devices Panel to order. I  
5 also want to request that everyone in attendance at  
6 this meeting be sure to sign in at the attendance  
7 sheet that is available outside the door. I would  
8 note for the record that the voting members present  
9 constitute a quorum and is required by 21 CFR Part 14.

10 At this time I would like each panel  
11 member at the table to introduce himself or herself  
12 and state his or her specialty, position title,  
13 institution, and stages on the panel.

14 I'll begin with myself. Some of you have  
15 already figured out that I'm not Dr. Mehta. Thanks to  
16 the vagaries of air travel and weather, Dr. Mehta is  
17 unable to be here but is joining us by speaker phone.

18 I'm Geoff Ibbott. I'm a medical  
19 physicist. I work at the University of Texas, M.D.  
20 Anderson Cancer Center in the Department of Radiation  
21 Oncology and Radiation Physics. I'm a voting member

**NEAL R. GROSS**

COURT REPORTERS AND TRANSCRIBERS

1323 RHODE ISLAND AVE., N.W.

WASHINGTON, D.C. 20005-3701

(202) 234-4433

[www.nealrgross.com](http://www.nealrgross.com)

1 on this panel and have been for several years.  
2 Obviously I'm standing in as chair for this meeting.

3 Then, Charles, let's start with you and  
4 we'll go around the table and introduce ourselves.

5 MR. BURNS: Charles Burns, Professor of  
6 Radiologic Science at the University of North  
7 Carolina. My primary expertise is Imaging Diagnostic  
8 Physics and I'm a nonvoting consumer representative

9 DR. IBBOTT: Thank you.

10 DR. MOORE: I'm Deborah Moore. I'm the  
11 Vice President of Regulatory and Clinical Affairs for  
12 Proxima Therapeutics. I'm the industry representative  
13 for the panel and a nonvoting member.

14 DR. STARK: I'm David Stark. My current  
15 title is President of MRI of Dettum in Massachusetts.  
16 I'm a clinical radiologist. I've been a chairman for  
17 close to nine years and I know many of you. I'm  
18 pleased to be here. Thank you.

19 DR. TRIPURANENI: Prabhakar Tripuraneni.  
20 I'm head of Radiation Oncology at Scripps Clinical in  
21 La Jolla, California. I have a practice and full-time  
22 clinician radiation oncologist and I am a voting

**NEAL R. GROSS**

COURT REPORTERS AND TRANSCRIBERS

1323 RHODE ISLAND AVE., N.W.

WASHINGTON, D.C. 20005-3701

1 member. I think this is my first or second date on  
2 the panel.

3 DR. DOYLE: I'm Bob Doyle. I'm the Exec.  
4 Sec. of this panel.

5 DR. BLUMENSTEIN: I'm Brent Blumenstein.  
6 I'm a biostatistician in private practice. I'm  
7 normally on the General and Plastic Surgery Panel.

8 DR. SOLOMON: I'm Steve Solomon. I'm a  
9 radiologist at Johns Hopkins Hospital. I'm a  
10 consultant to the panel.

11 DR. FERGUSON: I'm Tom Ferguson, professor  
12 emeritus of cardiothoracic surgery at Washington  
13 University School of Medicine, St. Louis. I'm a  
14 temporary voting member on this panel. I'm on the  
15 Cardiovascular Device Panel.

16 DR. CONANT: I'm Emily Conant. I'm the  
17 Chief of Breast Imaging at University of Pennsylvania  
18 and sort of half research and half clinical at this  
19 point. I'm a voting member.

20 DR. KRUPINSKI: I'm Elizabeth Krupinski  
21 from the University of Arizona. I'm a research  
22 professor in the Department of Radiology. My area of

**NEAL R. GROSS**

COURT REPORTERS AND TRANSCRIBERS

1323 RHODE ISLAND AVE., N.W.

WASHINGTON, D.C. 20005-3701

1 expertise is observer performance and image perception  
2 studies. I'm a voting member.

3 MS. BROGDON: I'm Nancy Brogdon. I'm not  
4 a member of the panel. I'm the liaison to the agency.  
5 I'm the Director of the Division of Reproductive  
6 Abdominal and Radiological Devices.

7 Dr. Mehta, would you like to introduce  
8 yourself?

9 DR. MEHTA: Yes, please. I'm Minesh  
10 Mehta. I'm a radiation oncologist in terms of  
11 specialty and I'm the Chair of the Department of Human  
12 Oncology at the University of Wisconsin. Generally  
13 when I'm there I'm chair of the panel but today I  
14 guess I'm listening in.

15 DR. IBBOTT: All right. Thank you,  
16 everyone. Mr. Doyle would now like to make some  
17 introductory remarks.

18 DR. DOYLE: Well, first on the agenda here  
19 is appointment of the Acting Chairperson. Pursuant to  
20 authority granted under the Medical Devices Advisory  
21 Committee Charter dated October 27, 1990, and as  
22 amended August 18, 1999, I appoint Geoffrey Ibbott,

**NEAL R. GROSS**

COURT REPORTERS AND TRANSCRIBERS

1323 RHODE ISLAND AVE., N.W.

WASHINGTON, D.C. 20005-3701

(202) 234-4433

[www.nealrgross.com](http://www.nealrgross.com)

1 Ph.D., as Acting Chairperson of the Radiological  
2 Devices Panel Meeting on February 3, 2004. This is  
3 signed by David Feigal, the Director of the Center of  
4 Devices and Radiological Health.

5 Now I would like to read the appointment  
6 of temporary voting status. Again pursuant to the  
7 authority granted under the Medical Devices Advisory  
8 Committee Charter dated October 27, 1990, and as  
9 amended August 18, 1999, I appoint the following  
10 individuals as voting members of the Radiological  
11 Devices Panel for the meeting on February 3, 2004, and  
12 they are as follows:

13 Brent Blumenstein, Ph.D., Thomas Ferguson,  
14 M.D., Elizabeth A. Krupinski, Ph.D., Stephen Solomon,  
15 M.D., and David Stark, M.D.

16 For the record, these individuals are  
17 special government employees and consultants to this  
18 panel under the Medical Devices Advisory Committee.  
19 They have undergone the customary conflict of interest  
20 review and have reviewed the material to be considered  
21 at this meeting. Again, signed by David W. Feigal for  
22 the Center of Devices and Radiological Health.

**NEAL R. GROSS**

COURT REPORTERS AND TRANSCRIBERS

1323 RHODE ISLAND AVE., N.W.

WASHINGTON, D.C. 20005-3701

(202) 234-4433

[www.nealrgross.com](http://www.nealrgross.com)

1                   Finally, the conflict of interest  
2 statement. The following announcement addresses  
3 conflict of interest issues associated with this  
4 meeting and is made part of the record to preclude  
5 even the appearance of impropriety.

6                   To determine if any conflict existed, the  
7 agency reviewed a submitted agenda for the meeting and  
8 all financial interest reported by the committee  
9 participants. The agency has no conflicts to report.

10                   In the event that the discussions involved  
11 in any other products or firms not already on the  
12 agenda for which an FDA participant has financial  
13 interest, the participants should excuse him or  
14 herself from such involvement and the exclusion will  
15 be noted for the record.

16                   With respect to all other participants we  
17 ask in the interest of fairness that all persons  
18 making statements or presentations disclose any  
19 current or previous financial involvement with any  
20 firm whose products they may wish to comment upon.

21                   Now, if there is anyone who has anything  
22 to discuss concerning these matters which I have just

**NEAL R. GROSS**

COURT REPORTERS AND TRANSCRIBERS

1323 RHODE ISLAND AVE., N.W.

WASHINGTON, D.C. 20005-3701

(202) 234-4433

[www.nealrgross.com](http://www.nealrgross.com)

1 mentioned, please advise me now and we can leave the  
2 room to discuss them. Seeing none, the FDA seeks  
3 communications with industry and the clinical  
4 community in a number of different ways,

5 First, the FDA welcomes and encourages  
6 pre-meetings with sponsors prior to all IDE and PMA  
7 submissions. This affords the sponsor an opportunity  
8 to discuss issues that could impact the review  
9 process. Second, the FDA communicates through the use  
10 of guidance documents. Toward this end, the FDA  
11 develops two types of guidance documents for  
12 manufacturers to follow when submitting a premarket  
13 application.

14 One type is simply a summary of the  
15 information that has historically been requested on  
16 devices that are well understood in order to determine  
17 substantial equivalence.

18 The second type of guidance document is  
19 one that develops as we learn about new technology.  
20 FDA welcomes and encourages the panel and industry to  
21 provide comments concerning our guidance documents. I  
22 would also like to remind you that the meetings of the

**NEAL R. GROSS**

COURT REPORTERS AND TRANSCRIBERS

1323 RHODE ISLAND AVE., N.W.

WASHINGTON, D.C. 20005-3701

(202) 234-4433

[www.nealrgross.com](http://www.nealrgross.com)

1 Radiological Devices Panel for the remainder of this  
2 year are tentatively scheduled for May 18th, August  
3 10th, and November 16th.

4 You may wish to pencil these dates in on  
5 your calendar but please recognize that these dates  
6 are tentative at this time. I'll repeat them in case  
7 you didn't get those. May 18th, August 10th, and  
8 November 16th.

9 DR. IBBOTT: Thank you, Mr. Doyle.

10 At this point Nancy Brogdon, who is  
11 Director of the Division of Reproductive, Abdominal,  
12 and Radiological Devices of the Office of Device  
13 Evaluation has a few words she would like to say.

14 MS. BROGDON: Thank you, Dr. Ibbott. We  
15 have three panel members whose terms just expired on  
16 January 31st. They are not present today but we  
17 wanted to recognize publicly their contributions to  
18 the panel.

19 The first is Mr. Ernest Stern. Mr. Stern  
20 was the Chairman and CEO of Thales Components located  
21 in Totowa, New Jersey, and he was the industry rep on  
22 the panel for the past four years. He is now retired

**NEAL R. GROSS**

COURT REPORTERS AND TRANSCRIBERS

1323 RHODE ISLAND AVE., N.W.

WASHINGTON, D.C. 20005-3701

1 from Thales.

2 Mr. Stern effectively represented various  
3 industries served by this panel and used his position  
4 on the panel to apprise other panel members of  
5 commercial considerations that they should take into  
6 account when making recommendations on the various  
7 applications under review.

8 Second is Dr. Wendy Berg. Dr. Berg was  
9 the Director of Breast Imaging in the Department of  
10 Radiology at University of Maryland at Baltimore. She  
11 served on the panel for four years as a voting member.

12 Dr. Berg brought to the panel a high degree of  
13 expertise in the field of mammography.

14 That was continually called upon as novel  
15 mammography related devices were reviewed by the  
16 panel. In addition, when asked, she provided written  
17 reviews of complex devices applications that the  
18 agency used as part of our in-house review process.

19 Third is Dr. Harry Genant. Dr. Genant is  
20 Professor of Medicine and Epidemiology, Orthopedics,  
21 and Surgery at the University of California at San  
22 Francisco. He also served as a voting member for four

**NEAL R. GROSS**

COURT REPORTERS AND TRANSCRIBERS

1323 RHODE ISLAND AVE., N.W.

WASHINGTON, D.C. 20005-3701

1 years. Dr. Genant brought to the panel a brought  
2 spectrum of expertise with special emphasis on bone  
3 densitometry. His probing questions and insightful  
4 comments on the pros and cons of the devices being  
5 considered were very helpful to the agency as it  
6 reviewed the safety and effectiveness of new devices.

7 We thank all of these past panel members.  
8 each will be sent a thank-you from the commissioner  
9 along with a mounted service plaque. Thank you.

10 DR. IBBOTT: Thank you.

11 Dr. Robert Phillips, the Chief of the  
12 Radiology Branch of the Office of Device Evaluation  
13 will now give a brief update on the FDA radiology  
14 activities. Dr. Phillips.

15 DR. PHILLIPS: Well, good morning again.  
16 As you can see by the absence of meetings between  
17 December '02 and now, we have not had a whole bunch of  
18 brand new PMAs that we've brought to the panel. In  
19 fact, in the last year we have not approved any PMAs.

20 However, there have been some changes in  
21 the branch itself and we have brought four new people  
22 on board as reviewers. These are Nancy Wersto who

**NEAL R. GROSS**

COURT REPORTERS AND TRANSCRIBERS

1323 RHODE ISLAND AVE., N.W.

WASHINGTON, D.C. 20005-3701

(202) 234-4433

[www.nealrgross.com](http://www.nealrgross.com)

1 comes to us from industry. She's a radiological  
2 physicist and her interest area is in radiation  
3 therapy products.

4 Then we have Kish Chakrabarti who comes to  
5 us from the mammography side of the center. He is a  
6 physicist. His area of interest is mammography and  
7 imaging systems. Kish, are you here today? No.

8 Dr. Barbara Shawback comes to us from  
9 outside. She's a medical officer and her area is  
10 study and design in rheumatology.

11 And then we just had a new employee come  
12 on board, Sophie Packerel. She is a physicist who  
13 comes from the University of Chicago and her area is  
14 CAD systems.

15 Those are the four people that have come  
16 on board and ends my talk. Thank you.

17 DR. IBBOTT: Thank you. We'll now proceed  
18 with the first of two half-hour open public hearing  
19 sessions for this meeting. The second half hour open  
20 public hearing session will follow the panel  
21 discussion this afternoon.

22 Both the Food and Drug Administration and

**NEAL R. GROSS**

COURT REPORTERS AND TRANSCRIBERS

1323 RHODE ISLAND AVE., N.W.

WASHINGTON, D.C. 20005-3701

1 the public believe in a transparent process for  
2 information gathering and decision making. To ensure  
3 such transparency at the open public hearing session  
4 of the advisory committee meeting, FDA believes that  
5 it is important to understand the context of an  
6 individual's presentation.

7 For this reason, FDA encourages you, the  
8 open public hearing speaker, at the beginning of your  
9 written or oral statement to advise the committee of  
10 any financial relationship that you may have with the  
11 sponsor, its product and, if known, its direct  
12 competitors.

13 For example, this financial information  
14 may include the sponsor's payment of your travel,  
15 lodging, or other expenses in connection with your  
16 attendance at the meeting. Likewise, FDA encourages  
17 you at the beginning of your statement to advise the  
18 committee if you do not have any such financial  
19 relationships. If you choose not to address this  
20 issue of financial relationships at the beginning of  
21 your statement, it will not preclude you from  
22 speaking.

**NEAL R. GROSS**

COURT REPORTERS AND TRANSCRIBERS

1323 RHODE ISLAND AVE., N.W.

WASHINGTON, D.C. 20005-3701

1           No individual has given advance notice of  
2 wishing to address the panel. If there is anyone now  
3 wishing to address the panel, please identify  
4 yourselves at this time.

5           Seeing none, I would like to remind public  
6 observers at this meeting that while this portion of  
7 the meeting is open to public observation, public  
8 attendees may not participate except at the specific  
9 request of the chair.

10           We can now begin the first open public  
11 portion of the meeting. We will now, as I said,  
12 proceed with the open committee discussion portion of  
13 this meeting that has been called for the  
14 consideration of PMA 030012 for a computer-aided  
15 detection, CAD device, that assist a physician in  
16 identifying actionable, solid nodules in CT images of  
17 the lung.

18           The first presentation will be by Dr.  
19 Robert F. Wagner of the FDA who will give an overview  
20 of contemporary ROC methods such as may be used in  
21 measuring the effectiveness of the CAD and other  
22 imaging devices.

**NEAL R. GROSS**

COURT REPORTERS AND TRANSCRIBERS

1323 RHODE ISLAND AVE., N.W.

WASHINGTON, D.C. 20005-3701

1           The sponsor, R2 Technology, Inc., will  
2 then state its case for the PMA and they will be  
3 followed by the FDA with its review of the device. We  
4 will proceed now with Dr. Wagner's presentation.

5           DR. WAGNER: Cybersource as I am, let us  
6 see if I can -- okay. Progress or regress? Let's not  
7 start from the back. Marvelous.

8           Thank you very much, Bob. I'm glad we  
9 planned this together this way. Good morning to the  
10 members of the panel, my colleagues and visitors  
11 today. I must acknowledge the fact that Dr. Bill  
12 Sacks and I were awakened by our respective wives at  
13 our respective homes every two hours this morning to  
14 see what the weather would be like to see if we would  
15 be able to make it and what time we should really get  
16 up. We are working against that as our background.

17           I would also like to thank my colleagues  
18 for giving me this opportunity to present this  
19 tutorial information on an overview of the  
20 contemporary ROC methodology as it is used today in  
21 the field of medical imaging and computer assisted  
22 devices.

**NEAL R. GROSS**

COURT REPORTERS AND TRANSCRIBERS

1323 RHODE ISLAND AVE., N.W.

WASHINGTON, D.C. 20005-3701

1           Of course, most of us know what the  
2 letters stand for. ROC stands for receiver operating  
3 characteristic. This is the historic name that comes  
4 down to us from the field of radar in signal detection  
5 studies where the problem is you're looking at a field  
6 of clutter and the question is is there an airplane in  
7 that clutter.

8           In the field of psychology and this  
9 perception in eye and brain coordination studies, this  
10 subject is often called the relative operating  
11 characteristic. Some people are just weary of the R  
12 and just refer to this as the operating characteristic  
13 because that's really what it is.

14           Those of us in the field of medical  
15 imaging have retained the name of receiver operating  
16 characteristic. I think it is because of our devotion  
17 to the classic literature from about 30 years or so  
18 ago that we have just retained, the conservative  
19 people that we are. I see a person who has worked in  
20 this field looking back at us.

21           Well, now here is an outline of the talk.  
22           We will spend a few minutes talking about efforts

**NEAL R. GROSS**

COURT REPORTERS AND TRANSCRIBERS

1323 RHODE ISLAND AVE., N.W.

WASHINGTON, D.C. 20005-3701

1 toward consensus development on the present issues.  
2 Then we'll move right into the ROC paradigm. We'll  
3 talk about how it gets complicated by the problem of  
4 reader variability. How the multiple reader multiple  
5 case, or so-called MRMC ROC paradigm, arose to address  
6 this problem of reader variability.

7           Since the ROC is a measurement, you have  
8 to have a meter stick of some kind so we'll talk about  
9 measurement scales. There will be a categorical  
10 scale, patient management or action scale and a  
11 probability scale that we'll talk about.

12           Then for today's submission, and  
13 submissions like it, there are additional  
14 complications from the problem of location  
15 uncertainty, from the problem of not really knowing  
16 the truth and dealing with uncertainty in the truth.  
17 Since the truth is uncertain, you really don't know  
18 how many effective number of samples you really have.

19           When you have a system that's going to cue  
20 readers about the possibility of lesions on a case,  
21 there is a problem of reader vigilance that we will  
22 discuss. Finally, we'll give a little wrap-up which I

**NEAL R. GROSS**

COURT REPORTERS AND TRANSCRIBERS

1323 RHODE ISLAND AVE., N.W.

WASHINGTON, D.C. 20005-3701

1 won't have to give because Bob Phillips just presented  
2 it for me.

3 Let's start off now with efforts toward  
4 consensus development on the present issues. The fact  
5 is that at the moment we do not have an explicit FDA  
6 guidance on how to review, how to submit and review  
7 issues like the present one. There's been a lot of  
8 work going on and deep background as to how did we get  
9 here.

10 The basic idea is how do you use the  
11 classic concepts of sensitivity, specificity, and ROC  
12 analysis to assess performance of diagnostic imaging  
13 and computer-assisted systems. Especially since there  
14 are many new issues and levels of complexity that come  
15 to the fore as more complex technologies emerge.

16 At the moment you see there is really no  
17 software to do the assessment task of the problem we  
18 have before us. That's why I would like to talk about  
19 piecemeal, all the different pieces and what is known  
20 and what does exist at the moment because the sponsor  
21 had to put together a creative combination of these  
22 many things. So continuing on this little laundry

**NEAL R. GROSS**

COURT REPORTERS AND TRANSCRIBERS

1323 RHODE ISLAND AVE., N.W.

WASHINGTON, D.C. 20005-3701

1 list. I'll give you an historical laundry list of  
2 efforts toward consensus development on these present  
3 issues.

4 That's RSNA. Most of you recognize that.  
5 That's the big Radiological Society of North America  
6 meeting that's held every year in November in Chicago  
7 that makes this weather look very mild today. Then  
8 following RSNA by a few months is the big SPIE medical  
9 imaging meeting. At the SPIE meetings we generally  
10 handle the more technical aspects of the issues that  
11 come up at the RSNA.

12 Then there's a society that meets every  
13 two years called the Medical Image Perception Society  
14 of which Elizabeth Krupinski on our panel has been  
15 president for 40 years I think it has been. Elizabeth  
16 is the President of the Medical Image Perception  
17 Society. We hold various workshops and literature  
18 every two years.

19 In all these meetings every few years we  
20 do note progress in this field. There is tremendous  
21 progress going on but it's without a doubt still an  
22 evolving work in progress. We are still not at the

**NEAL R. GROSS**

COURT REPORTERS AND TRANSCRIBERS

1323 RHODE ISLAND AVE., N.W.

WASHINGTON, D.C. 20005-3701

1 holy grail point that we would like to be at but a lot  
2 of progress has indeed been made.

3 At the good old FDA at our center in CDRH  
4 here at the FDA. One of the methods that I'll be  
5 talking about today is the so-called multiple reader  
6 multiple case, the MRMC scheme which has already been  
7 used for several submissions.

8 It was used to break the log jam that was  
9 holding back digital mammography from the market place  
10 so the MRMC scheme that I'll talk about in a few  
11 minutes was used there. It has been used for all  
12 successful submissions of digital mammography PMAs to  
13 our center.

14 This method that we'll talk about in a few  
15 moments has also been used for a successful submission  
16 in the area of a computer aid for lung nodule  
17 detection on chest x-ray film that is in some way  
18 analogous to the present submission but it's just on  
19 plain film.

20 NCI, National Cancer Institute, also has  
21 lung image database consortium and workshops. This is  
22 an NCI funded group of five universities and the

**NEAL R. GROSS**

COURT REPORTERS AND TRANSCRIBERS

1323 RHODE ISLAND AVE., N.W.

WASHINGTON, D.C. 20005-3701

1 principle director of that project, I though I saw him  
2 come in a moment ago. There he is, Larry Clarke.

3 There are five universities that work as  
4 part of this consortium and they are seeking consensus  
5 on a number of things, one of which is how to put  
6 together a database of annotated films of the kind  
7 that you would use, annotated CT slice images of the  
8 kind you would use to train and test a classifier in  
9 this field of computer-aided detection and diagnosis  
10 in lung cancer screening for nodules.

11 So that project is about half-way through  
12 its five-year history. A good two years underway  
13 right now. They are also addressing consensus on the  
14 many issues that you have to deal with when you want  
15 to deal with such a product.

16 For example, how do you keep score  
17 statistically? Once you know how to keep score, then  
18 you can start to design the size of a database. How  
19 do you outline the nodules? How do you keep score  
20 when there's a hit when there is just finite overlap  
21 between what is known of the lesion and what the  
22 reader marks? We'll talk about this in a few moments.

**NEAL R. GROSS**

COURT REPORTERS AND TRANSCRIBERS

1323 RHODE ISLAND AVE., N.W.

WASHINGTON, D.C. 20005-3701

(202) 234-4433

[www.nealrgross.com](http://www.nealrgross.com)

1                   Now, two of here in our center have been  
2 quite active members of this LIDC from the beginning.

3           Let me see if I have another comment here.   Yeah.  
4           The thing I would like to bring to your attention this  
5 morning is that there has been a great amount of  
6 communication among all these resources here.   A  
7 number of us in our center here are active members of  
8 the research community in this field.

9                   Many of us here and sitting just behind me  
10 have been very active in this area of applying these  
11 methods to several of the submissions in the area of  
12 imaging a computer-aided diagnosis.   Several of us are  
13 very active members, Larry Clarke's group here.

14                   What we have tried to do is see this as  
15 several quarters, four quarters if you will, if a  
16 quadrangle all holding the windows open to the others  
17 so the people who come in to us from industry at any  
18 given moment will know what is the state of the art  
19 from the academia, from our own center, and from the  
20 LIDC.

21                   We presented them all the papers, all the  
22 current drafts even, and made sure that everyone knows

**NEAL R. GROSS**

COURT REPORTERS AND TRANSCRIBERS

1323 RHODE ISLAND AVE., N.W.

WASHINGTON, D.C. 20005-3701

1 what's on the other people's mind methodology wise  
2 that is outside the area of anything that is  
3 proprietary. Anything that is not proprietary is all  
4 strictly methodology or statistics. We have tried to  
5 keep these communication channels as open as we could.

6 Here we go with the promised little  
7 tutorial and the fundamentals of the ROC paradigm  
8 itself. The idea is, of course, that you have two  
9 populations, one a population of actually diseased  
10 people. You might think of these as people with  
11 diabetes, for example, and a population of people who  
12 do not have the disease.

13 You would like to have a test that puts  
14 out a result something like a volt meter or a  
15 biochemical assay or, in the case of a simple blood  
16 sugar test, this would just be the blood sugar  
17 concentration. You would love to have the world such  
18 that the two populations would be separated and you  
19 could just drop a threshold in here and say these  
20 patients are diseased and these patients can go home  
21 and not worry about it.

22 Now, in the field of medical imaging those

**NEAL R. GROSS**

COURT REPORTERS AND TRANSCRIBERS

1323 RHODE ISLAND AVE., N.W.

WASHINGTON, D.C. 20005-3701

1 of us who have done work in that field you don't have  
2 a simple meter or biochemical assay. What you get is  
3 a reader looking at about a million pixels of a  
4 picture and trying to get the features out of it and  
5 reduce that through what we call the subjective  
6 likelihood, subjective judgment or likelihood that  
7 case is diseased.

8 Now, as I say, this is really not quite  
9 the way the diabetes blood sugar test works but if you  
10 think of what I am about to tell you in that context  
11 for the next few minutes, you won't be far off base.  
12 It's not precise but it wouldn't be misleading.

13 So here is what happens more typically.  
14 The two populations are not separated. The diseased  
15 population and the nondiseased population as far as  
16 their test result is concerned have a very great  
17 overlap. The idea is now who do you send home and who  
18 do you send on for further workup or people that you  
19 want to treat for a condition.

20 Those of you who have seen this before,  
21 what I've just done I've taken these two and dropped  
22 this population down so that you won't get mixed up

**NEAL R. GROSS**

COURT REPORTERS AND TRANSCRIBERS

1323 RHODE ISLAND AVE., N.W.

WASHINGTON, D.C. 20005-3701

1 with the colors. Now we have the nondiseased cases  
2 and the diseased cases on the same axis, the same  
3 relative position. Now in a practical situation with  
4 the overlap, now we have to set ourselves a threshold.

5 If this is a blood sugar test, for  
6 example, you could set it at 150 blood sugar level.  
7 If you do that, you'll pick up about half of the  
8 actual diabetic patients so we say we have a true  
9 positive fraction of 50 percent but you have to pay  
10 for this price. You have about a 10 percent false  
11 positive fraction so here is this point, 50 percent  
12 true positive and roughly 10 percent false positive.

13 We call this a less aggressive mind set  
14 and I think you'll see the reason for that in just a  
15 moment. So if we get a little bit more aggressive to  
16 try to pick up more patients in our sieve, we might  
17 set the threshold down here at 100 instead of 150.  
18 Now we get about 80 percent of the diabetic patients  
19 and now at the price of about 20 percent false  
20 positive or 25 percent. Here I've put this point  
21 about 80 percent and 25 percent.

22 Let's get even more aggressive and what I

**NEAL R. GROSS**

COURT REPORTERS AND TRANSCRIBERS

1323 RHODE ISLAND AVE., N.W.

WASHINGTON, D.C. 20005-3701

1 mean by that is I want to pick up more diseased  
2 patients in my sieve, the sieve being the test. If  
3 you set the threshold in the 90's, now we might get  
4 almost 95 percent of the patients in our sieve of the  
5 actual diabetic patients but then we have to pay the  
6 price of 50 percent of the nondiabetic patients picked  
7 up so now we have a 90 percent sensitivity and roughly  
8 a 50 percent sensitivity.

9 Now, you can take this to the extreme and  
10 we talk about this particular test all the time and I  
11 think this might not work because the threshold now --  
12 oh, it did work. Okay. We can put the threshold all  
13 the way to the left and call everybody to the right of  
14 this diseased and we would get all the diabetic  
15 patients. There's a little mark right up here. We  
16 would get also -- the price we would pay is we would  
17 have to call everybody who is not a diabetic a  
18 diseased patient here so we would generate that point.

19 I think you can see and let your  
20 imagination go wild that you can certainly fill in all  
21 these points. Don't blink, anyone. I saw Dr. Bob  
22 Doyle blink there so I have to go back and do that

**NEAL R. GROSS**

COURT REPORTERS AND TRANSCRIBERS

1323 RHODE ISLAND AVE., N.W.

WASHINGTON, D.C. 20005-3701

(202) 234-4433

[www.nealrgross.com](http://www.nealrgross.com)

1 again. Instead of working up more and more levels of  
2 aggressiveness, you could back off. You could start  
3 off with everybody at the sick point and then just  
4 back off, move the threshold the other way and fill in  
5 the complete ROC curve. You can see at this time of  
6 day I'm very easily amused.

7 Okay. Here is the overall picture now.  
8 This is the case of the schematic of, let us say,  
9 blood sugar as a test for diabetes. These are these  
10 two populations and the way they overlap and here is  
11 the corresponding ROC curve with the level of  
12 aggressiveness increasing.

13 Now, it can happen and, in fact, we've  
14 seen things like this in our center and you see this  
15 in the laboratory once in a while, the two populations  
16 could fall right on top of one another so that a test  
17 cannot actually discriminate between the two  
18 conditions so what we've done here is just drop this  
19 population and this population on top of each other.  
20 Now if you generate an ROC curve the way I just showed  
21 you, you would generate what we call the chance line  
22 or guessing line.

**NEAL R. GROSS**

COURT REPORTERS AND TRANSCRIBERS

1323 RHODE ISLAND AVE., N.W.

WASHINGTON, D.C. 20005-3701

1           Toward the other extreme you could have a  
2 test that separates the two populations very well. In  
3 that case, as we move the threshold across from less  
4 aggressive to more aggressive, we'll generate this ROC  
5 curve. Now we have the guessing line, we have the ROC  
6 curve corresponding to almost typical clinical  
7 laboratory test, and we have the ROC curve here for a  
8 very good test. We call this the level of increasing  
9 -- we call this direction the direction of increasing  
10 reader skill or increasing level of technology.

11           Now, many people like to have a single  
12 summary measure of ROC curve performance and what has  
13 traditionally been used is you take the area under the  
14 curve so the area under this curve, say the diabetic  
15 discrimination test, is something in the high 70s.  
16 Let's call it 78 percent or something like that.

17           If you use the area under the curve as a  
18 summary measure of performance, in effect, remember if  
19 you think of calculus, you're getting this area you're  
20 just integrating, you are effectively replacing the  
21 curve with a line that is fault at the level of that  
22 area.

**NEAL R. GROSS**

COURT REPORTERS AND TRANSCRIBERS

1323 RHODE ISLAND AVE., N.W.

WASHINGTON, D.C. 20005-3701

1                   In effect, what you've done is you have  
2 averaged the sensitivity with a true positive fraction  
3 over all false positive fractions. In effect, if you  
4 use the area of the curve you are given the  
5 sensitivity averaged over all false positive fractions  
6 or sensitivity averaged over all specificity,  
7 specificity coming from the other direction.

8                   Well, I hope it gets interesting now.  
9 That was the easy part. That's the idea. Let's see  
10 what really happens in the real world. In the real  
11 world in the last decade those of us who work in this  
12 field have been made acutely aware of the complication  
13 of reader variability.

14                   I'm going to show you some very famous  
15 data. I think Emily Conant knows this like the back  
16 of her hand from having worked with Craig Beam. For  
17 those of you who have not seen this before, I have to  
18 give a little build up to this.

19                   This is a set of data from Beam, Layde and  
20 Sullivan that I'm going to show you in which they  
21 studied 108 mammographers randomly chosen from around  
22 the United States. The mammographers in this study

**NEAL R. GROSS**

COURT REPORTERS AND TRANSCRIBERS

1323 RHODE ISLAND AVE., N.W.

WASHINGTON, D.C. 20005-3701

1 were given a set of mammograms. They were asked to  
2 set their threshold for action.

3 Remember when we were talking about this  
4 ROC paradigm we were moving a threshold and we wanted  
5 to set it at some place and the question is in a  
6 clinical laboratory test you could just dial that in  
7 somehow. How do you do it in medical imaging? You  
8 don't have a dial.

9 You have to deal with the human reader and  
10 they were asked to set their threshold between their  
11 sense of the boundary on the BIRADS scale, Breast  
12 Imaging and Reporting and -- Reporting or Recording?  
13 Anyway, Reporting and Data System. That's the  
14 American College of Radiology Scale that is used for  
15 managing patients in mammography.

16 These readers were asked to set their  
17 sense of the boundary between category 3, which is  
18 generally six-month follow-up recommendation, and  
19 category 4 which is highly suspicious and recommend  
20 consideration of biopsy. I'm sure I'm garbling that  
21 but you get the general idea. I wasn't asked to leave  
22 the room so I couldn't be too far off there.

**NEAL R. GROSS**

COURT REPORTERS AND TRANSCRIBERS

1323 RHODE ISLAND AVE., N.W.

WASHINGTON, D.C. 20005-3701

1                   Here's what happened.     This is a true  
2 positive fraction versus a false positive fraction for  
3 108 readers.    There are 108 points here.   Each one of  
4 these people thinks that they had set the boundary  
5 between category 3 and category 4.

6                   If you try to do public policy based on  
7 category 3 and category 4 and thinking that people  
8 have optimized that, the optimum is very broad.  
9 People have not figured out how to optimize that.  
10 That's a big problem.

11                   Let's look at this reader.   This is one  
12 out of 108 people.   This person has a sensitivity of  
13 70 percent and a false positive rate of about 25  
14 percent.   Now, this person thinks they are being as  
15 aggressive as they should be in the context but this  
16 person is more aggressive than this one, this reader  
17 is more aggressive than this one, this reader is the  
18 most aggressive on this bottom curve here, and these  
19 readers are less aggressive.

20                   Now, as we go in the other direction, we  
21 now see the variability due to the range of reader  
22 skill.   We can say that these readers have a greater

**NEAL R. GROSS**

COURT REPORTERS AND TRANSCRIBERS

1323 RHODE ISLAND AVE., N.W.

WASHINGTON, D.C. 20005-3701

1 skill at this task than these readers and these  
2 readers have the greatest skill yet.

3 At any level of reader skill we have  
4 different readers thinking that they have optimally  
5 set their threshold. This is a tremendous range of  
6 reader variability. There are 108 mammographers  
7 represented on this graph. This is classic work from  
8 Craig Beam, Peter Layde and Dan Sullivan.

9 What have I just told you? There is no  
10 unique ROC operating point. Each one of these people  
11 is set to be at a certain operating point. There is  
12 no unique ROC operating point. There is not even a  
13 unique ROC curve. There is only a band or region of  
14 ROCs as you can see. There is a very broad band.

15 I hope I've convinced you all now that  
16 this gets to be a more complex issue. In particular,  
17 here is the question. Suppose we have two  
18 technologies that manifest themselves in reader's  
19 hands with this level of variability?

20 How do you compare those two technologies?  
21 That's the issue before us with a whole class of  
22 problems that we've been discussing over the last few

**NEAL R. GROSS**

COURT REPORTERS AND TRANSCRIBERS

1323 RHODE ISLAND AVE., N.W.

WASHINGTON, D.C. 20005-3701

1 years and we'll be seeing more of over the next few  
2 years. How do you do it?

3 This is not an isolated example. People  
4 have gotten used to this and said this is really an  
5 extreme example. This is not the most extreme example  
6 we've ever seen.

7 In our group we have actually looked at  
8 over a dozen real world publicly available data sets  
9 and the example I just showed you is sort of in the  
10 middle. Sometimes things are a little bit better.  
11 Sometimes they are even much worse than what I just  
12 showed you. Sometimes things are a little bit better.

13 Sometimes they are even much worse than what I just  
14 showed you. The following is an example from Dr. Jim  
15 Potchen from plain chest x-ray picking up the disease  
16 on chest films. These are ROC curves. Dr. Potchen  
17 looked at over 100 radiologists and 71 residents. He  
18 averaged the score card ROC wise of his top 20  
19 radiologist. Here they are.

20 Then he presents here the average ROC  
21 curve for his radiology residents. There are 71 of  
22 them here representing this average line. The bottom

**NEAL R. GROSS**

COURT REPORTERS AND TRANSCRIBERS

1323 RHODE ISLAND AVE., N.W.

WASHINGTON, D.C. 20005-3701

1 20 radiologists in the study performed here. The  
2 range that we see here is comparable to what we saw in  
3 the Beam, et al. study for mammography. So this is  
4 the real world.

5 Well, you can imagine that if you wanted  
6 to keep score under that setting you have to use a lot  
7 of readers and a lot of cases. The paradigm that has  
8 emerged to address this is, thus, called, almost  
9 eponymously, I guess, if I could pronounce that word,  
10 the multiple reader multiple case, or MRMC paradigm.

11 There are a lot of designs for this.  
12 There are many ways to do it. Today we will just talk  
13 about something that is called the fully -- oh, I  
14 forgot my prop. We'll talk about the fully-crossed  
15 design. The fully-crossed design is one of many but  
16 it is the most efficient in some way so we will talk  
17 about it.

18 You match cases across modalities and you  
19 match readers across modalities. If I can pull this  
20 off. I'm used to having leaves of paper here. Okay.

21 You have a bunch of patients who have been imaged  
22 with modality A here. The same patients imaged with

**NEAL R. GROSS**

COURT REPORTERS AND TRANSCRIBERS

1323 RHODE ISLAND AVE., N.W.

WASHINGTON, D.C. 20005-3701

1 modality B so we say that the cases are matched across  
2 modalities.

3 If we were working with computer-aided  
4 diagnosis, modality A would be readers reading without  
5 the computer aid and modality B would be readers with  
6 the use of the computer aid. There is a stack of  
7 images here. Same patients.

8 We recruit a panel of radiologists,  
9 something like 15 of you people here. All of you read  
10 every patient case in both modalities. What we have  
11 then is we have the cases matched across modalities  
12 and we have the readers matched across modalities.

13 This design is the most statistical power  
14 for a given number of readers and for a given number  
15 of cases with verified truth. Thus, we say it's the  
16 least demanding of these resources. Around here in  
17 Rockville we speak of this as the least burdensome  
18 paradigm because you probably heard in previous  
19 meetings that the FDA has been commissioned by  
20 Congress to enable sponsors to seek and to find, if  
21 possible, the least burdensome path to the marketplace  
22 through the review process.

**NEAL R. GROSS**

COURT REPORTERS AND TRANSCRIBERS

1323 RHODE ISLAND AVE., N.W.

WASHINGTON, D.C. 20005-3701

1           So what we've done is we've always called  
2 this to the attention of incoming sponsors that this  
3 design is most powerful. You can use alternative  
4 designs and you can come close sometimes to the  
5 efficiency of this scheme but this is the most  
6 powerful in terms of the ground rules I have on the  
7 slide right there.

8           Well, if you are familiar with the  
9 literature in this field, you will say, you know, this  
10 is no modern big deal. This stuff has been known for  
11 a good 20 years or so. If you read the classic book  
12 by Swets and Pickett the whole idea is laid out there.

13          The trouble is there was no practical way to  
14 implement this scheme 20 years ago until people  
15 started to understand what's called the statistical  
16 approach of resampling strategies.

17          I probably shouldn't spend any time on the  
18 past history but the fact of the matter is in past  
19 years before they realized about resampling they just  
20 started to stratify the data and then you give up a  
21 lot of statistical power. In modern times in the last  
22 10 years people realized if you use the statistical

**NEAL R. GROSS**

COURT REPORTERS AND TRANSCRIBERS

1323 RHODE ISLAND AVE., N.W.

WASHINGTON, D.C. 20005-3701

1 resampling, you can use the data over and over again  
2 in a well-pedigreed way and get statistically valid  
3 inputs.

4 So the two most famous resampling schemes  
5 are called the statistical jackknife or the  
6 statistical bootstrap. The big break through came in  
7 this field in 1992. This is the classic so-called DBM  
8 paper. That's Donald Dorfman of happy memory whom we  
9 lost to our community very sadly two years ago. His  
10 colleague, Kevin Berbaum, and the well-known Charles  
11 Metz at the University of Chicago.

12 This paper broke the log jam in this  
13 field. They suggested using the statistical jackknife  
14 in combination with classical ANOVA and the  
15 statistical jackknife just being a leave-one-out  
16 method where you leave Mrs. Jones out one time and you  
17 leave Mrs. Smith out the next time and you generate a  
18 lot of data sets that way, submit it to classical  
19 ANOVA, and you can do your inference about the  
20 difference between these two competing technologies.

21 Well, it turns out this is a little bit  
22 more difficult to explain in any more detail than

**NEAL R. GROSS**

COURT REPORTERS AND TRANSCRIBERS

1323 RHODE ISLAND AVE., N.W.

WASHINGTON, D.C. 20005-3701

1 that. But the bootstrap method is very trivial to  
2 explain in some detail so I'm going to ask you to sit  
3 through that with me for the next minute or so.

4 The idea with the statistical bootstrap is  
5 that we are going to -- the bootstrap itself means you  
6 are going to resample from a set of data points with  
7 replacement. I'll show you that in a moment. We are  
8 going to bootstrap the experiment of interest. We'll  
9 draw random readers, random cases, and then carry out  
10 the experiment of interest many times.

11 Here is an example of some possible  
12 bootstrap samples from a set of -- suppose there are  
13 15 of you here. We might have a set of numbers one  
14 through 15. We start drawing them with replacement.  
15 If you wait long enough, you might get a list that has  
16 one, two, three, four, five, six, seven -- you have to  
17 wait a long time before that happens.

18 In the meantime you get more random  
19 looking samples like this. When I was thinking about  
20 this, you know, if you did this with letters this  
21 reminds you of that proverbial experiment where they  
22 have the monkeys trying to type out the soliloquy of

**NEAL R. GROSS**

COURT REPORTERS AND TRANSCRIBERS

1323 RHODE ISLAND AVE., N.W.

WASHINGTON, D.C. 20005-3701

1 Pollonius or something like that. It's going to  
2 happen but you may have to wait a long time.

3           Instead what you do is you get random  
4 samples like this. The number one never showed up in  
5 this group. The number two showed up once. Number  
6 three showed up a couple times. Number 14 showed up  
7 three times and so on. You randomly sample a number  
8 and then put it back. Write it down. This can go on  
9 for an astronomical number of times.

10           Then another example, the number one shows  
11 up, number 15 shows up and so on. You get a lot of  
12 these, a very great number of these but you don't have  
13 time to do them all so, in practice, people use about  
14 1,000. It depends on the complexity of the problem.

15           So you draw about a 1,000 bootstraps of  
16 readers and cases. The number of cases you draw is  
17 comparable to the experiment you are trying to mock  
18 up. Then what you do is with that bootstrap safe on  
19 the random case sample, you have all the readers in  
20 their bootstrap sample read all the cases in both  
21 modalities in that bootstrap sample, carry out the  
22 experiment of interest so you would get the

**NEAL R. GROSS**

COURT REPORTERS AND TRANSCRIBERS

1323 RHODE ISLAND AVE., N.W.

WASHINGTON, D.C. 20005-3701

(202) 234-4433

[www.nealrgross.com](http://www.nealrgross.com)

1 performance measure.

2 That's called area under the RC curve for  
3 the one. You get that number for the other. You take  
4 the difference. You do that 1,000 times and then you  
5 put them in order from the lowest different to the  
6 highest. Then it's very easy to get the mean and then  
7 you can take out the central 95 percent junk and that  
8 would give you a 95 percent confidence level. That's  
9 a simple way to explain the story.

10 In the jackknife plus ANOVA it's a little  
11 bit more elaborate than that but you can actually  
12 think of the jackknife as the first order of  
13 approximation to the bootstrap. So these two  
14 approaches are sort of in the same spirit but one is  
15 completely nonparametric and the other is -- the  
16 classical ANOVA is heavily based on the multi-variate  
17 normal so it's highly parametric.

18 As I just said, you obtain a mean  
19 performance over readers and cases but it's much more  
20 interesting. The mean is always easy to get no matter  
21 how you approach a problem. Well, it can be tricky.  
22 But the big thing you want is error bars that account

**NEAL R. GROSS**

COURT REPORTERS AND TRANSCRIBERS

1323 RHODE ISLAND AVE., N.W.

WASHINGTON, D.C. 20005-3701

1 for both the variability of readers and cases.

2 You know, in the DBM paper they quoted a  
3 quote that has become very famous from Jim Hanley.  
4 Many of us know Jim Hanley from McGill University in  
5 Montreal.

6 Jim Hanley says, "When you report the  
7 results of your experiment to your readership, it's  
8 not so important just to report the mean performance  
9 or the results you got in the very experiment at hand  
10 because, after all, this experiment will never be done  
11 again. No one will ever do this particular  
12 experiment.

13 What readers want is they want a sense of  
14 the range of performance to be expected if this  
15 experiment could be repeated many, many times drawing  
16 randomly, one hopes, from the same population from  
17 which the current samples were drawn. So that is the  
18 idea.

19 You ought to be able to report to your  
20 readership not just a p-value because we all know it  
21 takes p-value to get a paper published in a medical  
22 journal. You want to actually be able to explain the

**NEAL R. GROSS**

COURT REPORTERS AND TRANSCRIBERS

1323 RHODE ISLAND AVE., N.W.

WASHINGTON, D.C. 20005-3701

1 range of variability you expect to see if this  
2 experiment is done over and over again. That's what  
3 you get when you keep score this way.

4 Okay. We said that the ROC curve is a  
5 measurement. Above all else it is a measurement so  
6 you have to think about a measurement science. You  
7 have to think about the scale you'd be using for  
8 reporting and doing the measurements.

9 Historically -- I should just stop for  
10 moment to tell those of you who were not around in the  
11 late '70s and early '80s that the National Cancer  
12 Institute gave a contract to people in Cambridge,  
13 Massachusetts, Bolt, Beranek and Newman, where John  
14 Swets, David Getty, and Ronald Pickett and colleagues  
15 were working to develop a protocol for how to do ROC  
16 experiments and how to keep score and how to do the  
17 data analysis.

18 That is published in a paper in science  
19 1979. The book came out in 1982 and many of us have  
20 that book on our shelf. The protocol used at that  
21 time was so-called historic ordered category scales.  
22 There was no does this patient go to biopsy or not.

**NEAL R. GROSS**

COURT REPORTERS AND TRANSCRIBERS

1323 RHODE ISLAND AVE., N.W.

WASHINGTON, D.C. 20005-3701

1 You just looked at the case and you said this patient  
2 -- you use five or six categories.

3 One patient you might say this patient  
4 almost definitely does not have disease. There are  
5 several intermediate levels. The patient probably  
6 does not have disease, might have disease, probably  
7 does have disease, or almost definitely has the  
8 disease. That scheme of five or six categories was  
9 almost exclusively used and there was software for  
10 analyzing that for 25 years.

11 I'm being a little defensive because  
12 people may say why do people use that. That was  
13 approved by -- the experts in the field put it out and  
14 it was supported by NCI. There was a lot of science  
15 underneath it and today people say, "Why did people do  
16 that?" Well, that's what they had.

17 In the last 10 years in the field of  
18 mammography we have this BIRADS scale which is what we  
19 call an action item or a patient management oriented  
20 scale. In that idea you don't categorize the data.  
21 People think of the BIRADS scheme as a categorization  
22 scheme. Let's just put that to the side for a moment.

**NEAL R. GROSS**

COURT REPORTERS AND TRANSCRIBERS

1323 RHODE ISLAND AVE., N.W.

WASHINGTON, D.C. 20005-3701

1                   We'll just think of using the BIRADS scale  
2 to dichotomize patients. We'll say these patients  
3 will not be followed up at all versus these patients  
4 who will get a six-month follow-up. That's one way to  
5 dichotomize the data.

6                   Another way to dichotomize the data is to  
7 say we will try to make the break as we did with the  
8 Beam, et al. data. We'll make the cut in this  
9 dichotomization between those patients who would get  
10 six-month follow-up versus those who we think should  
11 be biopsied right now. So this is a patient  
12 management scheme. This is just a dichotomization  
13 scheme.

14                   About 10 years ago people realized for  
15 very technical reasons that it would be useful to use  
16 what they called the continuous probability rating  
17 scale, or quasi-continuous. It's a hundred-point  
18 scale, one, two, three, four, five, but you wouldn't  
19 get 1.5 for example so they call it quasi-continuous,  
20 hundred-point scale.

21                   Nobody expects anybody literally to use  
22 probability 13 or probability 17 or anything, but the

**NEAL R. GROSS**

COURT REPORTERS AND TRANSCRIBERS

1323 RHODE ISLAND AVE., N.W.

WASHINGTON, D.C. 20005-3701

1 idea is to scale your probability or your sense of the  
2 likelihood of disease along a probability scale. That  
3 seems natural to use something if it's a probability  
4 on a scale from zero to 100.

5 So this is the most popular scheme that's  
6 been used to generate ROC data in the last five or  
7 seven years or so. This felt strange to many people,  
8 especially people who are used to using the  
9 categorical scale. But I've talked to a lot of people  
10 about this and very few people outside of the  
11 mammographers have read the BIRADS document.

12 If you go through the BIRADS document and  
13 you go to category four, which is suspicious and  
14 recommend for biopsy, it actually tells you there that  
15 the radiologist should tell the referring physician  
16 their sense of the probability of cancer. There is  
17 actually a culture already existing in which you can  
18 use this kind of patient management action items like  
19 a BIRADS three, four, five, and at the same time give  
20 a continuous probability of disease rating.

21 I see some puzzled looks. I'm trying to  
22 figure out just what I should comment on next. So to

**NEAL R. GROSS**

COURT REPORTERS AND TRANSCRIBERS

1323 RHODE ISLAND AVE., N.W.

WASHINGTON, D.C. 20005-3701

1 make a long story short then, this continuous  
2 probability rating scale has been used for most ROC  
3 curves generated in this community for the last eight  
4 or so years. In the breast imaging --

5 Oh, I remember what I was going to say.  
6 That's why I'm stalling here. In the breast imaging  
7 community many people, it may not be more than half,  
8 but people do use this BIRADS scale. But it's really  
9 important to realize that this BIRADS scale was not  
10 generated -- was not designed to generate ROC curves.

11 People who have tried to use a five-  
12 category scale in this scheme and the BIRADS scale at  
13 the same time have met with a lot of confusion. It  
14 does not work out very well and I see somebody who may  
15 have witnessed people having that experience.

16 Well, I gave a lot of background here  
17 because I would like people to understand that this is  
18 a real issue for the community you would really like  
19 to have because every clinician says, "I want to know  
20 the patient management and I want to know the score  
21 card of the patient management." Every clinician you  
22 talk to, that's what they want.

**NEAL R. GROSS**

COURT REPORTERS AND TRANSCRIBERS

1323 RHODE ISLAND AVE., N.W.

WASHINGTON, D.C. 20005-3701

1           Everybody who measures ROC curves says, "I  
2 want to measure it as finely as I can. I want to use  
3 this quasi-continuous reporting scale." The best of  
4 both worlds would be to get both the quasi-continuous  
5 rating to get the ROC curve and the patient management  
6 action item to get a single sensitivity specificity  
7 point.

8           I'll get a little dramatic for a moment  
9 here. I've talked to many friends. I'm very familiar  
10 with the literature. I could find one example in all  
11 the literature at the moment that's in print where  
12 both of these were done. I could only find one  
13 example of where the best of both worlds was done.

14           This is a paper on classification, what  
15 Bill Sacks and others called CADx using a computer not  
16 to detect but to classify lesions on a film that are  
17 already known. I know that I have a stack of films  
18 here that have microcalcification clusters on them.

19           My task is just to say which ones are  
20 benign and which ones are malignant. That's the task.

21           But I'm going to keep score ROC wise and I'm also  
22 going to keep score patient management wise. I'll

**NEAL R. GROSS**

COURT REPORTERS AND TRANSCRIBERS

1323 RHODE ISLAND AVE., N.W.

WASHINGTON, D.C. 20005-3701

1 show you what they got in a moment.

2                   These authors -- Yulei Jiang, I guess, was  
3 expected here today from a group in Chicago under  
4 Kunio Doi. They studies this test and they had 10  
5 readers and they studied the complete ROC curves.  
6 They studied all the summary measures and they also  
7 studied the patient management or the action item,  
8 sensitivity specificity point.

9                   Here are the results. Here is the average  
10 of 10 ROC curves for 10 readers trying to make this  
11 dichotomy, trying to make this distinction between  
12 benign and malignant lesions. Here is the ROC curve  
13 in the unaided by computer condition. This curve was  
14 generated using the hundred-point probability scale.

15                   This is the curve in the computer-aided  
16 condition, again generated by the hundred-point  
17 probability scale. This point is the mean sensitivity  
18 specificity point generated just by making the  
19 threshold, dichotomizing the data. These patients  
20 benign, these patients malignant. This is a single  
21 dichotomy patient action point in the unaided  
22 condition.

**NEAL R. GROSS**

COURT REPORTERS AND TRANSCRIBERS

1323 RHODE ISLAND AVE., N.W.

WASHINGTON, D.C. 20005-3701

1           That's the same point in the aided  
2 condition. You would love these points to fall on top  
3 of the curves and, for all statistical purposes, they  
4 do because remember the mean -- I have to remind you  
5 of this famous joke that we use around here. There  
6 was a six-foot statistician. You know what happened  
7 to this fellow, right? He drowned while wading in a  
8 stream that had an average height of five feet. You  
9 have to know about the variability.

10           This is not about means, okay? This curve  
11 moves all over the place and this curve moves all over  
12 the place in practice. This is the average of 10.  
13 Same thing. This point moves all over the place as  
14 does this. For all practical purposes this is a great  
15 experiment. This point falls on that curve.

16           Well, it's the only case I could find in  
17 the literature. How come you don't see more of this?

18           When you live with these people that I live with,  
19 it's a great crowd of people and the clinicians say,  
20 "I want the action point." I say, "The committee  
21 wants to measure the ROC curve." Everybody says,  
22 "Let's do both." We are trying to come to that

**NEAL R. GROSS**

COURT REPORTERS AND TRANSCRIBERS

1323 RHODE ISLAND AVE., N.W.

WASHINGTON, D.C. 20005-3701

(202) 234-4433

[www.nealrgross.com](http://www.nealrgross.com)

1 position. Why don't we see more of it?

2 Well, the area under the ROC curve,  
3 remember, you have your ROC curve and you've got the  
4 area under it. You are essentially getting the  
5 sensitivity averaged over all specificities. Right?  
6 You're averaging. You're going to average away a lot  
7 of noise.

8 The variation -- the variance of the area  
9 under the ROC curve -- oh, my goodness. The most  
10 important number of my entire talk is missing. The  
11 variance of the area under the ROC curve is the  
12 binomial variance over two. There's a two here, a  
13 very important two. Those of you who know me know I'm  
14 an expert in factors of two. It's the binomial  
15 variance over two.

16 What's the binomial variance? Well, I  
17 thought if you had a group as we have here today,  
18 about a third of you -- maybe 40 percent of you as I  
19 look around -- know what the binomial variance is.  
20 Suppose we had this meeting next week and we drew from  
21 the same population from which you all came.

22 The next time we did it we might get 32

**NEAL R. GROSS**

COURT REPORTERS AND TRANSCRIBERS

1323 RHODE ISLAND AVE., N.W.

WASHINGTON, D.C. 20005-3701

(202) 234-4433

[www.nealrgross.com](http://www.nealrgross.com)

1 percent of you might know what the binomial variance  
2 is. If we do it three weeks from now and joint  
3 another group in, maybe 49 percent or 52 percent of  
4 you will know what the binomial variance is.

5 What we've just done is what Bill Sacks  
6 refers to. We just made a self-referential example  
7 here. The binomial variance is the variance I would  
8 experience if I did the experiment I just discussed  
9 with you. The area under the ROC curve experiences  
10 only half of that variance.

11 If I studied sensitivity by itself and was  
12 able to tell you ahead of time what the specificity  
13 was so you didn't have to estimate the specificity,  
14 the variance of sensitivity is the entire binomial  
15 variance.

16 In the real world you have to estimate  
17 both the specificity and the sensitivity so the  
18 uncertainty in the specificity propagates into that  
19 and the sensitivity so the variance for that. So if  
20 you wanted to estimate the uncertainty in that action  
21 item that I showed, that point, the circle or the  
22 triangle in the previous data, if you were to estimate

**NEAL R. GROSS**

COURT REPORTERS AND TRANSCRIBERS

1323 RHODE ISLAND AVE., N.W.

WASHINGTON, D.C. 20005-3701

1 that, you would have to live with an uncertainty that  
2 was greater than the binomial variance.

3 If you use area under the RC curve you get  
4 a great reduction. You get the binomial variance over  
5 that famous factor two. This is all approximate but  
6 it works out very well with very practical examples.

7 So what we say is that the variance of the  
8 ROC area is the least burdensome approach to putting  
9 quantification into this problem. I remind you that  
10 is something that we are supposed to enable sponsors  
11 to appreciate.

12 Another thing that we realize in many  
13 discussions with academics and within our house and  
14 with the sponsors and so on is if you want to live in  
15 both of these worlds, that requires consistent  
16 conventions. If you want to be able to either get  
17 categorical reporting and the BIRADS reporting, that's  
18 a lot of work to try to get people to be consistent  
19 that way. People have dropped the categorical scheme  
20 for all practical purposes.

21 Even if you want people to be consistent  
22 between BIRADS and the quasi-continuous scale, that's

**NEAL R. GROSS**

COURT REPORTERS AND TRANSCRIBERS

1323 RHODE ISLAND AVE., N.W.

WASHINGTON, D.C. 20005-3701

1 difficult. We've seen a lot of data in our own group  
2 and from some of the universities. When you train  
3 people, this can be done but not everybody is  
4 trainable right away to be able to do this so it's an  
5 issue. To get data in both worlds then, it's going to  
6 require some convention development.

7 My final point here says this may require  
8 consensus bodies to promote the practice. We would  
9 hope that the American College of Radiology, some of  
10 them other professional societies, and even the fact  
11 that this is of interest to NCI and the FDA, we would  
12 hope that some this would encourage people to try to  
13 do measurements so that we could get both the point  
14 and the curve. Then I think everybody would be happy.

15 Well, this brings us to a little interim  
16 here. Some of you are very familiar with the next few  
17 slides. These are what we call the most famous slides  
18 in the RC archives. Those of you who know Charles  
19 Metz have seen this many times and his followers will  
20 use these many times. Charles died using these slides  
21 over 25 years ago.

22 Here's the classic question. You have two

**NEAL R. GROSS**

COURT REPORTERS AND TRANSCRIBERS

1323 RHODE ISLAND AVE., N.W.

WASHINGTON, D.C. 20005-3701

1 diagnostic modalities, modality A and modality B.  
2 Which one is better? You look at them and you have  
3 people doing public policy thinking in their minds.  
4 Which one of those is better? You start calculating  
5 something you've seen in a statistical decision theory  
6 book.

7 But the way this is approached in the  
8 field of medical imaging is the following. There are  
9 several possibilities here. Those two points may lie  
10 on completely different ROC curves. In that case we  
11 say that modality B is unambiguously better than  
12 modality A because at any false positive fraction the  
13 sensitivity of A is lower than that of B.

14 There's a different scenario. The two  
15 points could fall on the same ROC curve. Then you  
16 have these same people scratching their heads and  
17 saying, "Where should they really operate?" Well, in  
18 principle we believe that readers can move their level  
19 of aggressiveness. Not on any fine scale but we know  
20 that they adjust depending on the risk group their  
21 seeing. Some people do move around on their ROC curve  
22 so in principle these two points are in equivalent

**NEAL R. GROSS**

COURT REPORTERS AND TRANSCRIBERS

1323 RHODE ISLAND AVE., N.W.

WASHINGTON, D.C. 20005-3701

1 modality.

2           As I say, people will for years say,  
3 "There must be one of these operating points that's  
4 better than the other." Remember when I showed you  
5 that data from Craig Beam you saw people at every  
6 level of aggressiveness. Each one of these people in  
7 some way thinks they've optimized.

8           This is what we call the expected utility  
9 function or the expected value function. Every one of  
10 those people thinks in some way they have found the  
11 optimal operating point but they disagree with each  
12 other so this is another reason for using the ROC  
13 method.

14           There's yet another scenario. ROC curves  
15 may actually fall in such a way that modality A is  
16 everywhere higher than modality B. For the same  
17 reasons we would say that modality A is the superior  
18 modality in this scheme. Three different  
19 possibilities. B higher, equivalent, A higher.

20           This is the motivation for trying to get a  
21 finer measurement on this hundred-point scale. Then  
22 if the clinicians really want to know about the actual

**NEAL R. GROSS**

COURT REPORTERS AND TRANSCRIBERS

1323 RHODE ISLAND AVE., N.W.

WASHINGTON, D.C. 20005-3701

1 operating point, that is another step and we are all  
2 for that if you can coordinate the measurements but  
3 it's very difficult to do that.

4 Well, I'm sure many of you are sitting  
5 there thinking what about if the ROC curves cross? We  
6 know if that happens the situation enters the world of  
7 ambiguity. Then you can no longer necessarily use the  
8 total area under the curve as a sufficient summary  
9 measure of performance.

10 Other summary measures may be necessary.  
11 There are any number of other ways to make a summary  
12 measure of curves that cross. You can use partial  
13 areas. There's actually software even for that today.

14 Or you can use parametric summaries of the curve and  
15 there are several other ways to look at this.

16 If you decided you're going to use other  
17 summary measures, if you anticipate this possibility,  
18 the study protocol is expected to address this because  
19 if you wait until after the study and say, "I was  
20 going to use the partial area in this region," we have  
21 a name for that. That's called data dredging. You  
22 have to build that into your study up front.

**NEAL R. GROSS**

COURT REPORTERS AND TRANSCRIBERS

1323 RHODE ISLAND AVE., N.W.

WASHINGTON, D.C. 20005-3701

(202) 234-4433

[www.nealrgross.com](http://www.nealrgross.com)

1 Otherwise, when people do not expect to see the curves  
2 cross in any real way, they tend to use the area under  
3 the curve as a summary measure.

4 Well, for submissions as are coming before  
5 us in the area of computer-aided detection schemes,  
6 there is a question of how do you keep score for the  
7 location scored. I must remind you this is shocking  
8 to people who have never heard this before.

9 The basic ROC paradigm is an assessment of  
10 the decision making at the level of the patient. You  
11 don't say, "Where does the patient have diabetes?"  
12 You say, "This patient has diabetes." Or you say,  
13 "This patient has TB." You don't say, "The TB is  
14 here." You say, "This patient has TB." So the score  
15 keeping until recent years has been based on decision  
16 making at the level of the patient.

17 In more complex imaging you want to do the  
18 assessment of the decision making at a finer level.  
19 You would like to assess how well the localization was  
20 done. Well, there are little errors there that come  
21 across funny. If you do localization, of course, you  
22 will be providing the experimenter with more

**NEAL R. GROSS**

COURT REPORTERS AND TRANSCRIBERS

1323 RHODE ISLAND AVE., N.W.

WASHINGTON, D.C. 20005-3701

1 information.

2           If you have more information in the study,  
3 you get more statistical power. The trouble is to do  
4 all this adds complexity to the experiment. I would  
5 just like to review for you a couple of the highlights  
6 of the issues that have come up when you try to do  
7 location specific ROC analysis, so-called LROC for  
8 location specific ROC analysis.

9           The biggest problem is that if you want to  
10 keep score of a hit, the measurement of the hit  
11 depends on the criterion you use for localization. If  
12 the legion really is here somehow and you draw your  
13 circle and you say the legion is here, there is a  
14 certain amount of overlap and you would be surprised  
15 to see how sensitive the measurements are to that  
16 degree of overlap to the criterion you use for that.  
17 That's a real issue. There's no unique result.  
18 There's no unique LROC curve at the moment for the  
19 state of the field.

20           There are a couple of subtle points here  
21 that are very technical. I would just like to mention  
22 one of them. People have studied this for 20 or 30

**NEAL R. GROSS**

COURT REPORTERS AND TRANSCRIBERS

1323 RHODE ISLAND AVE., N.W.

WASHINGTON, D.C. 20005-3701

1 years. For a certain class of problems if you study  
2 the ROC and if you study location specific ROC, the  
3 curves in the summary figures tract with each other  
4 monotonically.

5 If the one goes up, the other goes up. If  
6 one goes down, the other comes down. They might  
7 change at different rates but they go together  
8 monotonically. So people haven't felt bad about just  
9 using ROC analysis instead of LROC analysis if they  
10 were willing to invest the extra resources because you  
11 will lose statistical power.

12 But people have been willing not to go to  
13 this level of complexity and to go to that higher  
14 level of complexity requires more elaborate models,  
15 more elaborate assumptions. These are still debated  
16 until today. You can see in the SBIE handbooks that  
17 people are debating this back and forth, Charles Metz  
18 and Dave Chakraborty.

19 But I must mention that a lot of progress  
20 has been made in this field. The bottom line of this  
21 slide if you haven't followed any of this is that  
22 essentially there's a lack of validated software for

**NEAL R. GROSS**

COURT REPORTERS AND TRANSCRIBERS

1323 RHODE ISLAND AVE., N.W.

WASHINGTON, D.C. 20005-3701

1 analysis of such experiments. Now, Elizabeth and the  
2 MIPS, Medical Image Perception Society, website  
3 actually has software for several of these approaches.

4 The writers of that software feel very  
5 good about the state of their software but there  
6 continues to be discussions in the field about how far  
7 have they validated. Have they checked whether the  
8 alpha level and the reject rate are agreeing and what  
9 is the power and so on.

10 The debate goes on but I expect that  
11 people coming down from Pittsburgh any day or any week  
12 now saying, "You've got to start using this because  
13 it's been validated." That's the state of the  
14 knowledge right now. There is software there but  
15 there are still people discussing the condition of the  
16 validation of the software.

17 So a few years ago to find some kind of a  
18 happy medium Nancy Obuchowski of the Cleveland Clinic  
19 and colleagues said, "Why don't we just simplify the  
20 task? Why don't we do something called region of  
21 interest location specific ROC analysis. Let's only  
22 require localization to within a quadrant so you don't

**NEAL R. GROSS**

COURT REPORTERS AND TRANSCRIBERS

1323 RHODE ISLAND AVE., N.W.

WASHINGTON, D.C. 20005-3701

1 have to say there's a lesion here or a lesion here.  
2 You just have to say I see a nodule in this quadrant.  
3 You require localization only up to a quadrant."

4 Similarly for the other quadrants you  
5 could say, "Why didn't we do it for octants or 16 fold  
6 or 32 fold?" Well, you could. This is sort of the  
7 entry level, this problem, but as you add number of  
8 possibilities, then you get more into questions of  
9 overlap and ambiguity so people have decided, "Let's  
10 start at the level of just quadrants." As I say, sort  
11 of the entry into thus problem.

12 Continuing on discussing this so-called  
13 ROI approach, the location specific ROC analysis,  
14 right away Dave Chakraborty jumps into the literature  
15 and say, "Wait a minute. This doesn't correspond at  
16 all to the clinical task." People have debated that  
17 back and forth whether it does or not.

18 But from the other wing of this Greek  
19 chorus comes the methodologist to say, "Yeah, it may  
20 not be quite right but it's really straightforward to  
21 account for correlations without getting into these  
22 assumptions that people have debated for a while."

**NEAL R. GROSS**

COURT REPORTERS AND TRANSCRIBERS

1323 RHODE ISLAND AVE., N.W.

WASHINGTON, D.C. 20005-3701

1                   What do I mean by that? Here are four  
2 quadrants, the right side of the lung, the left side,  
3 the top, and the bottom if you will. Whatever is  
4 going on in this quadrant is expected to be correlated  
5 with what is going on in this quadrant, or at least  
6 could be, and similarly across the quadrants.

7                   After all, this is the same person, has  
8 the same genes, experienced the same environment, and  
9 had a picture taken with the same imaging system. One  
10 has to allow for the possibility that these quadrants  
11 are correlated. The nice thing is that Carolyn Rutter  
12 and others came by another year later and said, "Wait  
13 a minute.

14                   All you have to do to preserve those  
15 correlations is when you resample you resample on a  
16 patient basis. You can't start resampling products  
17 this one from this person and this one from that  
18 person. You have to resample on a patient basis so if  
19 I sample you, all four quadrants from you come into  
20 that sample and so on. When you do this, you actually  
21 preserve the correlation structure and you are said to  
22 be using the patient as the independent statistical

**NEAL R. GROSS**

COURT REPORTERS AND TRANSCRIBERS

1323 RHODE ISLAND AVE., N.W.

WASHINGTON, D.C. 20005-3701

1 unit here.

2 Well, that's all I'll be saying about  
3 location specific score keeping and now to one of the  
4 real problematic issues in the submissions as we'll be  
5 seeing in the next couple years. This is the problem  
6 of uncertainty of truth state. There's a classic  
7 paper that all of us have almost memorized by now from  
8 Revesz, Kundel, and Bonitatibus 20 years ago.

9 This is Harold Kundel known to many of us  
10 as one of the pioneers of this field, the mentor of  
11 someone on our panel today, who was at the Temple  
12 University, and now is at the University of  
13 Pennsylvania emeritus. These authors, what did they  
14 say? They included various ways of obtaining panel  
15 consensus truth.

16 They actually did a study comparing three  
17 different ways of doing chest imaging and they had the  
18 truth but they set the truth aside. They said instead  
19 of depending on the truth to keep score, let's get a  
20 truthing panel. What they found out was they had  
21 several ways of obtaining consensus from that panel.

22 They could either use unanimity. They

**NEAL R. GROSS**

COURT REPORTERS AND TRANSCRIBERS

1323 RHODE ISLAND AVE., N.W.

WASHINGTON, D.C. 20005-3701

1 could use majority. They can use some kind of expert  
2 review. They have three or four ways of reducing this  
3 panel to truth. They compare three imaging  
4 modalities, as I said, and here's what they found.  
5 Any of the three imaging modalities could be found to  
6 out perform the others depending on the rule you used  
7 for reducing the panel to truth.

8 So this sobers a lot of us in the field  
9 about using a panel as truth. However, today the  
10 target of this experiment we'll be discussing today is  
11 not to say this is a nodule that is a cancer. It is  
12 only to say this is a target. This is a region that a  
13 panel of experts would consider to be an actionable  
14 nodule.

15 We're not trying to keep score based on  
16 the truth. We're trying to keep score based on what  
17 would a panel of experts do? Would they cue this  
18 region or not? Nevertheless, even though we changed  
19 the target, this classic reference above tells us that  
20 there's going to be additional uncertainty because of  
21 this panel. The panel will have variability in it and  
22 if you go to RSNA over the last few years, you'll hear

**NEAL R. GROSS**

COURT REPORTERS AND TRANSCRIBERS

1323 RHODE ISLAND AVE., N.W.

WASHINGTON, D.C. 20005-3701

(202) 234-4433

[www.nealrgross.com](http://www.nealrgross.com)

1 papers on this subject.

2           What we've said to incoming sponsors is  
3 that we strongly encourage you to resample, to come up  
4 with some resampling schemes to resample the panel to  
5 get a feel for the additional uncertainty that comes  
6 into this problem over and above the MRMC paradigm,  
7 over and above due to the fact that there is noise in  
8 the panel. You can start to see why there is no  
9 canned software to do this problem.

10           Well, since the truth is uncertain, it  
11 turns out that leads to uncertainty, in effect, in the  
12 number of samples you have. Let's talk about  
13 designing an experiment for a moment. Suppose you  
14 want to design experiments that are going to have very  
15 tight error bars on the sensitivity. Everybody know  
16 that if you want to do that, you want to have a lot of  
17 actually diseased cases to tighten up the error bars  
18 this way.

19           If you want to tighten up error bars the  
20 false positive way, you wouldn't have a lot of  
21 actually non-diseased cases. If your endpoint is the  
22 area under the RC curve, what distribution should you

**NEAL R. GROSS**

COURT REPORTERS AND TRANSCRIBERS

1323 RHODE ISLAND AVE., N.W.

WASHINGTON, D.C. 20005-3701

1 have between nondisease and disease cases? Well, it  
2 turns out it should be some kind of average between  
3 the two. It turns out that the number you should be  
4 using is the harmonic mean of the numbers in the two  
5 classes.

6 The numbers in the two classes is going to  
7 depend on the panel, right? Because some of the panel  
8 members will say these are diseased and others will  
9 say these are diseased. The actual number of diseased  
10 cases depends on the panel. We have uncertainty in  
11 truth that leads to uncertainty in the number of  
12 samples.

13 This is almost a trivial curve and I'm  
14 just going to tell you about the highlights because we  
15 think it might factor in today. Suppose you are told  
16 you can design an experiment with 100 patients. You  
17 say, "How should I distribute them?"

18 Well, you distribute them, let's say, at  
19 the beginning of an experiment like this so that you  
20 have 20 that are actually nodule containing cases, 80  
21 non-nodules, 20 nodule containing sites so we have an  
22 80/20 break.

**NEAL R. GROSS**

COURT REPORTERS AND TRANSCRIBERS

1323 RHODE ISLAND AVE., N.W.

WASHINGTON, D.C. 20005-3701

1                   This effective number, this harmonic mean  
2 of those two numbers, is 32. Whereas if I make a more  
3 even split, 60/40, 50/50, for 60/40 it would be up in  
4 the 40s the effective number. On a 50/50 split the  
5 effective number of samples for that experiment would  
6 then be 50. That's not surprising.

7                   The reason we're showing this is suppose  
8 you start out with an experiment like this and you are  
9 requiring unanimity in the panel to declare a nodule-  
10 present. Then suppose you relax that criterion and  
11 say instead of requiring unanimity, we'll just require  
12 two out of three. Then you expect that whatever the  
13 number was before you're going to move up this curve.

14                   So you are sampling variability, losing  
15 power, but gaining samples. You may tend to cancel.  
16 We don't know this. We are speculating about this.  
17 We'll discuss this. What I just said is if you want  
18 to get into the realm of resampling your panel, you  
19 could start by relaxing the panel criterion from  
20 unanimous to majority and there are several other ways  
21 of doing this.

22                   This is just, again, an entry level. When

**NEAL R. GROSS**

COURT REPORTERS AND TRANSCRIBERS

1323 RHODE ISLAND AVE., N.W.

WASHINGTON, D.C. 20005-3701

1 you do this, this gets you into the game. This allows  
2 you to resample, to assess the variability, but it may  
3 also increase the effective number of samples. These  
4 effects may tend to cancel. This is, again,  
5 speculation just based on the direction of these  
6 effects.

7 The last thing I want to talk about today  
8 is the problem of controlling for reader vigilance.  
9 When you do an experiment, with my two little pads of  
10 paper here, when you read in the unaided reading  
11 condition versus reading the aided reading condition,  
12 there are some people in this room who may be  
13 competitive.

14 If you're reading in the unaided reading  
15 condition you say, "The computer is about to tell me  
16 what it thinks." If you are a little bit competitive,  
17 you are going to say, "I've got to be careful when I  
18 read this." You may increase your vigilance.

19 How do you mock up? How do you do this  
20 experiment? This is a challenge that hasn't been  
21 quite sorted out. Any measurement setting has an  
22 artificial condition compared to the actual real world

**NEAL R. GROSS**

COURT REPORTERS AND TRANSCRIBERS

1323 RHODE ISLAND AVE., N.W.

WASHINGTON, D.C. 20005-3701

1 of practice. What I just described to you is the  
2 possibility that some readers might be more vigilant  
3 in their unaided reading because they know they are  
4 subject to the site.

5 Well, when you turn a modality lose in the  
6 real world, just the opposite could happen, right?  
7 The readers might be less vigilant in the real world  
8 because they know, "Well, I can brush through this.  
9 The computer is going to give me what it thinks in  
10 just a minute." In the real world the vigilance could  
11 go down. In some experimenters it could go up and I  
12 think we've seen experiments when the vigilance didn't  
13 change but I'm sure you can guarantee that.

14 The only thing we've seen in the practical  
15 solution to this problem, Heang-Ping Chan and  
16 colleagues about a dozen years ago wrote a paper in  
17 which they said, "Look, this is a real issue, this  
18 vigilance.

19 How do you do a controlled experiment  
20 controlling for reader vigilance?" They said, "Well,  
21 just simply control the time available to readers in  
22 the unaided reading condition to mimic the actual

**NEAL R. GROSS**

COURT REPORTERS AND TRANSCRIBERS

1323 RHODE ISLAND AVE., N.W.

WASHINGTON, D.C. 20005-3701

1 clinic. That was a suggestion I made. I don't know  
2 how many people have tried that yet but that's in the  
3 air.

4 Well, you can all take a deep breath now.

5 We're in the summary. Here we are. This field has  
6 been going on for 30 years. In the last 10 years the  
7 whole issue of reader variability has complicated it  
8 and there have been ways to promote it to address the  
9 issue of reader variability.

10 In the last few years we've had to deal  
11 with the complications from location uncertainty, from  
12 uncertainty in the truth, this issue of reader  
13 vigilance. What we've tried to do is this is like a  
14 quadrangle, as I said. We hear it sitting at the FDA  
15 and also doing some research here.

16 We have our academic colleagues doing  
17 research in academia, industry sponsors doing research  
18 on all these issues in another side of the quadrangle,  
19 and NCI and the Lung Image Database Consortium that  
20 we've been very actively working with and who are very  
21 interested in these issues.

22 We've tried to hold the windows open so

**NEAL R. GROSS**

COURT REPORTERS AND TRANSCRIBERS

1323 RHODE ISLAND AVE., N.W.

WASHINGTON, D.C. 20005-3701

1 that this quadrangle from all courts has been open to  
2 everyone. Whenever industry sponsors have come in  
3 with issues like this we've said, "Look, the windows  
4 are open.

5 Here's what is known from all these  
6 quarters. Here are the papers. Here are the drafts  
7 that are not even published yet. Here's what we know  
8 at the moment. We don't have guidance. We can't say  
9 this is where the FDA or anyone is holding the bar but  
10 this is all the knowledge that we have at the moment."

11 There is no canned software. There's  
12 canned software for little pieces of this problem so  
13 any industry sponsor would have to be creative to come  
14 forth with a novel way of putting all these pieces  
15 together.

16 Well, that's the state of the world as we  
17 know it today. Thank you very much for your interest  
18 in this. Oh, there's some papers. The "tz" are  
19 obviously Charlie Metz's papers. There are a few  
20 papers from our own group in which we have actually  
21 worked with Charlie Metz and our own statisticians and  
22 our clinicians try to review the state of the world.

**NEAL R. GROSS**

COURT REPORTERS AND TRANSCRIBERS

1323 RHODE ISLAND AVE., N.W.

WASHINGTON, D.C. 20005-3701

(202) 234-4433

[www.nealrgross.com](http://www.nealrgross.com)

1                   This is the first LIDC document. It's  
2 going to come out in April. Then in your notes there  
3 are many other pages of references.

4                   DR. IBBOTT: Thank you, Dr. Wagner.  
5 Before you go too far, I would like to ask if there  
6 are any questions from the panel for Dr. Wagner.

7                   DR. KRUPINSKI: What's the consensus? I  
8 mean, the quadrant problem gets rid of the  
9 localization problem if you end up with a nodule in  
10 each quadrant. What it still hasn't addressed, what  
11 do you do, for example, when you've got two lesions in  
12 a quadrant?

13                   DR. WAGNER: That's right.

14                   DR. KRUPINSKI: You still have that basic  
15 uncertainty.

16                   DR. WAGNER: That's right.

17                   DR. KRUPINSKI: The flip side of that is  
18 what if there is a false positive in the quadrant  
19 along with a true positive? You've just simply  
20 squished it --

21                   DR. WAGNER: That's right.

22                   DR. KRUPINSKI: -- into a quadrant and you

**NEAL R. GROSS**

COURT REPORTERS AND TRANSCRIBERS

1323 RHODE ISLAND AVE., N.W.

WASHINGTON, D.C. 20005-3701

1 still have avoided the localization problem and the  
2 problem of a false positive and true positive.

3 DR. WAGNER: That's right. That's been  
4 sidestepped. As you know, the higher levels of  
5 software attempt to address this one way or another  
6 and I think the jury is still out on whether we are  
7 ready to use that. I think the inventors of those  
8 other methods think they are ready to go and they  
9 might be but we also know there are people in the  
10 wings saying I'm not sure about these assumptions and  
11 so on. That software does not have general providence  
12 right now. Maybe that's too bad. Maybe it should be.  
13 These are real issues.

14 DR. BLUMENSTEIN: I'm impressed by the  
15 MRMC study design. I think that's a nice step  
16 forward. I'm wondering if anybody has ever subjected  
17 the same reader to the same image multiple times and  
18 studied the effect of that so that you could get at  
19 this issue about how a single reader uses their own  
20 personal scale?

21 DR. WAGNER: Yes. That's a classic  
22 question. There are experiments on that. I'm making

**NEAL R. GROSS**

COURT REPORTERS AND TRANSCRIBERS

1323 RHODE ISLAND AVE., N.W.

WASHINGTON, D.C. 20005-3701

1 this up but this is the spirit in which I remember it.

2 David Getty has shown some data on this in  
3 mammography and I think that readers are correlated  
4 with each other in the 60 percent range and are  
5 correlated with themselves only 70 some percent on  
6 repeats. There is, indeed, a lot of reader  
7 variability intro.

8 However, you get more bang for buck -- if  
9 you want to spend so much time in radiology reading-  
10 wise, there's more bang for buck to get a different  
11 reader than to use the same reader over again because  
12 you are so correlated with yourself you get more  
13 independent information if you bring in a sample  
14 that's not so correlated with the preceding reads.

15 Bank for buck-wise people have said this  
16 is a question of reading time. People have not in the  
17 MRMC paradigm in general tended to have readers  
18 reproduce their readings. You can do it and there are  
19 terms in the model to accommodate that, of course.  
20 It's just not common.

21 DR. BLUMENSTEIN: Actually, you took my  
22 question as a suggestion maybe of changing the study

**NEAL R. GROSS**

COURT REPORTERS AND TRANSCRIBERS

1323 RHODE ISLAND AVE., N.W.

WASHINGTON, D.C. 20005-3701

1 design. I didn't make it clear. What I'm actually  
2 concerned about is whether the methodology that's been  
3 developed to give p-values, estimate variance, which  
4 you rightly point out are the big issues here, whether  
5 those properly account for intra-observer variability  
6 in their use of the scales?

7 DR. WAGNER: I believe it does and I'll  
8 tell you why. The full model has seven terms. I  
9 won't take you all through all of those seven terms.  
10 Pure case, pure reader, various interactions. One of  
11 them is a three-way interaction between modality  
12 reader and case.

13 That's the sixth term. The seventh term  
14 is what you're talking about. It's the lack of reader  
15 reproducibility. If you do enough experiments, you  
16 can identify so-called in statistical language. You  
17 can separate these two. If you don't do the right  
18 experiment, you can't but they get lumped together.  
19 The term you're trying to get at is the reader  
20 inconsistency. That is sampled in the experiment but  
21 it cannot be identified. It cannot be broken out but  
22 it is in there.

**NEAL R. GROSS**

COURT REPORTERS AND TRANSCRIBERS

1323 RHODE ISLAND AVE., N.W.

WASHINGTON, D.C. 20005-3701

1           In fact, the way we do it is we do it with  
2 a family bootstrap experiment so we can actually put  
3 out all these effects but we cannot pull out the MRC  
4 from the epsilon. They come together. That  
5 represents not only this three-way interaction but  
6 represents the inconsistency of all the data sets  
7 together. So that is actually in there. Are you  
8 surprised?

9           DR. BLUMENSTEIN: No, no, I'm not. But  
10 since you don't measure that in the experiment, you  
11 can't estimate it obviously. That's the issue. I  
12 guess what I've been concerned about ever since I  
13 first heard about the use of ROC curves where the  
14 reader is recording their result on a subjective scale  
15 either categorical or probability or whatever it is.

16           It's a device to get you to the point of  
17 being able to use ROC methodology. What has always  
18 concerned me was that there was this underlying source  
19 of variability that wasn't taken into account in the  
20 models that you are estimating. It's only if you do  
21 the experiment that way that you actually get an  
22 estimate of that intra-observer or whatever you called

**NEAL R. GROSS**

COURT REPORTERS AND TRANSCRIBERS

1323 RHODE ISLAND AVE., N.W.

WASHINGTON, D.C. 20005-3701

1 inconsistency or whatever.

2 DR. WAGNER: Right.

3 DR. BLUMENSTEIN: I just wondered whether  
4 the degree to which this has been studied in  
5 actuality.

6 DR. WAGNER: Not very much because of the  
7 bang for buck point. As you can see, if you are  
8 inconsistent with yourself, and everyone is, that will  
9 show up in case to case within a given experiment but  
10 you won't be able to peel it out but it's in there and  
11 it's accounted for in the inference. It's a subtle  
12 point but we can discuss it.

13 DR. TRIPURANENI: That was an excellent  
14 presentation, Dr. Wagner. We used the MRMC for the  
15 intra-observation. If you are looking at two  
16 different modalities such as a chest x-ray or a cat  
17 scan, have you looked at whether there is any  
18 difference in the intra-observation between one  
19 modality to the other modality?

20 DR. WAGNER: It turns out to be a really  
21 neat point actually. Our own group has three papers  
22 on this subject. In the first one, you want to know

**NEAL R. GROSS**

COURT REPORTERS AND TRANSCRIBERS

1323 RHODE ISLAND AVE., N.W.

WASHINGTON, D.C. 20005-3701

1 if you can see the difference in the variance  
2 structure between the two modalities. Is that what  
3 you're asking?

4 DR. TRIPURANENI: That's right.

5 DR. WAGNER: There's a model that has six  
6 terms. We were just talking about that. There  
7 another model that -- you would think you would have  
8 to go to 12 terms to do that. It turns out there is a  
9 parsimonious way to do it with just nine terms but two  
10 ways to do that.

11 When you do it you find out that the extra  
12 issues brought up by the wrinkles you were just  
13 discussing, they come in in such a way that they  
14 average and it's only their average that goes into the  
15 inference so you can forget about the issue. It's a  
16 really interesting issue. We have two papers on it.

17 But you could forget about it. You could  
18 from right off the metro just hear about this and say,  
19 "I'm going to use the DBM software." You could forget  
20 about the difference in the variance structure across  
21 the competing modalities and if you do, the inference  
22 is still the same inference. It doesn't matter. It's

**NEAL R. GROSS**

COURT REPORTERS AND TRANSCRIBERS

1323 RHODE ISLAND AVE., N.W.

WASHINGTON, D.C. 20005-3701

1 a really interesting point.

2 DR. IBBOTT: Dr. Solomon.

3 DR. SOLOMON: How do you -- I mean, I have  
4 a feeling this topic is going to be discussed  
5 throughout the day but how do you translate changes in  
6 ROC curves into clinical significance? Especially  
7 since if you look at an individual's change in the ROC  
8 one person might do worse and another person might do  
9 better and then how do you make that determination?

10 DR. WAGNER: Right. Well, you might have  
11 been a fly on the wall in many meetings. I mean, this  
12 is a real issue. Dr. Sacks will say something about  
13 it later on. All I can tell you is that the most  
14 statistical powerful method to get at these  
15 differences is the one I've discussed today.

16 We really would like -- well, I take you  
17 back to the Yulei Jiang stuff. We really do want to  
18 see those action items. You can't go from the curve  
19 easily to the action items if you haven't measured  
20 those action items. Is that what you're getting at?  
21 I'm not sure I see what you're getting at.

22 You want to know how we can go from this

**NEAL R. GROSS**

COURT REPORTERS AND TRANSCRIBERS

1323 RHODE ISLAND AVE., N.W.

WASHINGTON, D.C. 20005-3701

1 ROC summary and inference to an interference to the  
2 clinic. Is that where you're going? I think it's  
3 difficult. What we're saying here is what we are  
4 doing is we are making a measurement that averages  
5 over all these variabilities that we have talked  
6 about. It averages over all that and here's the  
7 summary.

8 If you want something more clinically  
9 relevant than that, you would have to actually measure  
10 the action item, the dichotomization, if you will, and  
11 give it error bars. When you finish the problem is  
12 here would be the action item sensitivity specificity  
13 for the one modality and here it would be for another  
14 one or this way. Now, what do you do?

15 Suppose they go this way? What are you  
16 going to do at this point if they don't match up  
17 sensitivity wise or specificity? What are you going  
18 to do? There are things you can do but you have to  
19 start getting into expected utility analysis. I  
20 didn't mention it but I have some very strong  
21 professional opinions on this.

22 I think it's impossible to do that because

**NEAL R. GROSS**

COURT REPORTERS AND TRANSCRIBERS

1323 RHODE ISLAND AVE., N.W.

WASHINGTON, D.C. 20005-3701

1 to do the expected benefit analysis you need to have  
2 an idea of the prevalence of the disease and that  
3 changes from risk group to risk group so that is a big  
4 uncertainty. You have to have a sense of something  
5 called the utility matrix, the number of false alarms  
6 that you are willing to trade for a hit, if you will,  
7 different from the positive predictive value.

8 You have to have a sense of that utility  
9 matrix and you have to actually know the ROC curve  
10 already because all these things come in. I think  
11 this is almost impossible to do without this being  
12 taken on at a national level.

13 You can see from the data of Beam, et al.  
14 each one of these people thought that they were  
15 working out the optimal operating point and have  
16 completely different points of view. What I'm saying  
17 is that's an important question. I think it's a  
18 societal question.

19 I think it's very complicated and it calls  
20 for a lot of wise people with a lot of data to sit  
21 down with professional societies and say, "Where are  
22 we and where do we want to be?" This is a really big

**NEAL R. GROSS**

COURT REPORTERS AND TRANSCRIBERS

1323 RHODE ISLAND AVE., N.W.

WASHINGTON, D.C. 20005-3701

1 issue. I don't have an easy answer. I insist to my  
2 colleagues there is not an easy answer.

3 DR. IBBOTT: Brent.

4 DR. BLUMENSTEIN: I think it is the key  
5 question. What we are asked to do here is to  
6 basically judge whether this difference in the area of  
7 an ROC curve --

8 DR. WAGNER: That's right.

9 DR. BLUMENSTEIN: -- has any translation to  
10 the clinical setting. What we're lacking we have a  
11 measure of the significance of the difference in the  
12 area of the ROC curve. What we don't have is a  
13 measure of uncertainty around the clinical  
14 interpretation of the ROC curve.

15 This is what is particularly bothersome to  
16 me is I don't know how to do that and I don't see any  
17 methodology that gives me that answer. I'm concerned  
18 that we have started building a building with a  
19 foundation using subjective scales to measure things  
20 so that we can use ROC methodology and we are using  
21 resampling methodologies to do this.

22 We're not taking into account all the

**NEAL R. GROSS**

COURT REPORTERS AND TRANSCRIBERS

1323 RHODE ISLAND AVE., N.W.

WASHINGTON, D.C. 20005-3701

1 various sources of variability and so forth so we are  
2 way out there and our foundation may be collapsing and  
3 not giving us what we need with respect to the  
4 clinical outcomes.

5 DR. WAGNER: Well, if this was broadcast  
6 on academic TV today, apoplexy would abound in the  
7 community because we all feel we are building, as you  
8 say. We're building on decades of people trying to  
9 measure complex perceptual phenomenon. This is  
10 where we are right now.

11 It may not be the ending point to which  
12 you would like to be but this is about the best of  
13 where we are at the moment. I tried to challenge you  
14 a moment ago if you wanted to work on any action  
15 oriented clinical endpoints, I think it's very  
16 difficult to sort that out.

17 It's very difficult because you'll get  
18 bigger error bars and it's very difficult because the  
19 expected utility problem is one that every person in  
20 this room has a different answer to that problem. I  
21 think it's very difficult. I agree with you that we  
22 are constantly besieged by our clinical colleagues who

**NEAL R. GROSS**

COURT REPORTERS AND TRANSCRIBERS

1323 RHODE ISLAND AVE., N.W.

WASHINGTON, D.C. 20005-3701

(202) 234-4433

www.nealrgross.com

1 would like to have better answers to this problem.

2 One case which is kind of unambiguous is  
3 the Yulei Jiang's data that I showed you had an ROC  
4 curve that went up. The unaided condition was lower.

5 The action item, the dichotomization went from a  
6 certain sensitivity to a higher sensitivity and a  
7 lower false positive fraction.

8 I think everyone loves that scenario.  
9 Wouldn't you say? That's the world we want to live  
10 in. Right? That doesn't happen a lot. These more  
11 ambiguous things happen more often. So what we can do  
12 is average over the relevant parameters and say this  
13 is what we found.

14 In principle if one ROC curve is higher  
15 than the other, in principle one can operate at a  
16 given false positive in one modality and increase the  
17 sensitivity. For every time B is higher than A, if  
18 the specificity is here and the curve is everywhere  
19 higher, in principle I can operate at a higher  
20 sensitivity. In practice how to do that, wide open.  
21 This is a professional society issue that is bigger  
22 than all of us. That is a really tough question. I

**NEAL R. GROSS**

COURT REPORTERS AND TRANSCRIBERS

1323 RHODE ISLAND AVE., N.W.

WASHINGTON, D.C. 20005-3701

1 agree.

2 DR. BLUMENSTEIN: And just to throw one  
3 more complicated issue into all this is that a lot of  
4 this stuff that you presented here assumed that the  
5 modalities were assessed independently. In other  
6 words, modality A versus Modality B but the  
7 experiments that we are asked to look at are modality  
8 B added to modality A.

9 DR. WAGNER: Right.

10 DR. BLUMENSTEIN: Where the experiment  
11 itself has built-in constraints with respect to how  
12 one behaves in doing that. I don't see that taken  
13 into account.

14 DR. WAGNER: No.

15 DR. BLUMENSTEIN: And I'm concerned about  
16 that.

17 DR. WAGNER: This is a point of confusion.  
18 I would disagree with you. The modality A here is the  
19 reader unaided. Modality B here is adjuvated, the  
20 reader aided by the computer aid. This a standard  
21 paradigm and it actually corresponds to an experiment  
22 in the real world that you would like to do.

**NEAL R. GROSS**

COURT REPORTERS AND TRANSCRIBERS

1323 RHODE ISLAND AVE., N.W.

WASHINGTON, D.C. 20005-3701

1           It may not line up exactly with the  
2           clinical setting but you actually would want to know  
3           something about the performance of readers unaided and  
4           then you want to know about how they would perform in  
5           the aided condition. That is actually the comparison  
6           of interest.

7           DR. BLUMENSTEIN: I realize that but the  
8           way in which the data are recorded is such that the  
9           judgment -- as I understand it, the judgment under A  
10          is there and has never backed off. You could only  
11          improve.

12          DR. WAGNER: Oh.

13          DR. BLUMENSTEIN: And that's not taken  
14          into account in any of these models that I see. All  
15          the models that you presented, everything that you  
16          said, is based on having an independent assessment of  
17          the two modalities.

18          DR. WAGNER: Well, you have also touched  
19          on something that we have had a lot of discussions on.  
20          These are real issues. I'm not making light of  
21          anything you're talking about here. One hopes the day  
22          will come when these modalities are really good.

**NEAL R. GROSS**

COURT REPORTERS AND TRANSCRIBERS

1323 RHODE ISLAND AVE., N.W.

WASHINGTON, D.C. 20005-3701

1 These computer aids are really good and then you'll be  
2 allowed to back off. You could depend more heavily on  
3 the modality.

4 Today people are being encouraged not to  
5 back off but the measurement doesn't require them not  
6 to back off. They are just encouraged, "Do not back  
7 off," and there is a basic reason for that I think Dr.  
8 Sacks will explain later on so people are encouraged  
9 not to back off.

10 But when the systems are really good as  
11 they are in mammography, these computer-aided systems  
12 in mammography are almost flawless for picking up  
13 clusters of microclassifications. They are far from  
14 perfect for masses but they are almost flawless for  
15 microclassification clusters so readers have thrown  
16 away their eye loops, a lot of them that are using  
17 these systems so they are willing to depend on the  
18 computer.

19 I'm just giving you the only anecdotal  
20 evidence. You have a really good point. I don't have  
21 a really good answer to it but in principle it doesn't  
22 have to be this way. At the moment it is this way.

**NEAL R. GROSS**

COURT REPORTERS AND TRANSCRIBERS

1323 RHODE ISLAND AVE., N.W.

WASHINGTON, D.C. 20005-3701

(202) 234-4433

[www.nealrgross.com](http://www.nealrgross.com)

1 DR. IBBOTT: I would like to remind  
2 everyone we will have time to discuss this specific  
3 proposal in front of us later on this afternoon.

4 DR. STARK: May I ask a question exactly  
5 the point of the presentation, I believe?

6 DR. IBBOTT: Yes, please.

7 DR. STARK: Using the classic -- thank  
8 you. That was an outstanding presentation.

9 DR. WAGNER: Thanks.

10 DR. STARK: Let me just get to the point  
11 because I know we are running short on time. With a  
12 better test the AB test in come context in terms of  
13 clinical utility, either one that had less scatter.  
14 You showed the Beam paper where the radiologist skills  
15 cause scatter in the distribution of the family of  
16 curves.

17 It would seem to me that there would be  
18 two criteria applicable here where we have a different  
19 choice where the test with the larger Az is not the  
20 better test if that test is less flexible -- I'm  
21 sorry, has a larger scatter in terms of variability of  
22 radiology performance, radiology implementation

**NEAL R. GROSS**

COURT REPORTERS AND TRANSCRIBERS

1323 RHODE ISLAND AVE., N.W.

WASHINGTON, D.C. 20005-3701

1 creating a management problem, the implementation  
2 problem and then the clinical utility problem where  
3 all of the fabulously sophisticated group here are  
4 focused on.

5 The other area where the larger Az -- so  
6 if there is more scatter in the test with the larger  
7 Az, it will likely be an inferior test, more  
8 cumbersome, more costly, less safe and less effective  
9 in clinical utilization.

10 The other thing is that if there are two  
11 tests with comparable scatter but is easier to train  
12 with experience or inexperience, so if you have a  
13 trained panel of readers like you do under these study  
14 conditions under very circumscribed conditions where  
15 they know they are in a test and are not distracted by  
16 clinicians, by the busy realistic environment of all  
17 mammography or chest CT practices, you can have a  
18 curve that is more pliant in the direction that you  
19 want doctors to either start at with distractions or  
20 to move into with experience so it does seem to me  
21 that the scatter or the flexibility of the  
22 performance.

**NEAL R. GROSS**

COURT REPORTERS AND TRANSCRIBERS

1323 RHODE ISLAND AVE., N.W.

WASHINGTON, D.C. 20005-3701

1           The ROC curve I think is unassailable and  
2 I have learned -- I have enjoyed a ton here learning  
3 from Dr. Blumenstein's analysis, yours, and those of  
4 you have seen whatever I wrote here. My group had to  
5 do this 20 years ago. We published papers on ROC  
6 analysis and I know we're on the right -- I believe  
7 we're on the right foundation.

8           I think this is the right place to start  
9 but the breath of the challenge facing us all here  
10 today is let's not get obsessed with the ROC curves.  
11 I know we have the whole day for this but the safety  
12 and effectiveness of this is going to be what happens  
13 when you drop into a clinical environment.

14           And we have a lot of experience with  
15 breast and this panel has a lot of people experienced  
16 on it but can you tell me if you would agree that we  
17 need to see the scatter in these Az plots and know how  
18 they respond to inexperience or training to really  
19 know of the larger Az is better.

20           DR. WAGNER: Well, I would say that I  
21 think there is a little bit of second order phenomena  
22 here that is important. Just because something is

**NEAL R. GROSS**

COURT REPORTERS AND TRANSCRIBERS

1323 RHODE ISLAND AVE., N.W.

WASHINGTON, D.C. 20005-3701

(202) 234-4433

[www.nealrgross.com](http://www.nealrgross.com)

1 second order doesn't mean it's not important. For the  
2 practical inferences that have been -- the endpoints  
3 of studies we've seen to date, it has been the  
4 performance in the mean.

5 People have addressed that. There is  
6 software. We have several papers on how to do just  
7 what you say and how to split out every piece so we  
8 can see how much variation is from the cases, from the  
9 readers, from the various interactions. There is  
10 actually software to do that and we are encouraging  
11 people who operate at a higher level, say NCI or some  
12 academic consortium, to address these very issues and  
13 we can see it. We know how to peel all this stuff  
14 apart. As far as the inference on the table today, it  
15 was not done.

16 DR. STARK: The burdens would be huge. I  
17 mean, the sample sizes, the whole time period, the  
18 number of people that have to be involved.

19 DR. WAGNER: That's right.

20 DR. STARK: That's why you talked about  
21 the need for national studies and we would all like to  
22 do that in oncology and everything but we have to

**NEAL R. GROSS**

COURT REPORTERS AND TRANSCRIBERS

1323 RHODE ISLAND AVE., N.W.

WASHINGTON, D.C. 20005-3701

1 treat people and make decisions today.

2 On the other hand, let me ask my final  
3 question. Are you aware, or is anybody aware of any  
4 evidence that a p-value or some other statistical  
5 measure comparing your test A, B under whatever  
6 conditions, today's conditions or the ones I am  
7 dreaming about, we hope it has some clinical relevance  
8 but couldn't it all be counter intuitive? I mean,  
9 this is a very subtle business and couldn't we be  
10 missing the forest for the trees here?

11 DR. WAGNER: Again, that's a very wise  
12 question and I think that is why we have several  
13 medical officers involved in our center on the panel  
14 here so I'll defer to them.

15 DR. STARK: So the p-value of .003 doesn't  
16 necessarily mean a thing.

17 DR. WAGNER: I defer to my clinical  
18 colleagues for that.

19 DR. STARK: Thank you.

20 DR. IBBOTT: I want to make sure that we  
21 give Dr. Mehta a chance to ask a question if he has  
22 one. Dr. Mehta, do you have any questions? He may

**NEAL R. GROSS**

COURT REPORTERS AND TRANSCRIBERS

1323 RHODE ISLAND AVE., N.W.

WASHINGTON, D.C. 20005-3701

1 not be able to hear me.

2 DR. MEHTA: No, I don't have any  
3 questions.

4 DR. IBBOTT: Thank you.

5 All right. We are a few minutes ahead of  
6 schedule at this point so we'll take a short break.  
7 Let's make it 10 minutes and we back at 10:50.

8 (Whereupon, at 10:40 a.m. off the record  
9 until 10:55 a.m.)

10 DR. IBBOTT: Take your seats, please. I'd  
11 like to continue the panel now if you will take your  
12 seats, please. For those of you who are like me are  
13 concerned, we are getting the heat turned down in this  
14 room. At least in one sense.

15 We will now proceed with the sponsor's  
16 presentation which will be introduced by Dr. Kathy  
17 O'Shaughnessy who is Vice President of R2 Technology.

18 Dr. O'Shaughnessy.

19 DR. O'SHAUGHNESSY: Thank you very much.  
20 Dr. Ibbott, we are very pleased to be here today to  
21 present our image checker CT CAD software. I would  
22 like to introduce the attendees that are here from R2

**NEAL R. GROSS**

COURT REPORTERS AND TRANSCRIBERS

1323 RHODE ISLAND AVE., N.W.

WASHINGTON, D.C. 20005-3701

1 and some consultants that we have come to -- we have  
2 asked to be here today to both present and answer  
3 questions from the panel.

4 Besides myself from R2 Technology there's  
5 Dr. Castellino, our Chief Medical Officer; Dr. Wood  
6 who is the head of our CT Products group; and Mr.  
7 Schneider who is the lead algorithm architect that  
8 designed the algorithm that we are reviewing today.

9 In addition, we have asked the following  
10 people to join us. Dr. Delgado was a beta user of the  
11 system so he can describe a little bit about his  
12 experience using the system at his facility. Dr.  
13 MacMahon is a thoracic radiologist from Chicago with  
14 extensive experience in both CAD and ROC research.  
15 Mr. Miller is a biostatistician for the study. Dr.  
16 Stanford was one of the site investigators where we  
17 collected cases from one of the sites.

18 Here is a brief overview of our agenda.  
19 After my introduction we'll go into the current  
20 clinical practice for some background on lung CT and,  
21 in particular, the detection and management of nodules  
22 and lung CT images. Then we'll describe the device

**NEAL R. GROSS**

COURT REPORTERS AND TRANSCRIBERS

1323 RHODE ISLAND AVE., N.W.

WASHINGTON, D.C. 20005-3701

1 both in terms of how it works and how the user uses  
2 it.

3 The clinical study will start first with  
4 how we collected the cases that were used and then go  
5 into detail into the methods and results from the  
6 clinical study. After that we'll have a brief  
7 discussion, presentation about the beta test that  
8 describes a little bit about the usability of the  
9 system. And I'll finally summarize.

10 Before we move into the presentation, I  
11 wanted to put out our proposed indications for use of  
12 this device. I thought it was important to go over  
13 this to sort of put what we are presenting today in  
14 context. The image check for CT is a computer-aided  
15 detection or CAD system designed to assist  
16 radiologists in the detection of pulmonary nodules  
17 during review of multi-detector CT scans of the chest.

18 It's intended to be used as a second  
19 reader alerting the radiologist after his or her  
20 initial reading of the scan to regions of interest  
21 that might have been initially overlooked.

22 I would like to ask Dr. MacMahon to come

**NEAL R. GROSS**

COURT REPORTERS AND TRANSCRIBERS

1323 RHODE ISLAND AVE., N.W.

WASHINGTON, D.C. 20005-3701

1 to the podium, please.

2 MR. MacMAHON: Thank you. Again, I'm  
3 Heber MacMahon. I should say I have a small equity in  
4 R2 Technology. The company has also paid my time and  
5 expenses for this meeting.

6 I would just like to make some brief  
7 comments about the actual clinical practice of  
8 radiology as it relates to thoracic CT scans and the  
9 importance of detection of pulmonary nodules.

10 Some of the common indications for  
11 performing thoracic CT scans would include  
12 characterization of an abnormal finding on a chest x-  
13 ray. In this situation an abnormality may have been  
14 detected and the purpose of the CT scan would be to  
15 characterize it as possibly a lung cancer. And in  
16 addition to detect additional abnormalities that might  
17 be relevant such as metastatic nodules.

18 We also used thoracic CT scans extensively  
19 for staging and monitoring lung cancer and other kinds  
20 of tumors. In this situation we are looking not only  
21 for pulmonary nodules, but also for enlarged  
22 mediastinal lymph nodes and upper abdominal

**NEAL R. GROSS**

COURT REPORTERS AND TRANSCRIBERS

1323 RHODE ISLAND AVE., N.W.

WASHINGTON, D.C. 20005-3701

1 abnormalities.

2 In the case of extra-thoracic tumors we  
3 are commonly also looking for pulmonary nodules and  
4 for enlarged lymph nodes in the mediastinum. Then  
5 there are a range of other applications of thoracic CT  
6 some of which are developing and will be used more  
7 extensively such as detection of pulmonary embolism.

8 However, in all these situations, although  
9 the pulmonary nodules are not the primary focus of the  
10 examination, there is an opportunity to detect  
11 pulmonary nodules that may be present in the lungs of  
12 these patients.

13 Finally, lung cancer screening which is  
14 investigational and depending on the outcome of the  
15 ongoing NLST study may be used more widely. And, of  
16 course, in lung cancer screening pulmonary nodules are  
17 the main focus of the investigation.

18 But the point I would make is that lung  
19 nodule detection is a requirement in every chest CT  
20 scan no matter what the original clinical implication.

21 Only when the radiologist has detected a nodule can  
22 he or she decide what course of action is then

**NEAL R. GROSS**

COURT REPORTERS AND TRANSCRIBERS

1323 RHODE ISLAND AVE., N.W.

WASHINGTON, D.C. 20005-3701

1 appropriate.

2           There are various management strategies  
3 that can be used to manage a pulmonary nodule. In  
4 order to determine whether it's an actionable nodule,  
5 we need to consider the size. Generally larger  
6 nodules are more dangerous and more likely to be  
7 cancerous.

8           We consider the shape whether it's  
9 spiculated, ground glass, and so forth, whether  
10 there's been integral change from a previous  
11 examination in the same institution and that would be  
12 part of the normal diagnostic process to make that  
13 comparison. We would consider, of course, the  
14 clinical context, the age and gender of the patient,  
15 smoking history, and so forth. There are a number of  
16 factors that play into that decision in addition to  
17 the image itself.

18           If the nodule is considered actionable, we  
19 can recommend a number of courses of action. One of  
20 the most common would be to obtain outside prior  
21 imaging studies from other institutions. If we can  
22 establish stability over a period of time, no further

**NEAL R. GROSS**

COURT REPORTERS AND TRANSCRIBERS

1323 RHODE ISLAND AVE., N.W.

WASHINGTON, D.C. 20005-3701

1 action may be necessary.

2 Follow-up CT scan might be prudent at  
3 anything from three months to 12 months depending on  
4 the nature of the nodule and the radiologist level of  
5 suspicion. Other kinds of imaging studies such as a  
6 PET scan may be applicable, especially in larger  
7 nodules that are in the range of 8 to 10 millimeters.

8 This may distinguish cancer from a benign nodule,  
9 Finally, we can consider biopsy, either transthoracic  
10 needle biopsy, bronchoscopy, or thoracoscopic  
11 resection.

12 Just to illustrate the clinical problem,  
13 here is an example of a very small pulmonary nodule  
14 which I think might easily be overlooked in clinical  
15 practice. It's almost indistinguishable on the single  
16 section from surrounding blood vessels but this is, in  
17 fact, a small lung cancer which was detected one year  
18 later, as you can see, at which time it is much more  
19 advanced.

20 So this is a very challenging problem for  
21 radiologists to visually attack these very small  
22 nodules and CT scans. We are aware that we do miss

**NEAL R. GROSS**

COURT REPORTERS AND TRANSCRIBERS

1323 RHODE ISLAND AVE., N.W.

WASHINGTON, D.C. 20005-3701