I think we have answered this at least in part. Let's go through it and make sure we don't have any other comments. Can CAD affect the diagnosis of other conditions?

I think we said yes in our discussion this morning. Is there anyone who wants to change that?

Can it alter the risk-benefit profile of the CAD device? We talked this morning about the co-morbidities and the data suggesting that satisfaction of search is a problem, and that finding something may make finding a second abnormality that may be more clinically relevant difficult.

Do we still believe that, or are there any discussion we need of that?

So the answer is yes. Are there specific conditions that should be represented? If we go back to our discussion of standalone, let me read you some of the things that we talked about.

Scarring, pneumonia, air space consolidation, interstitial disease and emphysema were three confounding variables that we suggested be included in the standalone test to see their effect on the CAD

devices.

Are there any other conditions that we want to add to that list at this time? Yes?

DR. SPINDELL: Just looking at the question, I agree, yes, yes, to both. The question going to the risk-benefit profile really depends on what the indication for use of the CAD device is, and I don't know deep a discussion you can get into on the risk-benefit if we are not sure what the benefit is if we don't know what the indication for that particular device is.

CHAIRMAN GLASSMAN: Well, you are right except that I think there is data that suggests that whatever -- if there is a use, that use may hinder the finding of a coexistent other condition due to satisfaction of search. So it may not really matter so much.

DR. SPINDELL: I totally agree that it can influence. The question is the benefit depends on what you are looking for, and I think that is the question. So it is hard to -- you can define the risk which I definitely agree we have to define, but to define the benefit without knowing what the device is intended use

is a little more difficult.

CHAIRMAN GLASSMAN: Any comments?

DR. STEIER: I agree. The group of findings we described before of the test database with nodules, small nodules, sarcoid, et cetera, might lend itself toward this category as well.

DR. ZISKIN: Just one point, Dr. Glassman.

CHAIRMAN GLASSMAN: Dr. Ziskin? There you are.

DR. ZISKIN: Just one point. If the question had been relative to lung cancer rather than just nodules, I would say we should have non-cancerous nodules in the test base.

CHAIRMAN GLASSMAN: I think we are ready for a summary, but I don't know what it is.

DR. BERRY: Can I react?

CHAIRMAN GLASSMAN: Yes please, Dr. Berry.

DR. BERRY: So I distinguish between standalone and reader studies in this regard, and especially with respect to the last question here. I've been viewing standalone studies as sort of proof of concept, the pivotal study being the reader study.

The reader study should include clinical practice representation. It should not have specific conditions that are hard or easy or eliminated. It should imitate clinical practice so we know what is going to happen when this gets out into the world.

DR. STEIER: Well, I guess what I was trying to list is the things that I see in clinical practice that would be the real things people would see when they go out into the real world.

DR. BERRY: Yes, they would be there, but if they are very rare, then we shouldn't care about them as much.

DR. STEIER: No, no. I mentioned sarcoid, septic emboli, pneumonia, things like that which are relative common, at least in my practice.

DR. BERRY: So certainly they should be included if they represent clinical practice.

CHAIRMAN GLASSMAN: And we did include them in our list, yes. Okay?

DR. BERRY: But don't enrich for them. That is, in the enriched population you are enriching -- or in the enriched cases, you are enriching the total of cancer

cases, let's say, if we are talking about cancer. But we are not enriching particular subsets. We are including a representation of the case population.

CHAIRMAN GLASSMAN: I think that is what we agreed to, yes. Dr. Rosenberg, you had a comment?

DR. ROSENBERG: You would include, for instance, trauma?

CHAIRMAN GLASSMAN: That was not something that we included.

DR. ROSENBERG: Would that make sense?

DR. STEIER: In the evaluation of nodules or in what context?

DR. ROSENBERG: Well, in the context of routine CTs are done in patients with trauma and people get diverted away from looking at the trauma, and they will miss the nodules. So if you are looking for where the device might be more useful, it can be where you are not looking for nodules rather than where you are looking for them.

DR. LEITCH: And that would be a great example of where you -- you know, the really fortuitous thing is, basically, a healthy 20-year-old that gets shot

in the chest and that you could see something where the person does not have other things that are going on other than the trauma.

DR. STEIER: Okay, sounds good to me.

CHAIRMAN GLASSMAN: Okay, so Ms. Brogdon, we want to add trauma to our list of confounding things to be tested.

Summarizing L7: We believe that the effectiveness of CAD would be affected by other disease presence such as the ones we have mentioned, and they should be represented in the test database. And we have listed a number of them.

Is that sufficient?

MS. BROGDON: Yes, thank you.

CHAIRMAN GLASSMAN: Thank you.

This ends our discussion on lung CADs. We are now going to hear an FDA presentation on general issues related to CAD devices and their future developments. This will be followed by the second Open Public Hearing session to give the public an opportunity to once again direct questions to either the Panel or the FDA.

We will now proceed with Dr. Bilek's presentation. Stacie, are you here? Oh, there you go. I got the wrong glasses again. I'm sorry. I really couldn't see you in that white outfit.

DR. BILEK: I am going to go ahead and get started though while Sunder loads this because the first couple of slides were just the outline.

We began our presentations yesterday with an overview of the science behind CAD. We described what a CAD is, the basic components of a CAD, and the clinical use of a CAD, the tools and methods to evaluate CADs.

We then asked the Panel to discuss questions related to the data necessary to evaluate CADs for three types of radiological imaging: detection of regions of interest on mammographic images, detection of polyps on colon CT, and detection of lung nodules and cancer on CT or chest X-ray.

The information provided on these topics will be invaluable as the agency works with industry to continue to bring these technologies to market in a least burdensome manner and in the development of a future guidance document.

We are going to ask the Panel to spend the remainder of your discussion focusing on three areas.

(1) We have several remaining questions related to the evaluation of CAD in general. These were the 'G' questions in the questions that we posed.

(2) The prior discussion was intended to allow you to provide recommendations and advice on specific CAD devices. However, we believe that your recommendations can be applied to other types of CAD; and

(3) As users and researchers, you have insight into types of CAD or CAD-like devices that the agency may see in coming years.

First, the remaining general questions in CAD. Next slide, please.

To briefly revisit some of the concepts reviewed during our background presentation on CAD in general, the basic building blocks of a CAD detection algorithm are outlined here. The digital data is acquired, processed, and segmented. Then features are identified, classified, and finally annotated for the user.

This sequence and the details differ between

algorithms making each relatively unique. We would like you to discuss the extent of information manufacturers should provide regarding their algorithm, its training, and its stability.

Evaluation of a CAD may include standalone testing which was outlined in this diagram from Dr. Petrick's presentation yesterday. Evaluation of a CAD can include reader performance testing, which is outlined in this diagram, also from Dr. Petrick's presentation.

Some reader study designs include multiple readings of the same cases by the same radiologist. Radiologists tend to have long term recall of cases they have previously seen. We would like you to discuss methods for reducing the bias created by this recall, such as delays between the reads which is often term a washout period.

In reader studies, the control or comparison group for CAD-aided computer aided reading has typically been the unaided reading by a single reader. There are alternative controls, however.

They include an unaided double reading by the same reader. The reader is asked to read once again

and subsequently to look again to mimic CAD assistance; unaided double reading by two readers, which is when the unaided reads are made independently by two readers; reading aided with a sham control -- excuse me, with a sham CAD. A sham CAD randomly places marks on the image. We will be asking you a question about these controls.

Once a dataset has been collected, an important consideration is whether or not that test dataset can be reused in the evaluation of subsequent algorithm revisions. The ideal approach is to develop the CAD algorithm, collect test cases, and apply the CAD, then report the standalone and/or the reader performance.

This keeps the testing completely isolated from the training process. However, on subsequent algorithm revisions companies may want to compare performance using the same test cases or an expanded version of this dataset.

It is possible that the CAD developer learns something by simply knowing how the original CAD performed on the test data. This could then be used to produce a revised algorithm using this knowledge.

Therefore after testing a CAD once, the test cases are no longer completely isolated from the training. We would like your feedback on the constraints that should be applied to the reuse of these datasets.

The use of small enriched datasets frequently leads to study populations that do not match the target population in key clinical characteristics; for example, mass size or breast density in mammography CAD.

The distribution of clinical variables varies from study to study, limiting the comparability across studies of the observed CAD performance. A possible but as of yet, unutilized approach for mitigating this lack of comparability is to standardize the statistical analysis by weighing observations according to a designated standard distribution of the clinical variables. We are looking to hear feedback on the feasibility of using such techniques.

The types of analysis methods discussed in this meeting do have their limitations. We recognize that research into methods for assessing and analyzing reader performance continues, and we would like to take

this opportunity to encourage continued development in these areas.

Moving on to the application to other CADs. We have discussed issues related to the demonstration of safety and effectiveness for three families of CADs: mammography, lung, and colon CADs. We have had the Panel discuss the types of testing that are needed for each device type.

In other words, do they need standalone or reader studies, the testing dataset, the study endpoints, ground truth, reader paradigms? We would like the Panel to discuss the application of these same concepts to other image analysis devices.

Computer based technologies have become essential in the practice of radiology. These technologies can incorporate a wide variety of possible functions from the relatively routine, such as image archiving or annotation tools, to complex functions with important clinical ramification, such as a level of suspicion score.

Some of the types of functions you may be used to seeing in your practice include simple display

functions or more complex, semi-automated and automated evaluation tools. An example of this would be organ or vascular segmentation, computer prompting tools such as the CAD detection devices we have been discussing here.

I described on the first day the intent of this meeting was more or less to focus on CAD detection devices. However, CAD diagnosis devices are on the horizon and these could be used on physician identified candidates, examples being ultrasound evaluation of the breast or evaluation of lung nodules or computer identified candidates. Again, these would include lesion rankings on mammography CADe or the probability of malignancy score for a lung CADe, or it could be brain perfusion for the diagnosis of Alzheimer's or stroke.

A spectrum of testing is possible for assessing the safety and effectiveness of these computer based technologies. The spectrum of testing can range from relatively straightforward validation and verification testing. It could also include bench testing with phantoms or limited clinical images.

We have spent a great deal of time talking about standalone and performance testing, and reader

performance studies, and it is also possible that specialized clinical trials may be necessary.

We would like you to discuss the applicability of the evaluation methods discussed in this meeting to various the computer based technologies just described including CAD detection devices other than mammography, colon and lung, and (2) CAD diagnosis devices.

CAD detection and diagnosis has the potential to reach into many areas within medicine. Some examples related to radiological images include CADs that would be used to search for cancer in other parts of the body or to be used with other imaging modalities, CADs which might be used to guide biopsy and CADs that might be used to identify non-cancerous abnormalities.

Finally, CAD might also be used for monitoring the response to therapy or disease progression or to provide some sort of diagnostic assessment.

We ask the Panel to spend the remainder of its time to provide the agency with potential areas of CAD development that we should be prepared to see in the future. Anticipation of future developments allows the

agency to respond proactively.

As you consider these questions, we would like to remind you of the agency's obligation to be least burdensome in our requirements. Thank you.

CHAIRMAN GLASSMAN: Thank you. Are there any questions? Thank you very much.

If there are no further questions, we will now hold the second Open Public Hearing session for this meeting.

You are reminded that the same identification processes, disclosure suggestions and five minute time limit announced for the first Open Public Hearing session this morning still apply to this session as well.

We can now begin the second Open Public Hearing session, and our first speaker is Dr. Akira Hasegawa from Fujifilm Medical Systems.

DR. HASEGAWA: It is five minutes?

CHAIRMAN GLASSMAN: That is correct, five minutes.

DR. HASEGAWA: Thank you very much. I am Akira Hasegawa from Fujifilm. The title of my

presentation is "CAD Evaluation by ROC?"

In this presentation the specific CAD type we would like to talk about is computer aided detection for second read. Currently, one of the FDA requirements for clinical endpoint of CAD approval is to demonstrate a statistically significant improvement of ROC or FROC curve of readers by using CAD.

We question the logic of using the ROC or FROC to evaluate the effectiveness of CAD for second read.

Yesterday I explained how CAD for optional second read in my talk, but I would like to summarize it again. CAD for optional second read helps readers to reduce oversight. Here, oversight includes only perceptual oversight and does not include any cognitive error which is misinterpretation.

It is effective only when readers overlook some ROIs. Here ROIs include cancers, biopsy proven benign, and any suspicious areas. When the radiologist uses CAD as labeled by the manufacturer, more specifically, CAD is used as second read. CAD will not have any effect if there is no oversight.

Let's consider what conditions are necessary to obtain a statistically significant improvement of ROC or FROC in a reader study.

First, we need room to improve. So a statistically significant number of ROIs need to be overlooked by readers. Again, this oversight must be perceptual oversight. Secondly, CAD must help readers reduce this oversight. This is what we want to prove in the reader study and to make this happen, the first condition has to be satisfied.

Only when these two conditions are satisfied, can we logically obtain a statistically significant improvement of ROC or FROC.

Now let's consider what we can conclude from no statistically significant improvement of ROC or FROC observed. This is the converse of the logic in the previous slide. The derivable conclusions are either the CAD did not work as expected and did not help readers, or the readers did not miss a significant number of ROIs although CAD worked as expected.

It is important to realize that an ROC or FROC study for second read CAD cannot identify which of

these two events occurred. It is impossible to estimate the contradiction of perceptual oversight in a reader study because oversight occurs like an accident.

It occurs unpredictably. It is very difficult to reproduce in controlled environment. It is often not case dependent. If it were, oversight would be repeatable, but it is not. It often depends on the environment, reader's physical/psychological conditions. So oversight is a random event and not controllable.

Summary: When there is no statistically significant improvement of ROC or FROC, we cannot conclude that CAD did not work as expected. Perceptual oversight is a random event and not controllable.

While ROC analysis for CADx may be appropriate, ROC or FROC analysis for second read CAD may not make sense if we do not know whether or not oversight occurred.

Thank you very much.

CHAIRMAN GLASSMAN: Thank you.

Next we have -- I hope I get this right -- a unique opportunity. Philips Medical and General Electric are making a single presentation. Roel Truyen from

Philips Medical Systems and Stephen Slavens from GE Healthcare, and they have five minutes each.

Are you coming up one, and then the other?

MR. TRUYEN: Yes. We do our show a little bit like this. So you can imagine if Philips and GE can join forces here, then something will come out, we hope.

So my name is Roel Truyen. I am an employee of Philips Healthcare, and we are speaking, both me myself and Stephen, are speaking on behalf of industry as represented by MITA.

MITA would like to take this opportunity to discuss with the Panel some requirements for data submission of CAD devices. We will mention some general principles and then follow it by a specific example of colon cancer.

Now submission data should provide scientific evidence for the claims made on the device. Dependent on those claims, the type of study or experiment to generate this evidence can range from controlled standalone experiments to full blown observation studies involving clinical readers.

As in all good science, methodology should

be used as published in peer-reviewed journals and other publications. The fact that these methodologies have been reviewed and accepted by scientific peers is sufficient to allow using them for generating this evidence.

While multiple methodologies exist to generate evidence, there is no reason to insist on using one particular methodology for data submission.

MITA supports the use of standardized methodologies and wants to actively collaborate with FDA to define these. Until then, the sponsor may select the least burdensome methodology to generate scientific evidence.

It is, however, not necessary to extend the experiments beyond the claims. Although scientifically interesting, MITA judges that this extension of science is not required for data submission.

As an example, let's discuss colon CAD. The clinical data used can come from either retrospective or prospective studies. In the case of retrospective, the identified cases we propose, informed consent to be extended from the patient or already given for the study.

In the second half of the presentation, we will dive into this.

As there is more than one way to provide scientific evidence, there is also more than one way to define the ground truth. We agree with the panel definition made yesterday for colon CADs, but there are also other ways to do this; for example, an expert reading panel is also acceptable.

That option should also stay open for the manufacturers because optical colonoscopy will become less available in future screening studies because CT has proven its value by now.

So we ask the Panel to also consider this and supplement the recommendations made yesterday.

CAD devices are often claimed to work for a certain lesion size only. Notable methods appear in literature on the best way to measure size, but the scientific debate is still going on. Data submission evidence should, however, not be the place to solve these scientific debates.

We also apply the same general principle to the reader paradigm claimed in the CAD device. As

mentioned before, several choices exist, ranging from the most common second read CADs or the concurrent reads to the maybe more exotic first read CADs.

Our opinion is that the pre-described use should be tested. Other uses can be interesting from a scientific point of view but should not be included in the data submission.

Also, the way of performing these studies should be done according to the state of the art. For example, the time separation between independent and CAD assisted read in the second reader paradigm is usually not done in literature and doesn't have a strong scientific basis.

During the design of a device, we take into account feedback which is from clinical users of the device. The device is meant to improve their clinical practice. Although these improvements are real and are very much appreciated, they cannot always be measured in a simulated study environment.

While every functionality should, of course, be tested by the manufacturer, they should not all be accompanied by clinical evidence. Some obvious examples

of these types of functionalities are visualization, automated measurements, image filtering, more generally speaking, all types of automation and image processing.

Actually, we consider them as not being CADs, and actually more important in the semantics of being CAD or not being CAD, is the fact that we would like to propose the Panel to not consider extensive reader studies to validate these but to consider other methodologies.

Now generating evidence leads to an evidence paradox. Clinical users require flexibility from our devices so they can use the device in a way that best suits their needs.

An example is the color or appearance of the CAD marks. We heard yesterday that they should be large enough --

CHAIRMAN GLASSMAN: I'm sorry, but your five minutes are up.

MR. TRUYEN: Okay, thank you. I would like now -- I would like to hand over the rest of the discussion to my colleague from GE, Stephen Slavens.

CHAIRMAN GLASSMAN: Thank you.

MR. SLAVENS: Yesterday the Panel gave clear support for the use of registry data in CAD device training and FDA submission studies.

Currently, FDA requires that studies for submissions are subject to human subject protections requiring patient informed consent to protect patient confidentiality. However, current regulations do not specifically permit informed consent for exceptions for de-identified images and clinical data in repositories.

FDA has not accepted some recent CAD applications using retrospective data from registries without study-specific informed consent. By definition, it is virtually impossible to locate patients in de-identified collections to consent them. After all, the patient identities are not linked to their clinical data.

As a result, some sponsors have had to conduct prospective studies in support of their applications.

Now concerning the FDA human subject protection -- that is, informed consent -- in the privacy rule HIPAA, many clinicians, research hospitals, and companies view the requirements of informed consent for

IVD studies, for example using leftover specimens, as unnecessary for the protection of human subjects and as overly burdensome and costly, and IRBs broadly agree and support waivering informed consent for radiological imaging studies.

Given the situation, the problem is that for example, in NCI/NIH, in research institution, and industry-supported registry, sufficient data are currently not being accepted in many CAD applications to FDA.

This could affect the well known DMIST mammography and ACRIN colon studies. This fails to leverage the public's and industry's investments and delays the availability of devices aimed at improving the public health.

This is not the least burdensome approach to safeguarding human subjects in bringing new CADs into clinical practice. Retrospective studies of de-identified data do not impact diagnosis or treatment of subjects, are not a health threat, and preserve patient confidentiality, and there is a waiver provision in FDA IRB regulations, 21 CFR 56, that could permit IRBs to

give these waivers.

What MITA encourages the agency to consider is to apply the principles in their guidance on informed consent for in vitro diagnostic device studies using leftover human specimens that are not individually identifiable and, specifically, to declare FDA will exercise enforcement discretion for the requirements of consenting for de-identified patient data to include both the patient images and associated diagnosis in the definition of data, provided both are de-identified; permit IRBs to review and waive the informed consent for de-identified retrospective cases; to advise the sponsors what procedures they should use prior to conducting clinical trials to protect subject identity and confidentiality; to advise sponsors what records they need to keep regarding the conformed consent issue.

In summary of this two-part presentation, CAD submission data should provide scientific evidence for the claims of the device. The sponsor should not be obliged to provide evidence for functions that are not claimed.

To generate scientific evidence, state of

the art methodology should be used as published in peer reviewed journals.

The choice of the scientific methodology should lie with the sponsor.

Procedures must protect patient confidentiality and de-identified data accomplish both the HIPAA and the IDE intent. Clinical studies should thus be allowed to use retrospective data under IRB waiver of informed consent.

Excessive requirements for CAD clinical data and data for advanced visualization software that is not CAD are delaying the introduction of useful innovations to health care.

The recommendations that this Panel provides to FDA are essential to developing guidance FDA and industry can rely on to advance the technology and clinical benefits of CAD.

We thank the Panel for their thoughtful deliberations on this important CAD issue.

CHAIRMAN GLASSMAN: Thank you very much.

Our next speaker will be Pat Milbank.

MS. MILBANK: Thank you. I am speaking here

today on behalf of MITA and at the invitation of Medipattern, a MITA member and one of my clients.

I have been a regulatory attorney and a consultant for 30 years, and I have focused my practice on software medical devices and CAD products for the past 15 years.

The purpose of this presentation is to ask the Panel to provide further clarification on two issues that we have been discussing during this Panel meeting and among ourselves and with the agency for the past two years.

The first issue is the question of whether sponsors should be required to conduct off-label studies for approval of their devices. It should be noted for the record that Section 513 of the Food, Drug, and Cosmetic Act provides that the FDA shall establish the safety and effectiveness of a device based upon the indications for use proposed by the sponsor.

The agency may, of course, require additional labeling or warnings regarding potential off-label use, but in light of yesterday's mammo CAD discussions regarding reader paradigms, the Panel should

be advised that it is inappropriate under the statute to require sponsors to study their products for off-label use.

For example, products indicated only for second readers -- second reading models should not be required to be studied under a concurrent reader model as well. Companies should not be perceived by the public as promoting off-label use of their products or required to study and publish the results of so called off-label studies.

We ask the Panel to clarify that off-label studies are not required for approval of these devices, as required by the statute.

The second issue we wish to raise involves the recent requirement of the agency to conduct studies with washout periods between reading sessions.

Yesterday in Slide 25 from Dr. Smith, he cited the requirements for establishing effectiveness of a device. The regulation states that effectiveness is based upon testing the device in the target population for its intended use and under its intended conditions of use.

Therefore, we request that the Panel consider whether the effectiveness of washout studies is appropriate for use in CAD reader models.

We recommend that the proper study method to establish effectiveness of these devices is the real time intended use study design which was originally designed to meet the regulatory requirements for approval of these CAD devices. Next slide.

This slide provides two very recent examples over the past year. They are still pending. The first example describes a study design proposed by the agency for colon CAD devices.

Now I want to point out that the sponsor specified that this product would be marketed for second read only, and that concurrent read would be contraindicated. However, the agency required a study design at baseline reviews, a second read, a sequential review and followed by a concurrent review which is now an off-label use. This matter is still pending.

In the mammo CAD field where you heard yesterday we have four PMAs. They have identical labeling, and they were tested under identical

methodologies. Newcomers to the field, however, are being asked to agree to conduct studies as described here and to provide non-comparative studies to the four preceding manufacturers.

As you heard -- next slide, I'm sorry. These requirements for multiple-arm studies with various washout periods do not satisfy the least burdensome requirement imposed by Congress.

It is also worthy of note, and Sophie mentioned it this morning, that requiring off-label studies may also require a re-engineering effort, including verification and validation to confirm the product has been designed correctly to conduct the off-label study which the company will not be able to use.

Last slide. In conclusion, we ask that the Panel carefully consider and support the recommendations of industry; that the design of studies should be tested in correspondence to the claims that are being made; that the sponsors not be required to study off-label uses of their products, in accordance with the law; that the standard study design, which for most CAD studies is a second reader study, be conducted in a simulated clinical

use environment; and --

CHAIRMAN GLASSMAN: I'm sorry. Your time is up, thank you.

MS. MILBANK: Thank you.

CHAIRMAN GLASSMAN: Our next speaker is Dr. Maryellen Giger from the University of Chicago.

DR. GIGER: Thank you. I am from the University of Chicago. I will be speaking on beyond computer-aided detection going toward computer-aided diagnosis and quantitative image analysis.

I am representing myself. My research is supported as shown here, and I receive research funding from R2/Hologic.

Okay. The potential of CAD is expanding. Beyond computer-aided detection, it has a potential to reduce interpretation errors, reduce variation between and within observers, improve the visualization of the image data, improve efficiency of the interpretation, and yield quantitative measures.

Basically, computer image analysis is becoming an integrated step in the diagnostic decision making process.

For example, shown here is computer aided diagnosis in the work-up of suspect lesions. A computer is being used to help characterize the lesion and potentially indicate a computer determined probability of malignancy or a malignancy score, leaving the final decision to patient management.

You see the benign and malignant. The one in the middle is a little confusing. That is a malignant case.

So let's look at ones that reduce interpretation errors. Various studies have been shown. This one by Jiang has shown that the computer can help radiologists improve their interpretation of clustered microcalcifications.

Besides just giving a malignancy score systems are now incorporating online databases that can be searched based either on the lesion characteristic or on the estimated probabilities of malignancy.

An example is shown here, where the case in question are the upper images, and the computer can show either a malignancy score, similar cases where the outline in green is benign, the outline in red is

malignant, or a malignancy score histogram where the unknown case is indicated in this case as showing that it is most likely benign, where here the green is benign and the red is malignant.

This is being extended for ultrasound and also MRI now. As breast imaging goes multi-modality, so does CAD. And this study has been shown also to aid readers in an observer study.

Computer aided diagnosis research is also being performed in chest CT, here as in distinguishing between malignant and benign lung nodules, and once again an improvement. This is a study by Lee showing the computer added.

It is also being used to reduce variation between readers. Studies have shown, for example here, that use of CAD reduces disagreement between readers, attendings, residents and so on.

Also, it can be used to help efficiency of the interpretation. For example, in breast MRI where we have 4D information, we have information overload going to the radiologist; and with CAD we can take the lesion segmentation, the extraction of the relevant area of the

breast lesion, look at kinetic data, morphological data, merge it into a likelihood of malignancy, all within a matter of seconds now, and this could help the overload.

So I just want to end with: to help translate these, we are going to require many of the things we did for computer aided diagnosis including an independent technology assessment institute, and we need to consider, and I would like the Panel to consider, what the potential for computer aided diagnosis as a concurrent read.

Basically it would be another clinical tool along with other tests, both image data and clinical data, in the diagnostic work-up. Of course, the final decision would be the radiologist's, who would interpret all these tools, whether it be image-based, information-based, computer-based.

We need to separate the diagnostic test performance from the user performance.

Going back to this technology assessment institute, I am concerned that with all these new CAD devices being developed and submitted to the FDA, we are going to run out of cases. We are going to run out

readers.

With this institute, it could be tasked with the performance assessment of new or improved CAD devices.

These would be for standalone tests. It would be consistent and standardized. When one needed something tested, they could randomly extract a subset from the institute's large database, many of them which were talked about today which we need to protect them soon.

The subset could be selected so it matches the desired population that is being tested. It could report only the overall performance scores instead of performance on individual images, and all these would help maintain the integrity of the test set and help this industry move along to get the improvements in technology to the public.

Thank you.

CHAIRMAN GLASSMAN: Thank you. Is there anyone in the audience who would like five minutes to speak at this time? Please identify yourself when you get to the podium.

MR. VASTAGH: Thank you, Mr. Chairman. Not five minutes, perhaps one.

I am Steven Vastagh representing MITA and the manufacturers.

We were pleased, and I was pleased personally, to hear in Dr. Bilek's presentation a few minutes ago as she invited your advice and she said she would be working with industry alongside with her colleagues to work on this matter.

We, the industry, have heard it, and we have been offering to work with the FDA on this matter, and would like to reinforce that response to work with FDA as we go forward to evaluate these issues and come to a guidance document and resolution of these important matters.

On behalf of the industry, I thank the Panel for your work in these two days. Thank you.

CHAIRMAN GLASSMAN: Thank you very much. Do any of the Panel members have any questions for any of our speakers? Dr. Berry?

DR. BERRY: Is it okay that I react rather than question?

CHAIRMAN GLASSMAN: Why don't we save the reaction for the general discussion if you don't have a question for them.

DR. BERRY: Well, this is for a very specific statement and person.

CHAIRMAN GLASSMAN: Still, let's keep it for the general discussion which will be happening very soon. Any questions? Yes?

DR. CARRINO: I can make my reaction in the form of a question.

CHAIRMAN GLASSMAN: If it's a good question.

DR. CARRINO: With the GE-Philips, Philips-GE combo presentation --

CHAIRMAN GLASSMAN: Would the two speakers please come to the podium?

DR. CARRINO: I had two main questions. One was a general question. They suggested alternative methodologies, and if they can expound upon what those methodologies, that would be one question.

The second question is with regard to the washout period which actually is a well-established tool that is used by people who do observer performance

studies. If they meant that you could obviate having a washout period if you did a second reader type study, where the first initial naive or unaided reading becomes what you would consider your unaided reading and then the second read is the enhanced reading -- so I wanted them expound upon those two things.

DR. TOURASSI: Actually, I would have the same confusion. How is it possible to test a concurrent paradigm or the first reader paradigm without the washout period?

CHAIRMAN GLASSMAN: We will let you try to answer that.

MR. TRUYEN: Well, I will start with the third and last question. In the case of a concurrent paradigm you indeed need an independent read, but in the case of a second reader paradigm, you can -- the reader will read the case unaided, immediately followed by review with the CAD results. And in that case, we estimate it is not necessary to do an independent read.

DR. TOURASSI: I am in full agreement with that. Sequential reading is well accepted, but you presented it in a more general way that the washout

paradigm is -- it is not acceptable. Yes, everybody wants to save time, but if a sponsor comes and proposes the concurrent read paradigm or the first reader paradigm; yes, blue skies, how can this be proven without the washout?

MR. TRUYEN: Sorry for the confusion that I caused on that, but it is true that there's not many options there left, but still the length of the washout periods is still not yet determined. And if I talk to our radiological collaborators to say, once you have seen a case, you remember it even --

DR. TOURASSI: Well, if we go by literature, there seems to be that rule of thumb of one month.

DR. CARRINO: Yes, 30 days, it's pretty standard. It would be hard -- I mean even to get somebody to reread them in a shorter period of time is logically hard. I don't think that is burdensome.

CHAIRMAN GLASSMAN: I think, rather than a discussion we have had the answer. You did have another question, though, about alternative methodologies.

DR. CARRINO: Yes, that ROCs may not be suitable so expound upon the alternatives.

MR. TRUYEN: Well, in the case, specifically of what I call lung CADs or, let's say, ultimate measurements, I think the Panel in a previous discussion; it was already indicated that some standalone testing possibly using phantoms, scan phantoms, can be sufficient in there.

CHAIRMAN GLASSMAN: Thank you. Any other questions for -- Oh, yes?

DR. KIM: In terms of -- You advocate that we could substitute an expert panel instead of colonoscopy as ground truth. Are you saying that for like, say, greater than 10 millimeter lesions?

MR. TRUYEN: Well, depending on the type of lesion that you want to study. I heard yesterday larger than 6 millimeters, possibly larger than 10 millimeter.

The point that I wanted to make there is currently optical colonoscopy is still done for larger than 10, but indeed of smaller ones they will be much more into the follow-up mode; and instead of having to wait for three years or whatever to follow up, then I would propose to also use an expert panel for those cases. That, in the future will probably not be sent to

optical colonoscopy right away.

DR. KIM: Okay.

CHAIRMAN GLASSMAN: Another question?

DR. SPINDELL: Same presenter. In your presentation you had mentioned CAD --

CHAIRMAN GLASSMAN: Microphone, please.

DR. SPINDELL: Okay, you mentioned in your presentation CAD, and then you mentioned at the end something about non-CAD. I got a little confused. So could you just explain what you meant by non-CAD? I know what you meant by CAD, but what did you mean by non-CAD?

MR. TRUYEN: Well, there was a presentation yesterday, also MITA presentation, where they made it a little bit complicated looking at assessment tree. There, non-CADs were defined based on the human intervention in that and also the risk analysis was on there.

If you do remember that scheme, on the bottom right there were some techniques. They say computer aided measurement or automated measurements of size, length, volume, visualization, volume rendering follow-up measurement, growth rate measurements.

So there is a lot of, well, automation into image processing that is happening, that has been happening for years now that has proven its value; and we kind of would like to propose to keep on evaluating those techniques, as we have until now, not necessarily using reader studies for that. But I agree that the border between CAD and non-CAD, I think, in that assessment is relatively clear, if we talk about CAD, the CAD types, mammo, lung and colon that were discussed today and yesterday.

CHAIRMAN GLASSMAN: Did you have a question?

DR. STEIER: I have a question, actually, for Dr. Giger -- a couple of quick questions. I'll be brief.

In one of your slides you talk about requiring training of CADx users for proper use. Who would provide this training? How would it be documented? What kind of competencies would be expected? What do have in mind?

DR. GIGER: Well, I think training for use of CAD should -- CAD is going to become an integral part of radiologists' life, and actually it should be

integrated from the beginning during residency. However, because we are starting with folks who are beyond residency now, there could be training courses.

I think it would best be run by academics as opposed to industry just to keep it unbiased. Radiologists vary in their performance, some of them because of lack of training or retraining.

DR. STEIER: Okay. Next brief question, you would require QA of CADx systems, quality assurance? Is there not quality assurance now?

DR. GIGER: Well, because CADx is new, it is not really out there yet. I just want to make sure that people just don't put a system in and not keep an eye on how it is going. It was more of a warning to make sure that is performed.

DR. STEIER: Okay, my last question is your other comment was the need to separate the diagnostic test performance from the user, i.e., radiologist performance. What do you mean by that?

DR. GIGER: Well, I believe in reader studies when they are very properly designed with a distribution of cases and the distribution of readers.

However, when it is useful just to have a standalone test to show improvement, sometimes -- I believe those would work better because you could have a lousy mammogram and a great radiologist, or a great image and a radiologist who is not that good at reading mammograms.

I think we have to remind ourselves to look at both the tool and who is using it; and if you do an observer study and your radiologists are not that good, you are going to get -- you will get a different result than if you do an observer study and you have very good radiologists.

So it was more of just everyone, just remember to keep these two separate, even though they end up being integrated.

DR. STEIER: Thank you.

CHAIRMAN GLASSMAN: Yes, Dr. Mittal?

DR. MITTAL: A question for Dr. Giger. Can you expand on your concept of technology institute, and I assume you are just talking about CAD, not all the technology. Who will be the sponsor? Who will be funding it in this time of funding cuts from Federal government? How do you envision that?

DR. GIGER: Well, I think some folks could see it going beyond CAD, but I just focused on that today. I have actually thought about this for multiple years, and it's always amazed me why no one has -- no entrepreneur has taken it up.

To me, it is similar to when you are at the Emmy, and they get the envelope and it has been verified by the accounting company or the lottery. You need a private, not-for-profit company that can collect all these cases to make sure -- a database is not just a collection of cases.

It is very careful annotation, verification of the truth, and as more and more companies come about and they are all trying to do this by working with this hospital, another one working with another hospital -- to me, it's a waste of resources.

If everyone worked together and put them in this institute, of course, the database would have to be large enough so that when you did have a system test, you could randomly select from the large pot cases that reflect distribution of your population; and that subset is what you would do your standalone test on and you

would get a score of 82 percent, and that was it.

If you reach the benchmark required by FDA or someone, then your incremental improvement in your CAD could go on and be implemented, and I see this as an efficient way and a resource savings way of getting technology tested, the standalone ones, into the public.

CHAIRMAN GLASSMAN: Thank you. Are there any other questions? Go ahead.

DR. D'ORSI: Maryellen -- Dr. Giger, I just wanted to clarify in my head what compilation you are recommending for CADx, and is it combined with CADe?

DR. GIGER: It is being -- well, I see it used right now in the diagnostic work-up where you are looking at mammograms, ultrasound, MRI. You have patient clinical data on all of it. You have different modalities. You have to -- well, I don't know. I'm not a radiologist. Radiologists have to interpret. So I have it at the work-up stage, to help with it.

CHAIRMAN GLASSMAN: Yes, Dr. Abbey?

DR. ABBEY: My question was for Pat Milbank. So you brought up, I think, a really important topic that we struggle with, and I would prefer not to have to think

about off-label use at all if I could. I imagine the rest of the FDA feels the same way.

So here is my question. If the FDA approves a CAD device and it goes out onto the market, and it is widely used off-label to the detriment of, say in mammography, the women who have the scans, who is responsible for that? Has the FDA failed those women? Has the company failed those women, or have the physicians failed those women, and how would you assign responsibility?

MS. MILBANK: If it is used off-label?

DR. ABBEY: Yes. With no proof -- there is no way to know from the software --

MS. MILBANK: Well, let me clarify that. You heard from Nancy Brogdon this morning that the FDA has no authority under what Congress has issued to require post-market studies on off-label uses.

The purpose of my talk was to clarify they also do not have that authority in pre-approval studies. Do I think those studies should be done if that's the way the product has been evolving over time?

I agree with what the Panel said yesterday

that we are headed in that direction, but to require those studies at this point in time when we carefully label and train, then that is a requirement beyond what Congress has authorized us to do. But should those studies evolve over time? Should those uses evolve?

The studies will have to be done. You will have to have washout periods though they are not well established. Have we just found that very smart people never forget anything they have ever read?

I checked with Charles Metz on this at University of Chicago, and that was his quote of the day.

I do agree that we have to watch as the future evolves. We also work in a business and a legal environment that we have to deal with, and the companies have to be able to market products in a way that means they can successfully provide them on a regular basis, despite our very unique legal environment in this country.

CHAIRMAN GLASSMAN: Thank you very much. Let me move on now. The book says that we are supposed to have a coffee break now, but we came back late from lunch. So I would like to defer that a little bit.

Actually, we started lunch late is the truth.

I would like to go into the general discussion of CAD and future devices. Dr. Berry, I know you have something you want to say. I would like you to go first, if you would.

DR. BERRY: This refers to Dr. Hasegawa's presentation that he suggested that we don't have to show a statistical significance of the ROC, that oversight is a random event and not controllable, and it makes no sense to try to prove effectiveness if we don't know whether or not oversight occurred.

I submit that if oversight hadn't occurred, then we don't need the CAD.

The standard of evidence-based medicine is proof. A hundred years ago we used to do things because some expert said it worked. Now we require that it be proven. Statistical significance is the standard. If we throw it out, we will be going back 100 years or more.

It is not necessary, however, that statistical significance be at .05. .05 is a completely arbitrary level. In some circumstances, such as treating a rare disease, .05 is too stringent. But more

generally, it is too liberal, and in the current setting I think you should take this as a gift, that you can show at .05 and not have to show something more stringent. It actually is quite a weak criterion.

So I don't want to move back into the dark ages.

CHAIRMAN GLASSMAN: Very good. Thank you. Other general CAD comments about future developments?

DR. CARRINO: I just wanted -- There was a statement made that IRB waivers for the de-identified data -- and I think that is totally suitable, and it should be supported. So if there is a dataset out there and the patient has already consented to an original study and is now put in this de-identified database, it is very common to waive getting consent again for a CAD-related study, and that should help facilitate doing these studies.

CHAIRMAN GLASSMAN: Ms. Brogdon?

MS. BROGDON: This is a very complicated area, and every time it comes up we have to go back and re-review the regulations, and why we have taken certain positions.

The studies that FDA reviews for the devices we regulate are subject to what is called 21 CFR Part 50, informed consent, and also Part 56 which is about institutional review boards, IRBs.

Part of Part 56 allows an IRB to waive written informed consent. So what we have advised firms is that there may be other ways to obtain informed consent other than strictly in writing. They can contact the patients and so forth.

Because images are de-identified sometimes creates problems for us because we still need to be able to audit the data. That means going back to the source records and comparing that information with what the companies submit to us.

So if there is no connection between the data that we have to review and the source records that creates problems for FDA's obligations to audit data.

So that is what we are dealing with, and there are also other regulations that these studies are not subject to from FDA's point of view that does allow waiver of informed consent. That is sometimes confusing to companies.

So what we encourage companies to do is to come and talk to us before they do their studies so we can discuss the issue of informed consent early.

CHAIRMAN GLASSMAN: Thank you very much. I think that clarifies the fact that we are in over our heads at this Panel if we try to discuss informed consent.

General questions or comments about CAD in the future? I know in the later questions -- in fact, it's the last question of the day -- we get into kind of the borderline between CAD and no CAD, and what measurement technologies -- you know, what level of independent intelligence for a measurement technology pushes it over into the CAD review.

I think all of us would agree that cardiac scoring for coronary artery calcification is probably a CAD. Is the automatic measurement of intimal thickness in a carotid ultrasound a CAD? That's sort of on the other end.

I think one of the things that we are going to be asked in a little while by me and by the question is to try and fit the line between CAD and no CAD in the

things that are in PAC systems or embedded in ultrasound machines or CT scanners that are very innovative measurement tools that require some artificial intelligence, and therefore, may require some different level of testing. Or will bench testing with a phantom be sufficient?

Things -- obviously, computer aided diagnosis is the next big area, and I think we have touched on the different level potentially of scrutiny for something that is meant to diagnose a specific disease rather than to simply identify something for evaluation.

Anybody have any comments about that?

DR. STEIER: Well, I have comments, but -- I guess they kind of tangentially related to that.

The two things that occur to me most is the issue of training and competency in these new modalities, and who is going to provide it.

We require our residents to be competency trained, our patient care assistants, our nurses and everybody else in the hospital; and the issue of training and competency as these new things are developed and

implemented is one thing that seems to be a striking issue.

The other is the preponderance of the off-label uses of the product as well, using it not always as a second reader but sometimes as an initial reader or a concurrent reader, and the issues that are inherent to that.

So I think those are two issues along with QA, which was mentioned -- so a third issue -- that are things that will really have to be fleshed out as the CAD process proceeds.

CHAIRMAN GLASSMAN: Dr. Bourland?

DR. BOURLAND: Sort of two comments general and two maybe a little more future.

One is in the FDA diagram Step 1 was acquired digital data. So the modalities are expanding, and there are issues of image quality and image fidelity for each of those. So a number of lines will, so to speak, point to that box or be within that box.

We talked a little bit about phantoms relative to lung, but in fact they are very useful tools throughout the entire process, and it should not be

forgotten that images can be finessed, adapted, revised, disease added or subtracted.

There are ways to do this so that essentially you have virtual phantoms by manipulation of digital image. That could be used for test cases. Maybe that applies best to standalone.

For the future, for lung in particular, we talked about screening diagnosis in early stages, and especially small nodules. So just say, the one thing we are all keeping in the back of our head relative to radiation treatment, is that we would very early detect lesions at a very small size, and then with hypo-fractionated treatment address those with perhaps ionizing radiation or some other type of ablative side.

The question is that, maybe is very much a CAD approached system that incorporates then, both the imaging and diagnostic focus, as well as treatment.

The one thing I thought about mammography is -- and this is a question for the radiology colleagues here -- that we talked a little bit about first reader. But I wondered about the use of mammography CAD relative to service for underserved populations, and whether there

was any opinion from radiologists on the appropriateness of that.

CHAIRMAN GLASSMAN: If I can answer that question briefly, I assume you are implying as an only read for a mammogram in an underserved population.

DR. BOURLAND: I don't know if it would be only, but maybe the first and not unattended, I guess I would say.

CHAIRMAN GLASSMAN: I think the proof now is that the better read is the CAD as a second read, and I don't think that the -- in terms of the speed, while it does make it a little slower, I think that in the absence of other data, it would stay that way for now.

Data may come that it can be done differently and done faster, but for now, I think in an underserved population a lot of the issue is access to equipment and patients coming in for the exams. It is not just availability of readers.

Other future comments or comments about the future? They are all future comments until you make them.

DR. SAHINER: So maybe what I will say will

be a paraphrasing of what you said in the first place. But I think there are many devices that are being designed to do measurements, and in the FDA document -- I am just paraphrasing from the document -- it says that CADx device is designed to process a specific finding in order to characterize the finding.

So, for example, if we have systems that measure the size of an abnormality over temporal images to see if the lesion is growing to characterize it as malignant or benign, or if we have some, again measurement methods to look at the response to therapy, would these be considered as CADx devices or CAD devices?

I think -- I don't know if we have the answer now, but I think this is an important issue to consider.

CHAIRMAN GLASSMAN: Other comments? Oh, yes, Dr. Rosenberg?

DR. ROSENBERG: I think the decision of what is a measurement tool, what is CADe, what is CADx, and how we divide those, will be an interesting question.

CHAIRMAN GLASSMAN: Yes. and we will get to it in just a little while. Yes?

MS. FINKEN: One comment in line with what Dr. Rosenberg said or, actually, I guess it was at the end of the table there.

I do think we need to keep in mind the quality assurance for all, that this very sophisticated equipment might not reach down to those levels of people who are either underinsured or not even insured or in areas that are too remote to take advantage of these systems. Just to add that into the comments on the future.

CHAIRMAN GLASSMAN: Thanks, Ms. Finken. Thank you.

Any other comments? I think now would probably be a good time to take a 10-minute coffee break instead of a 15. Can we do that and all get back on time? Thank you very much.

(Whereupon, the foregoing matter went off the record at 3:10 p.m. and went back on the record at 3:22 p.m.)

CHAIRMAN GLASSMAN: At this time we will begin our discussion on the FDA questions related to general methodologies and future CAD devices. Copies of

the questions are in the meeting handout and on the tables outside of this conference room.

So this is Question G1, G for General.

To what extent should sponsors provide algorithm descriptions, training dataset descriptions, standalone performance of the device on the training database, and/or stability analysis of the algorithm to training as part of the original CAD submissions or as part of subsequent algorithm updates?

I guess this looks at probably some trade secrets as well as information before the testing done by the agency or mandated by the agency.

Anyone want to begin? Dr. Ziskin?

DR. ZISKIN: Well, I am curious about algorithms, but I don't think that is as terribly important for me to know the details of it ahead of time, but I would care very much about the way it was tested and the test series and so on, about the analysis and everything beyond that, I think, is very important. It is just the algorithm itself -- I feel that is probably proprietary. I don't need to know that.

CHAIRMAN GLASSMAN: Yes, Dr. Tourassi?

DR. TOURASSI: I am in complete agreement. It doesn't really matter what is going on with the training set. Did they use 10 cases, 1000 cases? Was it robust? Wasn't robust? As long as the test set has been independent; there was no biased selection of what the test cases would be. That's what we care about, standalone test performance and later performance.

CHAIRMAN GLASSMAN: So just to make sure I understand, we care not about the training of the equipment but about its performance on the test sets.

What about stability of the algorithm?

DR. TOURASSI: On the training set?

CHAIRMAN GLASSMAN: Yes.

DR. TOURASSI: Because that is how the question is phrased.

CHAIRMAN GLASSMAN: Yes.

DR. TOURASSI: Doesn't matter.

CHAIRMAN GLASSMAN: Any other -- Dr. Dodd?

DR. DODD: I would agree that the testing is fundamentally what is important, but I am not sure I clearly understand the FDA's role for verifying that the company has done what they have done and whether they

have happened to pick the particular cut between the test and training set that is the lucky one, and if there is any role in that in terms of what they actually -- that they have to evaluate what the company has claimed they have done that they have actually done.

DR. BOURLAND: Comment?

CHAIRMAN GLASSMAN: Comment, yes.

DR. BOURLAND: I have a different opinion on algorithms, and I do not know how thorough this needs to be, but often algorithms are previously tested. Some have been published.

Yes, they might be protected previously by patents and things like this before disclosure and use, et cetera. However, in general, a statement such as mutual information or something tells the user about the method that is being used within that algorithm, and hopefully, the user may have a sense for limitations and strengths that are associated, and then these would be borne out by testing.

So I think at least the name of the algorithm, even if it just says least squares fit gradient function or whatever it might be, is of value.

CHAIRMAN GLASSMAN: You said value to the user. Did you mean the user or the agency?

DR. BOURLAND: I guess both.

CHAIRMAN GLASSMAN: Any comment about that statement?

DR. TOURASSI: I am not sure whether there is going to be value to the user. How many radiologists would necessarily understand concepts with this information?

DR. BOURLAND: Perhaps it would be a smaller group of people called physicists or something.

CHAIRMAN GLASSMAN: Dr. Watt?

DR. WATT: As the end user, I am far more interested in the labeling that tells me that it has been tested and tested appropriately. I am not interested myself in knowing the algorithm itself, and I could care less about that.

I want to know the equipment, that it is reliable and is going to be functioning in a standard methodology. So therefore, I have to rely upon the FDA testing and labeling.

CHAIRMAN GLASSMAN: Yes?

DR. SAHINER: I agree with what has been said about the algorithm descriptions and the training dataset descriptions maybe, but for stability I think I have a different view.

If an algorithm is not stable to the training set, it means that the next time the company comes up with a supplement, maybe training with a slightly different dataset or changing some of the parameters, and you see a huge performance in -- you might see a big change in performance, then it becomes an issue.

So I do believe that some analysis of stability to the training might be important, and especially -- not only to maybe the training dataset but also how the parameters are selected.

DR. BERRY: Can I?

CHAIRMAN GLASSMAN: Dr. Berry, yes please.

DR. BERRY: With respect to telling the user that its least squares or whatever, I don't think that is essential, and I would worry about the patent stuff.

The role of the FDA -- I mean, historically -- the FDA, of course, knows what their role is much

better than I and knows where they are going much better than I.

Historically, the FDA hasn't worried about those kinds of things, you know, how you got to where you are. The same thing is true in drugs.

I don't want to talk about drugs very much, but in drugs we screw up royally in the early phases, and the FDA has now realized that they should be helping companies to do better in the preclinical and the early clinical trials. So they have something called the critical path initiative, which is precisely defined to go back in and help companies to develop things in a more efficient way so that they have a better, more focused, what they call Phase III trial.

The same thing, I would encourage the FDA, CDRH, to do the same thing here to help companies, to teach them, because they have much more experience at this sort of thing than the companies do, where they are going and to configure themselves so that they get there in the most -- in the least burdensome way.

CHAIRMAN GLASSMAN: Ms. Brogdon?

MS. BROGDON: One of the purposes of our

asking this question was to find out whether you believe that FDA needs this information. It wasn't only a question about what should be presented to users. It is also whether you have opinions on what information about algorithms and other related things should be sent to FDA.

CHAIRMAN GLASSMAN: Dr. Swerdlow?

DR. SWERDLOW: One of the things I would like to know as an end user regarding standalone performance is, as we have discussed at length over the last couple of days, are there particular subsets or types of calcifications or villous adenomas that the CAD does not perform as well at than others, even if there is not as much statistical power because thee are fewer of them? I think that is very important to know as a user.

CHAIRMAN GLASSMAN: So that would be a labeling issue, but what about the specifics of algorithm and the effects on -- the results on training database? Is that something that the FDA should have access to, or should their interest begin with the testing phase rather than the training phase? I think that is the nature of the question.

DR. TOURASSI: I am so confused with the amount of information that is requested because even if somebody comes and says this algorithm uses technique X to do segmentation and then Bi propagation neural network to do training, yeah, great.

All of this is great information for the scientists, but still what is the value to that, even to the FDA? They cannot -- If they don't have access to the neural network in the actual training data to double-check everything, does it really matter? Does it give more value to you to know that it was a neural network versus some --?

CHAIRMAN GLASSMAN: Ms. Brogdon?

MS. BROGDON: I would refer you to Dr. Petrick's talk from yesterday about the stability of the algorithms, and if you would like, I would imagine Dr. Petrick could address some of this if you would like to hear that.

CHAIRMAN GLASSMAN: Could you just make a one or two-minute comment? If you can keep it to that, about --

DR. PETRICK: So let me just clarify. So we

are talking about -- we could be talking about two different things. What should be presented to the user? What should be presented to the agency?

The agency can receive patented information, proprietary information that is kept out of the public database, so that it is not something that necessarily has to go out to the public.

The question then is, is the information on those descriptions useful not just for the original submission but again, for subsequent submissions that come in? How would we know whether an algorithm changed or not, based on what information is provided to us? That is the basis of the question.

DR. BOURLAND: I have a comment on this.

CHAIRMAN GLASSMAN: Dr. Bourland?

DR. BOURLAND: So I will change what I said about this being of value to both. I think it would be of value to FDA, in particular, and it allows an opportunity for risk assessment.

My experience has been very different with algorithms. I sort of want to know what they are, but I have had -- yes. So this has been of value because

limitations and the range over which an algorithm may be possible for use might be limited in some scope.

Well, what if FDA, so to speak, is aware of that relative to that algorithm and then has a dataset submitted, and the question is do they compute or not? They ought to match up and things like this.

So I could see value for those who are assessing risk for the device.

CHAIRMAN GLASSMAN: Dr. Berry?

DR. BERRY: Just quickly, I completely agree, and I think it is important for the FDA not only for what Dr. Petrick said but for competitive circumstances. You know, another company comes through with another neural network that is doing exactly the same thing. They will have the intelligence from the previous setting to be able to guide and possibly help in the development process.

So I think that they've got to build a data bank to know what the basis was.

CHAIRMAN GLASSMAN: Dr. Rosenberg?

DR. ROSENBERG: I am not sure if it is more of a question or a comment. But in terms of stability,

it seems like it could be important, given that there are subtle changes in the technology, and would that inform the FDA as to things to be more concerned with?

DR. TOURASSI: I think stability is important, but stability on the testing set, and if we'll go back to the idea that Dr. Giger proposed, if there is this repository, repository of images where more cases can be selected; it is more valuable to see the stability of whatever is the algorithm to different subsets or random selections of that collection, rather than say, yes, it was stable on the training set.

So what? Does that stability translate to the testing set?

CHAIRMAN GLASSMAN: Dr. Sahiner?

DR. SAHINER: Dr. Tourassi, I think the question is asking about the stability of the algorithm to the training, not on the training set but how you train.

DR. TOURASSI: On the training database.

DR. SAHINER: Yes, but the second part is "or stability analysis of the algorithm to training." So I read it as not the stability of the algorithm on the

training set but how does training affect its stability on the test set?

DR. TOURASSI: Okay, I read it differently.

CHAIRMAN GLASSMAN: Dr. Spindell, do you have any comments that may be helpful to us?

DR. SPINDELL: I might. I guess the overall answer is I don't think it is a big deal to send the algorithm to the FDA because you might want to have some of the specific manufacturers more because it is still protected information.

If the FDA feels that would help them in processing the application in a quicker manner, I think the manufacturers would not be upset with that.

CHAIRMAN GLASSMAN: Thank you for that observation. We are still stuck between what matters is the testing phase, not the training phase for the agency, or that the agency should have access to the training database as an advantage.

I can't possibly summarize that as a coherent answer, and I, sitting here, see – certainly, I see both sides of it. So I can't even shade the answer to the way I believe.

So are there any other comments? Dr. Abbey or Dr. Garra?

DR. GARRA: No, I agree.

CHAIRMAN GLASSMAN: To which side?

DR. GARRA: Well, I don't know if there is really two sides to it. I mean, the companies, I think, should, can, and if they are able to, provide the algorithm to the FDA, but I think that the comments to the user, the algorithm is not going to be so important.

CHAIRMAN GLASSMAN: This has nothing to do with the user. We are talking about to the FDA.

DR. GARRA: Yes, to the FDA, but is it required? Even the FDA would have trouble evaluating some of these algorithms, in particular, neural networks. Beyond knowing that it is a neural network, there is no way to look inside there and see what is going on.

So the performance during training and the performance on the tests are what are really going to be important.

CHAIRMAN GLASSMAN: Training or testing? You threw both together.

DR. GARRA: Both.

CHAIRMAN GLASSMAN: Does the FDA need to know the performance on the training?

DR. GARRA: They need to know the parameters of the training set.

CHAIRMAN GLASSMAN: The parameters, but not the performance?

DR. GARRA: Yes, because that gives you an insight into how stable it is going to be.

CHAIRMAN GLASSMAN: Other comments, please? Dr. Abbey?

DR. ABBEY: So there is various ways, I think, to do this. I'm sort of imagining sort of our cross-validation approach where you would take case sets, put it into a training set and a testing phase, and train your algorithm on the training set, evaluate it on the test.

I guess the stability is in, if you don't like the answer, can you start re-deciding which one is the training and the testing, and I think the idea here is to say -- is to get away from a specific training set and a specific testing set.

I think it is not much more -- at least as I

understand it, much more burdensome to do a sort of a generalized cross-validation approach where you get the stability as well. So it strikes me -- if I am not mistaken in that, I would say it is not much of an additional burden to request stability. But somebody correct me if that is not the case with the specifics.

DR. SPINDELL: I have a question for the FDA.

CHAIRMAN GLASSMAN: Yes, Dr. Spindell.

DR. SPINDELL: As part of the evaluation submission, I was under the assumption you would be allowed to request something like the training dataset as part of your evaluation. Is that not true if you needed that information for the approval?

MS. BROGDON: I believe that is true, and we have done that. I need to ask the staff if there is a follow-up question here.

DR. PETRICK: This is Dr. Petrick. I would just, I guess, clarify. The details of the algorithm may be associated with really, the complexity of the algorithm. So understanding every single detail -- I'm sure the FDA staff probably isn't going to understand

everything that goes on.

The issue is again, understanding details about the algorithm would give some indication of the complexity of the algorithm. And the question again, is whether that is important for the FDA to understand again, not just on the original submission, but how to go back on subsequent submissions that come along for the same device?

DR. GARRA: If I could just make a quick comment here?

Although knowing the algorithm per se may not be specifically helpful, the FDA could be pointed toward -- usually, algorithms like this are based on some literature that has been published; oftentimes maybe a dozen years before or something, or it is in some military declassified document.

If the FDA is made aware of the source of the algorithm, even though they don't know the specifics of this particular implementation, oftentimes there is follow-on papers that talk about stability or instability in a various algorithm, and it might be helpful for their evaluation.

So that information should be provided.

CHAIRMAN GLASSMAN: Dr. Sahiner, I think you had a question.

DR. SAHINER: Yes, well I think it is the same point that Dr. Garra made a moment ago that complexity and stability are actually intertwined. So the more complex an algorithm is or the more parameters it has, it may become less and less stable.

So from that perspective, I think I agree with the fact that not the algorithm description: that this is this and this is that and this is how they come together. I think an order of magnitude of about the number of parameters used in the algorithm is important to evaluate the stability.

DR. GARRA: That is a very good comment.

CHAIRMAN GLASSMAN: Ms. Brogdon?

MS. BROGDON: Maybe to simplify things, I should just turn the question around. Is there anybody who feels that FDA should not be requesting this sort of information?

I heard Dr. Tourassi express some

reservations about some of this.

DR. TOURASSI: I still don't understand. Okay, you get information about the algorithm. There are many parameters. I agree with you. It is complex.

Past performance is really good. What do you do next? The next algorithm, the next revision comes back even more complex, or some parameters are tweaked more. But the test performance supports using the device.

Yes, it's great to have the information, but it needs to be used in a meaningful way other than subjective assessment. It is complex, moderate complex, not very complex.

CHAIRMAN GLASSMAN: Any other comments? Dr. Ziskin?

DR. ZISKIN: I wanted to address the question that was brought up of the value of knowing an algorithm change to see whether it is really important or not.

I don't see where the knowledge of that is adequate. You would then have to test it anyway to see whether or not the performance changes. In other words,

is a change big or small? Well, it depends upon did it affect performance.

So you are back again to the testing of the thing, and the details of the algorithm don't really matter that much. I think it's the testing that really matters.

CHAIRMAN GLASSMAN: Dr. Abbey, you look like you are about to press the button.

DR. ABBEY: My only concern here is that we have been saying that for a modification of the algorithm, we might accept only standalone testing; and so then if we don't know what the algorithm is, how do you assess the magnitude of a modification as to whether it goes back for standalone testing or whether it requires -- it's a substantial enough change in the algorithm to actually require a revised reader, say?

CHAIRMAN GLASSMAN: Dr. Sahiner?

DR. SAHINER: I think this relates to another question that will come up. I think the next question, the appropriate constraints on the use of a test set. So some algorithms can be modified in such complex and particular ways that they may work very well

on a test set that has been used multiple times.

If you are just looking at that particular test set and standalone performance on that test set, then I think it would be a dangerous thing to do.

DR. TOURASSI: I am in complete agreement with you that the next point is far more important than the first one, and somehow both are related. But going back to Dr. Abbey's comment, if this extra information, the details about the algorithm are going to be used to assess whether future changes in the algorithm are substantially not for naught. Yes, then I understand the value, but just for the sake of getting more information? I don't see it.

So if it is going to be useful because you need a quantitative measure for that later on to assess what is substantial change or not, yes. But then, of course, you will have to come up with measures. So what is substantial and based on the complexity of the algorithm?

CHAIRMAN GLASSMAN: Not this afternoon, we won't. But let me try to summarize because I think we have come to an agreement here, that to the extent that

the agency needs the data to evaluate the stability and future changes, the data from the standalone performance and the algorithm details should be available to the agency as part -- so to evaluate now and for subsequent algorithm updates.

Is that -- do I think we are there now? I think so. Ms. Brogdon, is that an acceptable -- Is that a sufficient answer to your question?

MS. BROGDON: Thank you, yes.

CHAIRMAN GLASSMAN: Okay. now let's get on to this important one, G2. Not that they are not all important, but I have just been told this one is critical.

What may be appropriate constraints on the reuse of test data in order to balance data integrity and data collection for CAD assessment? Appropriate constraints, Dr. Tourassi?

DR. TOURASSI: Will I start, or will you go?

CHAIRMAN GLASSMAN: Please start.

DR. TOURASSI: It will go against the least burdensome approach, but anytime test data is used to evaluate where the system fails so that future upgrades

need to be made and tweaking the algorithm, as I say, the same test set should not be used again.

So I don't know if Dr. Sahiner wants to elaborate on that.

CHAIRMAN GLASSMAN: I think he does.

DR. SAHINER: I think that -- I agree that it is a dangerous thing to use the same dataset over and over again, and just then it becomes part of your training set. But on the other hand, I think some of the knowledge gathered from the previous version has to be compared with what happens in the next version. And some very general descriptions like if a company finds that their performance on calcifications is this much but their performance on masses is worse, and they have to concentrate their efforts on the detection of masses, then I don't regard this as reusing the test dataset as part of training because it is so general.

DR. TOURASSI: I am in complete agreement with that. What I am saying is that the test set for that next phase should be supplemented with new cases somehow and not rely on exactly the same cases from which we realize that, yes, the algorithm is not doing as well

with masses so let's go back and focus on these type of masses with this size parameters or this type of morphology -- whatever is the knowledge we are extracting from failing in a subset of the test set.

CHAIRMAN GLASSMAN: Dr. Sahiner?

DR. SAHINER: I agree, but it also depends on how general that information that was used is. Yes, you would say you know masses detected at this location in the image are -- or masses at this location of the image are undetected, then I think it would be a very bad use of the test dataset.

So I do agree that datasets should be supplemented, but I don't know how much.

CHAIRMAN GLASSMAN: Oh, I think that would depend on how big the change to the algorithm would be. With the example of it is a major issue of masses, then the supplement should be a major issue of masses.

On the other hand, if it is very minor, then you might not need to supplement as much to prove the point. Is that a reasonable way to look at it?

DR. TOURASSI: Yes, but we are supposed to come up with certain guidelines. As you realize, G2 is

extremely difficult, and we knew from the beginning that that is a sticky point here.

CHAIRMAN GLASSMAN: Right.

DR. LIN: Can I ask, should the test data be replaced or just supplemented if there is a major change? I mean should we have the old test data still present and then just add cases to that because that can contaminate?

DR. TOURASSI: What is major change? That goes back to that. I mean ideally it should be supplemented or changed.

CHAIRMAN GLASSMAN: Dr. Garra?

DR. GARRA: These are all big, difficult issues, but one way to handle it is like what Dr. Giger said. That is, you have a test dataset available, and they get the results, a summary of the results, but they are not allowed to use that dataset to train. They don't get enough information back to train their algorithm, and then that protects the integrity of your dataset.

The point was made by my colleague here, Dr. Abbey, that -- well, resubmitting it over and over again, that is affecting training. What you can do is limit the

number of accesses to that dataset and then go to a different one for future tests.

CHAIRMAN GLASSMAN: Dr. Sahiner?

DR. SAHINER: I was just going to make the point also that, when you change something in the algorithm, you don't want the performance to become worse on the previous test set. So I do believe that maybe supplementing so that you can also look at how it did with the previous test sets would be important. But again, this is the difficult issue.

So does it mean that the test set is going to get larger and larger and will it be a burden? I don't know.

CHAIRMAN GLASSMAN: Yes, Dr. Berry?

DR. BERRY: My initial answer to the question was appropriate constraints are always to constrain and never supplement and always replace. That is sort of a most burdensome approach.

So I do think that there are circumstances when you could use some of the data. You might even argue I am going to use the same data. We can't here make those decisions. That would have to be something

that the argument that this is a scientifically high level could be made by a company to the FDA. My guess is that the FDA would say I'm sorry, go back and get another test set so that we can understand it.

If you are supplementing and you are trying to analyze the supplementary dataset, it drives statisticians crazy. I mean, it is almost impossible to do. A fresh dataset, you can do.

We've got to be building, we've got to be thinking about building settings where you can use data as they accumulate over time, but false positives and multiplicities just are everywhere, and the only way -- I shouldn't say the only way, but the cleanest way is to say start again; give me a fresh dataset -- maybe it doesn't have to be as big as the last one -- that shows that you've got the required sensitivity and specificity.

I think the answer to the question that we should give is that it is appropriate to place really rigid constraints and relax those only if the FDA is persuaded by the companies' arguments that those aren't necessary.

CHAIRMAN GLASSMAN: Dr. Dodd, any comment at

all?

DR. DODD: I would just agree with Dr. Berry that I think we should set the appropriate constraints. I think everybody understands the risks of reusing the test data although that said, I can foresee some situations in which you would allow limited reuse of the test data. But I do think that onus should be upon the company.

CHAIRMAN GLASSMAN: Any -- yes, Dr. Garra?

DR. GARRA: I just wanted to say that yes, I think it's okay to put the onus somewhat on the company, but there is not an unlimited supply of test data out there within the budget of most companies.

So I think the FDA needs to move forward in conjunction with other government agencies on building some of these test data sets, just like the American Board of Radiology is constantly asking us for new test questions to test the quality of applicants. That is something that the regulatory bodies need to have for these companies as well, uncontaminated test datasets that can supplement the company's efforts.

CHAIRMAN GLASSMAN: Any other comments about this? Go ahead.

DR. LIN: To me, this is analogous to developing scoring systems in outcomes research, like for example in NGI we have lots of scoring systems like the MELD score for liver failure.

You know, when we develop the scoring system, we use a set of data to develop the system, but then when we validate it, there has to be a completely separate set of data; and any cross-contamination, I think, is a major problem.

I recognize the burdensome concerns here, but I would agree with the previous speakers that we should really be very careful about letting people reuse test set data.

CHAIRMAN GLASSMAN: Dr. Abbey?

DR. ABBEY: One additional comment is that this is a little bit of bringing coal to Newcastle. I consider researchers at the FDA to be some of the leaders in the field of looking at effect of training and testing. So they have a lot of in-house expertise in this particular issue.

The other comment I would make is that the multi-reader, multi-case is intended to generalize to the

population. So at least statistically, it seems to me that that is the appropriate measure, to use all new datasets, use a multi-reader, multi-case approach.

I would totally agree that that is highly burdensome, at the same time. So that is, I guess, the crux of the issue.

CHAIRMAN GLASSMAN: Dr. Tourassi?

DR. TOURASSI: The whole issue was not necessarily for the reader paradigm. It was for the standalone performance, not for the reader study. Obviously, you are not going to use the same case here.

CHAIRMAN GLASSMAN: Okay, Dr. Bourland?

DR. BOURLAND: I have a question for FDA. Would you accept data from a nationally credentialed, maintained database, and does that make your job easier? Wait a minute. We don't want to think about that. But in other words, characterize -- fully characterized database, et cetera? Is that of value to FDA?

MS. BROGDON: I would like to ask Dr. Petrick to respond to that.

CHAIRMAN GLASSMAN: You must have gotten the short straw today, Nick.

DR. PETRICK: I think that there are certainly details that go into how those databases are developed, but in general FDA has used databases before. Tissue banks are one example of databases that have been used in submissions.

So I don't think there is anything inherently that keeps us from using that data. Again, it's the details of how it is collected, how it is controlled, and how it is maintained that would be the details of how it would be used.

DR. BOURLAND: And follow-up to the manufacturers -- is that of advantage to you in any way for least burdensome?

DR. SPINDELL: It would be, but I think you heard earlier that there is some concern about the FDA's accepting of data without a verbal informed consent; and some of those databases, it would be nearly impossible to go back and get verbal consent from all those people.

So I think that issue needs to be worked out as well.

CHAIRMAN GLASSMAN: And that is an issue beyond the scope of our discussion.

Let me try to summarize what we have said and give everybody a chance to comment on it, and then maybe we will come up. I think I've got the sense that the Panel has severe concern about the reuse of test data, and that optimally a new test set should be obtained. However, we realize that there will be certain circumstances where that will either be unnecessary or so burdensome that a lesser solution would be acceptable, knowing the quality of the people at the FDA and that they wouldn't accept anything less than what was adequate.

How does that sound to the Panel?

DR. BERRY: It sounds like you are a politician.

CHAIRMAN GLASSMAN: I'm trying to thread the needle here to say that we really think that new test data is the best. However, we realize that that may not always be possible, and so to make that a blanket statement is probably overly rigid given what I have heard over the last 15 minutes; and we are in Washington. Dr. Berry?

DR. BERRY: So why don't you say that

however, there may be circumstances that the FDA would accept a partial use of the former test data, something like that?

CHAIRMAN GLASSMAN: I think you said it better than I tried to say it, but that was what I was trying to come up with. Is that acceptable? Is that an answer that helps you or do you need more discussion?

MS. BROGDON: I think you have given about all you can.

CHAIRMAN GLASSMAN: Okay. Thank you. Let's go on to G3. I don't know quite what that meant, but we are moving on.

In a paired design, when each reader reads images with and without CAD, should there be a washout period between readings? Secondly, do you have any suggestions for improving paired designs for reader-CAD studies?

Why don't we take the second one first, because it may affect the first answer? Any suggestions for improving paired designs for reader-CAD studies over what is generally done now? Dr. Spindell?

DR. SPINDELL: Could we just clarify exactly

what the question is asking because is the question asking -- and this was the confusion we had before? Is the question asking that, if the study is reader and then reader with CAD -- do they mean a washout period between those two readings; or is it reading, reading with CAD and then a washout period before another read?

DR. TOURASSI: I think it is both.

DR. SPINDELL: You know, we are always instructed as a manufacturer to test the device in the intended use setting, and the intended use setting is the reader reads it and then turns the CAD on and reads it again.

DR. TOURASSI: Actually, I agree with you based on the discussions we had before. Sequential reading is the least burdensome. It is accepted. Yes, the document that we got from FDA outlines some really nice reading paradigms, but in terms of efficiency, to go through and do the randomized design, which was beautiful -- I really liked it, but it requires a lot of effort.

So at least for the second reader paradigm, sequential reading should be more than enough. Now when it comes to new paradigms of concurrent reading or the

first reader, then we have to go through that washout period.

CHAIRMAN GLASSMAN: Yes, I think the answer is depending on the intended use, and if the intended use is other than second read, then you don't have that internal control on the study.

So let's go back to the second part again. Any suggestions for improving paired designs for reader CAD studies, or are the paired designs that are done good and sufficient? Dr. Sahiner?

DR. SAHINER: Before that, I want to add one thing to the first discussion point. I know that the least burdensome way of doing it is the sequential reading. I just want to point out that, when you are doing sequential reading and the reader knows that CAD is going to follow, it actually affects the user's behavior and it may not be the same in the real clinical use of the system because when you are clinically using it, you have time constraints. It is a different way of reading.

So I just wanted to point out that there are some differences between doing a sequential reading versus a washout time period. But I think for the

purpose of being least burdensome without having a huge effect on the comparison, I still think that the sequential design is fine.

CHAIRMAN GLASSMAN: Dr. Berry?

DR. BERRY: I don't think -- I mean I agree with you, bottom line, but I don't think it is least burdensome. I think, as Dr. Spindell points out, it is the relevant study.

I don't understand what you said, Dr. Sahiner, about the timing or the fact that somebody who is in a clinical trial or a study is going to do things differently than they do in the clinical practice. I mean, that is something we always face in clinical research, but why the particular bias in this one?

DR. SPINDELL: May I answer?

CHAIRMAN GLASSMAN: Please.

DR. SPINDELL: So one example might be that a reader, for example, when he or she knows that CAD is going to follow and has an unlimited time to search, may search very, very carefully in the image for finding an abnormality; whereas, in the clinical practice, they may not have that luxury to spend that time, and they may

spend less time.

DR. BERRY: Okay, well I maintain that we may not come -- it may not be perfect, but it is as close as we can get; and if that is a major issue, then we should try to resolve it by either educating the physicians or the readers in the clinical trial or getting a large number of sites in the clinical trial so that we understand better this issue. But we should try to do what is relevant, as Dr. Spindell indicates, for actual use in the clinic.

CHAIRMAN GLASSMAN: I don't want to put words in your mouth, but I think you meant it as a comment rather than a substantial criticism.

DR. SAHINER: Exactly. I meant it as a comment. I agree that sequential reading is appropriate.

CHAIRMAN GLASSMAN: For non-sequential reading, what about the necessity of a washout period and if so, how long should it be? We heard 30 days discussed a little while ago as kind of a standard.

First, is it necessary; and second, how long a time?

DR. ABBEY: I think it is another difficult

issue. I'm informed by Harold Kundel who says, if it is an interesting case, I remember it my whole life. So probably 30 days is an acceptable standard. It is probably not perfect, but it probably fits in with what is reasonable.

CHAIRMAN GLASSMAN: The problem is Hal Kundel never forgot anything, interesting or not.

Dr. Berry?

DR. BERRY: So if we are trying to get somebody who doesn't remember, why don't we get somebody who wasn't there? Why do we insist, if there is going to be a subsequent read with CAD, that it be the same individual, the same reader?

CHAIRMAN GLASSMAN: Dr. Tourassi?

DR. TOURASSI: Because it is going to be very difficult to control experience levels, behavioral aspects of the different observers. It's dangerous to change the readers because all of these studies have a fairly limited number of readers, and there is so much variability among them in the behavioral aspects beyond the expertise and all that.

CHAIRMAN GLASSMAN: Any other comments? We

have obviously stayed on the first part, but that's good because nobody wants to talk about the second part yet.

Any other comments about the first part? Washout is 30 days. A reasonable number? Anyone think it is not? Dr. Garra?

DR. GARRA: I think it is reasonable given that it is being used, and at least there is some validity there.

I wonder if, actually, you could shorten it if the person is seeing a very large volume of cases. I think it's the number of cases you see between the time you saw the case in interest and the time you have to reread it. But I don't know of any studies that prove that. So I would stay with the one month until we can demonstrate an alternative.

CHAIRMAN GLASSMAN: Any conflicting comments about 30 days or the need for washout when you have a paired design study like that?

Then any suggestions for improving paired design studies? Dr. Dodd?

DR. DODD: I just have one question. Is it implicit that the ordering of the reads would be

randomized? I would certainly hope so.

CHAIRMAN GLASSMAN: I would assume so. That would be sort of good practice or best practice. Any other?

DR. BERRY: So just to be clear, you are not talking about the sequential reading?

CHAIRMAN GLASSMAN: No. This is--

DR. BERRY: The sequential reading would always be the reader unaided and then the reader aided.

CHAIRMAN GLASSMAN: Right. This is non-sequential. It's the other paradigm, okay? Can I try to -- this one, hopefully, will be easy.

When there is not sequential reading and the paired design is used, there should be a washout period; and the 30 days is standard and supported by the Panel; and we do not have any suggestions for improving paired design studies for reader CAD.

Is that a sufficient answer for the agency, Ms. Brogdon?

MS. BROGDON: We are deliberating.

CHAIRMAN GLASSMAN: Thank you. Is there something you want us to get to that we haven't gotten

to?

MS. BROGDON: I think Dr. Gwise has a clarification that he would like to ask for.

CHAIRMAN GLASSMAN: Thank you. Dr. Gwise?

DR. GWISE: Yes, the specific issue has to do with the control in the sequential read.

This is the unaided modality, and we have the scoring here. And this is the second modality. We are comparing the two modalities. The thing to notice is the first modality is also part of the second modality.

Now the question is do you see any possibilities for bias here?

CHAIRMAN GLASSMAN: I take it this is a sequential read situation?

DR. GWISE: This is -- yes, this is a sequential read. So the comparison is between the unassisted read to the total.

DR. CARRINO: I think you are implying that because the person -- the radiologist knows that there is a fail safe, that something is coming along, that they may not be as diligent in the unassisted scoring?

DR. GWISE: That is part of it. Are there

any other possibilities for bias in this situation? That's the whole point. The washout period here is essentially zero.

DR. CARRINO: We are just asking how lazy radiologists really are?

CHAIRMAN GLASSMAN: We don't want to go there, trust me on that. On the other hand, this is clinical practice. I mean, this is exactly the way this is done.

DR. GWISE: That's not -- the point is having the two modalities separate, comparing it to an unassisted read, which is one modality -- comparing the unassisted read to this and having the two separate.

CHAIRMAN GLASSMAN: Dr. Abbey?

DR. ABBEY: I think the question is, is it clinical practice when you don't have a CAD algorithm coming next to help you out, and that is the fear of bias entering in. I think that is what Dr. Sahiner tried to address and said that there may be some there. It's hard for us to assess that.

So the issue then would be you do one read with no CAD, have a long washout, go through the same

process, do another read, and then add the CAD in at the end. That would be the control assuming that the washout process was good. That would be the most appropriate thing to do.

So then the question is, is it burdensome? The burdensome rests on whether that bias is substantial or not, and we are struggling with whether there is evidence of bias.

DR. TOURASSI: I think I know where this is going. We are supposed to support the randomized paradigm that was in the document, which was really, really well designed, where basically the cases are randomized and the radiologist doesn't know if the CAD support will follow or not.

DR. GWISE: Actually, we are just looking for your opinion.

DR. TOURASSI: And that would be least burdensome.

DR. BERRY: That's good. That's what you should do, Dr. Gwise, is you know, either it's preordained and you open an envelop, or you flip a coin and you get the CAD or not. If you get "not," you can't

do the CAD.

CHAIRMAN GLASSMAN: I guess, though, the question is then the issue of burden and the issue of how significant is the bias. I think we would all agree that the way you have described the test, it would be superior to a simple sequential read. The question is how superior and how burdensome.

DR. TOURASSI: Exactly. When I first answered this question, I said I read some beautiful paradigms in the document, but I still wonder -- they don't fall under the least burdensome approach because that would require the same case to be read twice with these two paradigms, the unassisted followed by CAD.

CHAIRMAN GLASSMAN: I guess the question is -- or Dr. Spindell, and then --

DR. SPINDELL: I guess my concern here is before the agency was concerned about people not following package insert instructions and using concurrent reads. If we do it the other way, we are not really mimicking what is going to end up happening in clinical practice when people -- radiologists will know whether they have the CAD or not because they bought the

CAD.

So if you are doing that study, you are not really using it in the intended use or what the affect is going to be on the population in the results.

CHAIRMAN GLASSMAN: What do we, the statisticians here on the Panel, think the amount of bias in the sequential read with the immediate appearance of CAD is? How significant is that as a bias?

DR. BERRY: So we are statisticians. Where is the data? I mean, I have seen circumstances like this where -- not exactly this -- where the bias could go in either direction.

You know, you may have a reader who is really interested in making sure that he or she is not wrong, and so diligently makes sure that every base and every point is covered. You may have another one who says well, you know, CAD is coming along. So it could go in either direction, but we don't really have any data.

A trial like this -- I mean, I appreciate Dr. Spindell's point, but we can have both. Unfortunately, it means it is more burdensome in terms of the company, but we could do something -- you know, one

of my favorite things is be adaptive. And so you start out looking at the randomization and then morph into something, which is once you have established something about the bias that you morph into something which is a confirmatory aspect where you are just doing the sequential thing.

So the short answer is that, of course, we don't know, but it could be huge, and it could be in either direction.

CHAIRMAN GLASSMAN: Would it -- let me just ask -- or Dr. Ziskin?

DR. ZISKIN: I think that knowing that the CAD is going to be used will definitely bias, but I think it is biased in the right direction because this is the way that it is going to be performed.

The radiologist will know ahead of time whether they are going to use the CAD afterwards or not. So I think that it is the more practiced way of doing things, and my suggestion is to leave it alone.

DR. BERRY: So Marvin, the problem is that in the clinical practice you detect something, and there is a process, and you say yes or no; and nobody knows how

you got it. But in a clinical study you are going to have to write down, you know, this is what I did by myself and this is what I did with the CAD. That is very different.

There are some people that, you know, just don't want to be wrong and don't want to be corrected.

CHAIRMAN GLASSMAN: You asked for data, and I think it is a very interesting question. It certainly is obtainable data. I don't know how large a dataset you would need to do a non-industry study to compare these two reader paradigms for, let's say, mammography CAD or lung CAD, but certainly mammography CAD because there are any number of cases.

Would this be something that would be worth doing to put this issue to rest? And if, in fact, the performance is dramatically different, if the agency had the authority to require that kind of testing, there would at least be some science behind it.

On the other hand, if it was a very minor difference, then least burdensome becomes in play.

Dr. D'Orsi?

DR. D'ORSI: Would perhaps setting a time

limit for viewing on this help, sort of getting an average time? I know in mammo, there is a lot of data that shows you read them in about 40-45 seconds; and you could easily just have the picture blank out in 45 seconds, and then have them read with the CAD. I don't know if that would address some of the issues.

CHAIRMAN GLASSMAN: Dr. Garra?

DR. GARRA: That would address half the bias; the guy that is going to be over-thorough because he is competing with the system.

What I would suggest is that a subset be done because you don't know which direction the bias is going to be in in any given observer. It could be in favor of the CAD system or it could be against the CAD system -- that we will suggest to the FDA that they may want to require a subset of data for each observer to be randomized like that.

So sometimes they get the CAD; sometimes they don't, but not all of it -- not all the data would have to be that way, just enough to establish within approximate parameters what the bias of that observer is, and then move forward from there and do the correction.

CHAIRMAN GLASSMAN: If you have a number of observers, could you assume that the bias is randomly distributed and it would cancel itself out; and therefore, you wouldn't need to do that? Dr. Abbey is shaking his head vehemently, no. Okay.

Ms. Brogdon, please.

MS. BROGDON: If you want to move on to the next question, we wouldn't be opposed.

CHAIRMAN GLASSMAN: Okay, because it is already tomorrow morning.

Let's just say that on G3 the committee could come to no conclusion.

G4: What are appropriate control groups for reader performance testing? Dr. Dodd?

DR. DODD: I don't know that I was raising my hand.

CHAIRMAN GLASSMAN: You weren't.

DR. DODD: Are we talking about the standalone performance? I think we have already addressed that.

CHAIRMAN GLASSMAN: Reader performance control groups.

DR. DODD: Right. But if we are talking about standalone performance, I think we have just given a lot of discussion to what the appropriate control groups should be. Are we talking about the concurrent paradigm?

CHAIRMAN GLASSMAN: The question is sort of open-ended.

DR. DODD: Okay. I think, with regard to the concurrent reader paradigm, an appropriate -- Actually, I'm going to pass for a moment because I need to give a little more thought to this, sorry.

CHAIRMAN GLASSMAN: Thank you. Dr. Abbey, do you want to take a shot at it?

DR. ABBEY: So I think we are in the same position we were on the last question, actually. It is a very difficult thing to do where you are trying to balance burden against statistical appropriateness in the exclusion of bias.

So there is a multi-reader, multi-case statistical formalism that accounts for all of this. It is oftentimes very hard to do. If you can make assumptions about either effect size or biases, you can

simplify that considerably; and the question is probably best answered by data. Are these safe assumptions to make?

The truth is that some of that probably is in the literature, but some of it is probably still unknown.

CHAIRMAN GLASSMAN: Dr. Berry?

DR. BERRY: What is appropriate is what is appropriate. The control you want to use is whatever you would do without the CAD, and how do you do that? Is it going to be a parallel design? Is it going to be -- parallel designs tend to be shunned for appropriate reasons that they tend to be huge.

The appropriate control in a reader setting is the unaided. I mean, that is what we have been talking about, and I think that is least burdensome, unaided versus with the CAD.

Do we do the randomization? Would that be more appropriate, you know, in line with what Dr. Tourassi was talking about before? That would be better, but it does lead to bigger sample sizes.

In the setting where you are doing

concurrent read, the question is what would you do if you weren't doing concurrent? And you would do presumably, just unaided. So there, the appropriate control is unaided versus both. That has to be -- unless you can think of some sort of washout thing. That has to be a parallel design, and the sample size is correspondingly big.

CHAIRMAN GLASSMAN: Dr. Sahiner?

DR. SAHINER: I think one of the things that might be asked in this question is, for example, is it appropriate to ask a user, a radiologist, to read the image again without any CAD to see if there is any difference, you know, when they look at the image once and do their detection and then they are told to do it again, just like they would do in CAD but without any aid, and would that make any difference?

Another control could be compare CAD results reading with CAD to double reading. Would that be -- I think it depends on, as you said, what the device is intended for.

If the device is intended for doing something similar to double reading, then CAD results

could be compared to double reading results.

CHAIRMAN GLASSMAN: Are we simply dealing here with kind of the answer to G3 all over again which is, you know, a paired design with the reader reading in CAD only and being his or her own control 30 days later by reading the opposite way, either with CAD or without CAD?

DR. BERRY: I think the answer is yes. I mean, that is an answer. I don't know whether I would like that if I were on a Panel reviewing something because I hear Dr. D'Orsi suggesting that he forgets things because of his age. But I'll bet he remembers things like the person you were talking about earlier, and I don't know whether I would think that that was unbiased. Exactly what can a person remember?

I remember, you know, minutia; whereas, I forget all kinds of things. A parallel design doesn't have that characteristic, but it does have other characteristics. It loses the pairing.

A cross-over design, which is what we are talking about here, would be nice except that there is bound to be some sort of residual from the previous read.

CHAIRMAN GLASSMAN: Is that residual so important that it invalidates the results? I guess that is -- in the least burdensome range, that becomes the question.

DR. BERRY: So short answer is we don't know. It could be. The problem is we don't have any way of measuring it.

DR. ABBEY: You actually probably could do an experiment where you tested performance when you know CAD is coming next versus performance when you know CAD is not coming. You could design an experiment to assess bias in that, and it may be that that stuff is out there.

So I would just suggest that our recommendation to the FDA be that the most appropriate thing to do is, I think, the parallel studies you suggested. However, if there is evidence that this bias is small, then it may be acceptable to use a washout period.

DR. BERRY: So there is a long history of this in drugs where people thought that cross-over designs would be wonderful, that they would establish the pair, and what they found was that there was this sort of

residual effect, and they couldn't get rid of it.

They said, okay, so here is what we'll do. We will build a trial and, if it turns out that there is bias, then we will use only the first period so that half of them get the one first and the other half get the other first. So we will forget about the second period.

It turned out that they were forgetting about the second period most of the time, and so cross-over designs are now in the drug world pretty well not used because of these problems.

CHAIRMAN GLASSMAN: Yes, Dr. Dodd?

DR. DODD: I was just going to add to Dr. Abbey's point, that you could potentially embed that kind of test into your study by throwing in some percentage of sham CADs to evaluate whether there is an effect. That introduces other problems, but --

CHAIRMAN GLASSMAN: I think I don't have a summary.

DR. TOURASSI: It seems to me that we are complicating the issue too much. The question is what should be the control group. What Dr. Berry said in the beginning, the control should be whatever is the current

practice. Is it single reading? Is it double reading? Whatever that is, that should be the control group.

Now, how the study will be -- we are not going back to whatever point before. This is just what should be the control group? How the study will be designed, we addressed it before, hopefully.

DR. CARRINO: Yes, I agree. I think that has to be the way radiologists practicing with the unaided reading because we are having all this consternation about trying to get around it, but it is probably not valid. We are guessing. People are throwing in, I don't know, whatever, to try and get around it. But the bottom line is that that is probably the standard for what we know right now.

CHAIRMAN GLASSMAN: So would anybody object to the recommendation being just what Dr. Berry said. That is, the unaided normal conduct of the radiologist as the control group to the radiologist with CAD?

DR. TOURASSI: My only comment here is that we need to leave it more general, whatever is the standard practice. If the standard practice for whatever application is double reading or triple reading, whatever

is the standard practice.

CHAIRMAN GLASSMAN: The standard practice without CAD versus the standard practice with CAD.

DR. LIN: I think Dr. D'Orsi's point about time limitations is actually a good point. That might be a way to address this issue. This concern that we are having with people not behaving like what they normally would when they single read these films.

So, for example, somebody who takes 60 seconds to normally take a mammogram. We need to make sure that when they are in the study that they are not taking five minutes because they know that CAD is coming, and they don't want to look bad compared with the CAD.

So I just wanted to throw that out.

CHAIRMAN GLASSMAN: Does the rest of the Panel think that that is an important -- Dr. Spindell?

DR. SPINDELL: I also think if that is the case, we can't let them take five seconds either. I mean if they need to take 60 seconds, they need to take 60 seconds. If they take five seconds because they know they have CAD coming, that is going to bias the results also.

CHAIRMAN GLASSMAN: Dr. Garra?

DR. GARRA: Can I suggest that we instruct them carefully to take the normal amount of time and record the times?

CHAIRMAN GLASSMAN: Dr. Rosenberg, and then I want to try to sum up on this one.

DR. ROSENBERG: I think it is important, but I worry also that who the control group is, which radiologists, might even be a bigger issue. We haven't even talked about that.

CHAIRMAN GLASSMAN: Okay. Let me summarize. I like that. I like that.

The appropriate control group is a group of radiologists who practice as the standard of care without CAD of a professional background similar in experience to the group that reads with CAD. If they are all academic radiologists or all private practitioners or a mix, that that mix should be reflected in the other group.

Is that acceptable to the Panel? Is that acceptable? Does that answer the question for the agency, Ms. Brogdon?

MS. BROGDON: Yes, it does. Thank you.

CHAIRMAN GLASSMAN: Thank you. Okay. G5 is a long question, and we are getting close.

Please comment on the appropriateness of using a standardized weighted analysis as a primary or secondary analysis of a CAD study. The standardized analysis weights observations according to a standard distribution for important clinical variables thought to be representative of the target population.

Dr. Dodd?

DR. DODD: Since you are looking at me, I thought we had addressed this yesterday when Dr. Berry was suggesting that for reader studies we do some enrichment in terms of the disease status but we don't control the proportion and micro-manage the proportion of the different subtypes of interest within that.

So if that is the case, I don't think, particularly for a primary analysis that you need to do a weighted analysis. For a secondary analysis, I think it's fine because it is a secondary analysis.

CHAIRMAN GLASSMAN: Dr. Berry, does that accurately reflect what you said yesterday? And do you think that is the correct answer now?

DR. BERRY: Yes, I think that is the correct answer now. But I would distinguish between standalone and the reader; and in the standalone setting, which again I regard to be sort of proof of concept, then you could do the kind of analysis where you include extra trauma patients and include extra challenging patients and do some sort of a weighted analysis. But I don't regard that as being the definitive analysis that is going to get your product cleared.

CHAIRMAN GLASSMAN: Dr. Abbey, any comment?

DR. ABBEY: I guess I'm just back to the issue of burden again. If it is difficult to get enough cases in one category and you can weight it by knowing that, okay. We didn't have that many cases, but we will expand the effect -- in other words, we will give this a higher weight because we know it is important -- or we are able to weight it and then approximate that. I guess the concern is that do you ever really know what that standardized distribution should be? I wonder if it isn't easier.

CHAIRMAN GLASSMAN: Dr. Tourassi, a comment?

DR. BERRY: So if you are going to do, let's

say, a confirmatory study where you are weighing according to the challenging cases and you are trying to understand how many challenging cases are there in the population we get -- what we want to do is have a representative case mix, and the question is what is a representative case mix, and that would establish the weights?

It is an extra level of complication. It's hard, and moreover, it is kind of black box-ish. So when a company is presenting to a Panel like we are, the Panel wants to see things that are quite simple, you know, that are here is specificity and this is the way we calculate it; here is sensitivity, this is the way we calculate it and, oh, by the way, we used this weighting that was based on such and such.

So then the challenge becomes what about -- you know, this alphabet soup, BCSC or whatever, did it have the right -- it raises a whole other set of problems. So keep it simple.

DR. TOURASSI: Yes, I agree. That was my first impression. I was confused a little bit by your comment. It looks like an interesting mathematical

exercise, but it doesn't seem to serve the end user. That information is going to be meaningless to radiologists.

I think the substrata analysis that we discussed before, even for the standalone performance, is going to be much more useful to the radiologist to know explicitly this is the sensitivity and specificity for these type of cases or the other type of cases, much more than having one performance index according to some weighted distribution, which is totally theoretical.

CHAIRMAN GLASSMAN: Dr. D'Orsi?

DR. D'ORSI: Also I think the weighting might actually in some instances skew against what the results in a patient might be. So I think that is another negative effect that could happen.

For example, a small cluster of micro-calcifications, you can argue, if I miss it, it is probably DCIS. I'll pick it up next year, but you would get a higher weighting because it is a more subtle lesion.

So I think I agree that that is going to bring up a lot of problems.

CHAIRMAN GLASSMAN: Dr. Dodd?

DR. DODD: I also want to emphasize, I am not clear how much more burdensome it is to do it the way we are proposing because in order to do this, you also rely on having good estimates of what the effect is in the substrata. So by the time you do all of that, you may, in effect, have almost as many cases as you would have if you just did sort of a more random sample.

CHAIRMAN GLASSMAN: Dr. Abbey?

DR. ABBEY: So my concern is you have collected 1,000 cases, and for some bizarre reason you only have 10 micro-calcification cases in there. So you can't do the study even though you have collected 1,000 images. You've got to keep on going and get more until you get enough cases to build that representative distribution up.

So whereas with a weighted analysis, you would say, well, we only got 10, we should have had 50; so we will just weight the performance in that strata by a factor of five. So the weighting should be incidence- or prevalence-based, not based on subtlety or something like that.

CHAIRMAN GLASSMAN: Dr. Berry?

DR. BERRY: So the problem is that you didn't, in fact, get a representative sample, right? If you get a representative sample, you are bound to see some unusual things associated with it, but that is the luck of the draw. You run the study on the sample.

If the sample was really taken as representative in some practice, then even you got only 10 micro-calcifications and you should have gotten more, there was no particular bias in the way you got it. In fact, you don't know that the particular population that you are sampling -- breast cancer, for example, is changing incredibly over time for a number of reasons, and it is conceivable that the 10 is the right number.

The onus on the company should be to get something that is representative of the cancer and, if you did that and if you've done a quality job of getting that representative sample, then you run the CAD and the study on the basis of that, and the FDA should accept it.

CHAIRMAN GLASSMAN: So if I can try to summarize, and I may get this 180 degrees wrong because I got lost about 10 minutes ago. Let me try.

The Panel believes that there are many problems with weighted analysis and that the FDA should -- and the Panel is not in favor of weighted analysis as a statistical test and that the hope would be that the initial group of patients would be properly accrued so that this would not be necessary.

I open it up for comments by the Panel.

DR. BERRY: I like what you said.

CHAIRMAN GLASSMAN: Dr. Abbey, does that fly for you or would you like --?

DR. ABBEY: I guess, as long as we acknowledge that may be somewhat more burdensome in some cases.

CHAIRMAN GLASSMAN: And it may be more burdensome in some cases. Is that an answer that the agency can use?

MS. BROGDON: Yes, thank you.

CHAIRMAN GLASSMAN: Okay, F1. We are getting close. Oh, we are doing great. Don't let me down.

Future Issues with CAD -- we have focused thus far on devices that are used primarily for computer

aided detection. Do you have comments on the types of testing needed for computer aided diagnostic devices, compared to the types of testing that we have talked for computer aided detection?

I think we have said that the bar would have to be higher, but how high and in what way potentially for diagnosis, which in fact takes some of the radiologist function away, is the way I see diagnosis.

I mean, the radiologist will still ultimately have the final say if we are talking about a radiology CAD, but that there will be a lot more data which may help or not the radiologist in terms of probability, of say cancer, if we are dealing with looking at nodules in one place or another.

Should the testing be more rigorous, and if so, how? Should the endpoints be different? Any comments from anyone? Oh, Dr. Leitch?

DR. LEITCH: Well, for the tests where you do have a lexicon of diagnoses, you can hold the device to that standard. So you could say, on its BI-RADS category predictions, do they come out as the proportion of cancer cases as one would expect for that BI-RADS

category? And hold it to that standard because that is a standard you hold the radiologist to.

So you can -- that can be the standard that you have where you have the lexicon that's pretty easy like that. Where you don't, it can then be more complex and may come down to kind of making it perform a little better than average.

CHAIRMAN GLASSMAN: If the device doesn't give a definitive answer but a likelihood which is probably more likely than a binary, you know, BI-RADS-2/BI-RADS-3 kind of thing where it is a .7 probability of carcinoma as opposed to .3 for another lesion, it would be harder to do that.

How do we assess that performance? Is there a statistical way? Do we compare it to an expert panel who has looked at the same case? Do we require pathology gold standard proof for the first couple of things that come along? Dr. D'Orsi?

DR. D'ORSI: We do have data on the BI-RADS separations, 3, 4A, 4B, 4C, and 5 with the percentages of cancer in each one of those. So you have a database that is in the literature. It is young yet, but it is in the

literature.

So you have some numbers to compare with. What you have to do, I think, is do what you do in clinical practices and compare PPVs with and without this device and see how often it hits cancer. And then compare it to the levels of concern that the operator puts on via BI-RADS.

Now BI-RADS is ordinal, but it is not equal; and that can skew ROC analysis a little bit, but at least it is ordinal.

The other thing with diagnostics -- and I'm just talking about mammo; I think mammo is kind of unique in this particular sphere -- you are going to need a very, very large database to have this system work with any degree of confidence for the user, and that doesn't only mean cancers. It means cancers in various-- it varies with various features, and the robustness of that dataset is directly related to how this CADx is going to perform.

So I think we have almost everything in place to test it, and it is just going to be getting enough data to produce a robust algorithm.

CHAIRMAN GLASSMAN: Of course, in the more general sense, though, we have not only breast, but we have lung. We have colon. We have things we haven't talked about yet.

So in a general sense, what kind of testing do you think we need? Dr. Garra?

DR. GARRA: Yes. I think you hit the nail on the head, that we are going to need pathology correlation more often than we require for just detection especially for lung nodules.

A consensus panel determining whether a nodule is there or not is one thing, but if you are trying to make a diagnosis, you are going to have to have pathology more often, maybe not in every case, but in a lot of them.

CHAIRMAN GLASSMAN: Pathology or reasonable term clinical follow-up.

DR. GARRA: Yes, it depends on the indication or what they are claiming that they can do. If they are claiming that they can diagnosis histoplasmosis reliably, then you are going to have to get proof on some of the benign lesions as well.

CHAIRMAN GLASSMAN: Yes, Dr. D'Orsi?

DR. D'ORSI: Do lung and colon have the robustness of features indicating benign/malignant that are present in mammograms; and if not, what are we getting a CADx for then? Is it only size, presence of calcification, flatness of the polyp, size of the polyp? Is that about it for CADx, or not? Is there more?

CHAIRMAN GLASSMAN: Dr. Garra?

DR. GARRA: I'm not sure I can speak to the polyps. I think we saw some evidence earlier today that flat polyps are at higher risk, but I don't think the answer is in on that.

For lung cancer, I think there are a lot of robust features. They are actually very similar to the ones for breast, so irregular margins, you know, architectural distortion, all those things and, of course, lymphadenopathy.

DR. D'ORSI: So you could use a paradigm similar for chest that you do for lung? I mean for lung and breast.

DR. GARRA: Yes, I think so.

DR. D'ORSI: It's getting late.

CHAIRMAN GLASSMAN: Any other comments about this? Yes, Dr. Leitch?

DR. LEITCH: So for colon, though, again this thing because you have a pre-malignant thing; actually I think it might be easier in colon because the lesion is not so weird. It is kind of a pretty simple lesion because you do want to find a benign polyp. That is what you want to find. The size discrimination is more of the thing on that.

So that is -- I don't think that is actually going to be as complex as the ones where there is more variety to the lesion, particularly if the lexicon is not worked out for -- I don't know; in lung, could you say there is a lexicon that says if it looks like this, the probability of cancer is thus and such?

Part of it is what do you get with the radiologist right now? You know, what are the standards? What do you get with the radiologist? That is how the -- that is going to be what the device has to perform to.

So if the radiologist isn't 100 percent, so you don't expect the device to be. But it's got to fall pretty close to it to replace the radiologist.

DR. WONG: I'm not too sure that you are going to add that much more to the colon. If you identify the polyp, obviously the size of the polyp is going to kind of determine as to whether it is going to be a malignant or a tubulovillous type of polyp. Then if it is an annular lesion, it is more likely to be a cancer.

There is a fair amount of data already because there are so many centers that have been using CTC that it is conceivable that there may be characteristics about the cuts through these masses that may identify characteristics that could tell you maybe that there is some histologic pattern that you can see from these CTCs. But again, that would be purely conjectural.

CHAIRMAN GLASSMAN: Obviously, the manufacturers would have to go ahead and provide that data as part of their submission, I think.

So far, and I will come to it in just a second. So far what I have heard is from a sort of generalized view that the major difference so far we have come up with would be the need for pathologic correlation

in a much higher percentage of cases if we are talking about diagnosis than if we are talking about detection.

There would be some other surrogates such as, you know, time of stability or negative PET scan, but that an enriched dataset with a number of cancers would probably need a fair amount of pathologic proof.

DR. CARRINO: I think my general statement would be that I would take it to another level. The computer assisted diagnosis -- I would look at it like a prediction model or these decision rules that are being used, and the methodology for validating those is a lot more -- more rigorous and more intensive; and I think that is way beyond the time that we have to talk about it.

It's definitely, I think, a separate topic, and I think it is going to be much more rigorous, and I would look to those things that are done for like decision models, prediction models, decision rules.

DR. KIM: I think the colon is really a special case because it is relatively easy to get pathology.

CHAIRMAN GLASSMAN: Right, and breast too,

relatively easy although the ladies in the group may not agree.

DR. KIM: And I think the issue is a little bit different in the colon in the sense that the diagnosis is between a true polyp and a pseudo-polyp related to stool. That's the diagnosis you are trying to make. Is it truly a soft tissue polyp?

CHAIRMAN GLASSMAN: So the colon doesn't particularly lend itself probably to computer aided diagnosis so much as computer aided detection, but there are other areas in the body that will?

DR. KIM: Right, and they are definitely intertwined, but I would say that the diagnosis is that, yes.

CHAIRMAN GLASSMAN: Okay. So we've got greater emphasis on pathology and follow-up, and I think to look to the examples of clinical guidelines and pathways for things like, if I can put it in -- nobody has mentioned yet, head trauma, ankle trauma and the way that they were developed to look to the proper way to evaluate diagnosis, knowing that it will be a much more rigorous bar for the companies rather than detection.

I think that is about where we are. Does that answer the question or is there something more specific that you need?

MS. BROGDON: I don't think we had anything more specific that we needed.

CHAIRMAN GLASSMAN: Okay, F2. Emerging CAD areas that the FDA should be aware of? Comments on the types of testing needed for other CAD devices, present and future, compared to the testing we have discussed.

First, let's deal with what is coming down the pike that any of us have heard of, have friends who are dealing with, other organ systems, prostate or things for CAD?

DR. LEITCH: Well, I wouldn't say, quote, I've "heard" of this exactly, but I think any imaging test that has functional parts to it, uptake of materials into lesions, that sort of thing, how it washes out -- those are tests that lend themselves more, I think, to computer interpretation because there can be numbers that are attached to that and then probabilities can be attached to those numbers that those lesions will be malignant.

It is actually harder for the radiologist to "work through" all that, and there are many images involved in that. So studies where there are a lot of images that are taken and you have a functional component to it; that's the type of test that would lend itself to computer diagnosis, I think.

CHAIRMAN GLASSMAN: I'm sure you are thinking of breast CAD.

DR. LEITCH: So MRI, PET.

CHAIRMAN GLASSMAN: Yes, certainly, the breast MRI CAD, two of which have been approved by the agency are primarily functional analysis tools with some detection capability. Dr. Abbey?

DR. ABBEY: The one thing I have heard coming down the pike and may be further along than I know is the use of CAD, both CADe and CADx, to exclude from reading very easy cases so cases that are primarily non-dense, fatty or fatty replace breasts that are not to be then read by human eyes. The CAD algorithm would actually exclude them from the reading.

CHAIRMAN GLASSMAN: Any other sort of secret information out there?

DR. LIN: One suggestion might be a four-phase abdominal CT scan for liver cancer screening because that is a dynamic study. It seems to lend itself to, you know, what you were talking about just now with computer assisted detection because liver cancer screening is becoming more and more important as well.

CHAIRMAN GLASSMAN: Dr. D'Orsi, you had a comment?

DR. D'ORSI: There are people thinking now with tomosynthesis CAD as well even though the technology hasn't been tested yet, but that is what people are speaking about. So that is something else.

One other point, I think CAD for MRI is more an intelligent work station than CAD. It only gives you the DX for one minor mode which is not -- it is not a real DX. It is an intelligent work station, I think.

CHAIRMAN GLASSMAN: Okay, any other future -- yes, Dr. Wong?

DR. WONG: You know another area that we've found with CTC is that you find an equal number of extra colonic tumors. We are finding that to be a major area of interest. Obviously, you've got a CAD for

intraluminal problems, but that doesn't really take into account the extra colonic aspect or the intra-abdominal.

So that might be another area that CAD could look at. It is, obviously, very complex, because you got a lot of organs in the abdomen.

CHAIRMAN GLASSMAN: We call those radiologists where I work.

DR. LEITCH: Ultrasound, breast in particular.

DR. SAHINER: Two of the areas that I am aware of are upper urinary tract cancer on CT urography and pulmonary embolism detection, and also I don't know if it qualifies as CAD or not but any type of temporal analysis, change analysis over time.

CHAIRMAN GLASSMAN: That may have more to do with the next question.

Are we at this point prepared to suggest any different testing methods for these other conditions or would that be too preliminary until we know exactly what they are designed to do? Nobody wants to come up with one? Okay.

For F2, emerging areas of CAD: studies that

have kinetic significance, studies to check easy cases in mammography, the fatty breast cases, liver cancer, tomosynthesis, breast ultrasound -- I think Dr. Giger showed us a little bit of that -- renal carcinoma, pulmonary embolism, and things that go swish in the night, things with temporal change where computers are very good at measuring changes in wash-in and wash-out are the kinds of future areas that we see potentially coming down the pike.

I think, until they get here, the statistical analysis is beyond what we are willing to take on.

Is that good to the Panel? And is that good to the Agency?

MS. BROGDON: That is helpful, thank you.

CHAIRMAN GLASSMAN: Okay,F3. Your cars are outside. So think about that.

Comments on the levels of testing for the different types of computer based technologies compared to testing that we have discussed.

So here we are talking about the kind of near-CAD and borderline CAD, the measurement packages

with intelligent design to them or intelligent software, and I guess I can start by simply saying would phantom testing be enough if it is limited to sort of intelligent measurement; in other words, making a judgment about what the structure is and then measuring it, or would you need reader testing? Would you need just standalone testing of clinical cases?

Where might you draw the line for things that are short of detection but long on other things, measurements? Anyone?

DR. ZISKIN: Well, it would depend upon the nature of the output. If it were a quantitative number, it would only require a phantom or a test object. But if it required a person to visualize something, then you have to test the performer.

CHAIRMAN GLASSMAN: Okay, any other? Yes, Dr. Rosenberg?

DR. ROSENBERG: If we were talking about, for instance polyp testing where sizes matter, would phantom testing be adequate or would you actually have to compare it to pathology?

CHAIRMAN GLASSMAN: I would think pathology

in that instance because it is critical to the next step, and it is also relatively simple to do.

Any other comments? We almost don't have anything here. Dr. Garra?

DR. GARRA: There are a couple of image processing capabilities that are on the horizon, various forms of speckle reduction that will require observer studies, and to show that it improves the observer's perception of pathology.

So there is definitely a whole raft of those waiting out on the horizon. They largely come from the spy satellite program.

CHAIRMAN GLASSMAN: Hopefully, not the one we had to shoot down.

Any other comments?

DR. BOURLAND: Comment here. So testing always starts simple and heads to complex. It goes from geometries toward anthropomorphic, and those would be appropriate based on whatever the device is designed to do, whatever its modality energy source is, and whoever the receiver is on the end.

CHAIRMAN GLASSMAN: So I think if I can

summarize, there are many different types of near-CAD technologies coming down the road, some of which based on their improvement of image quality, will need human eye testing.

Others may need just phantom testing or standalone testing, and it would be sort of device and function specific.

I know that is very general, but is it sufficient?

MS. BROGDON: Yes, thank you.

CHAIRMAN GLASSMAN: Let me then thank the Panel and those of you from the FDA and others who have stayed through these two days.

I think the Panel has done a fantastic job on a very different kind of open-ended structure, and hopefully when the agency takes the next 12 or 15 hours to think about what we did, will agree. But anyway, you all have my thanks.

I formally adjourn this session.

(Whereupon, the foregoing matter went off the record at 5:03 p.m.)

- - -