Formatted: Right: 0.25"

patient and cancer characteristics in a screening population. The relevant characteristics are cancer size, breast density, finding type, histologic type, and palpability. And these are very important, as Dr. Petrick and Dr. Gwise discussed earlier, when designing studies. Whether you are doing enriched or non-enriched studies, it's very important to look at these characteristics in the final population of patients.

With regard to cancer size, the approximate distribution of cancer size on screening mammograms is, approximately, as follows: 35 percent less than or equal to 10 millimeters, 60 percent less than or equal to 15 millimeters and 75 percent less than or equal to 20 millimeters. Larger cancers are more readily identified and characterized on mammography.

With regard to breast density, approximately 10 percent of patients have either almost entirely fatty breasts or

Formatted: Line spacing: single

Formatted: Right: 0.25"

extremely dense breasts, and approximately 40 percent of patients have scattered fibroglandular densities and 40 percent have heterogeneously dense breasts. Greater breast densities are associated with lower sensitivity for breast cancer detection and a higher incidence of interval development of breast cancer following a negative mammogram.

With regard to finding type, approximately, 30 to 40 percent will be masses, 30 to 40 percent will be microcalcifications, 10 to 20 percent will be a combination of a mass and microcalcifications and 10 to 20 percent will be architectural distortion or focal asymmetry.

With regard to histologic type, approximately, 70 to 80 percent are invasive cancers and 20 to 30 percent are Ductal Carcinoma In Situ or DCIS.

By definition, patients who undergo screening mammography are asymptomatic. The

Formatted: Line spacing: single

Formatted: Right: 0.25"

retrospective studies have shown that approximately two to five percent of patients who undergo screening mammography actually have symptoms that were unknown at the time of the examination.

There are two types of mammography devices currently on the market: Screen film devices and digital devices. And there are two types of digital devices: The DR and CR. CR is a device that uses a photostimulable phosphor.

There are two standard mammographic projections that are obtained of each breast:: the craniocaudal (CC) view on the left side and the mediolateral (MLO) oblique views on the right side, and these are typically displayed side-by-side right and left.

Breast cancer is detected on the basis of four types of mammographic findings: The characteristic morphology of a mass; the shape and spatial configuration of microcalcifications; distortion of breast

Formatted: Line spacing: single

Formatted: Right: 0.25"

tissue architecture; and asymmetry between the left and right breast.

Mammography is unique among imaging tests as it must be performed and even interpreted in accordance with the Mammography Quality Standards Act or MQSA, but MQSA does not apply to mammography CAD devices.

With regard to interpretation of mammograms, the CC and MLO projections in each breast are considered complementary and necessary for interpretation. Mammography studies are always interpreted by examination of the CC views from each breast in the side-by-side manner and likewise for the MLO views.

When a finding is identified on a single view, whether it is the CC or the MLO, the corresponding region on the complementary view is examined in order to confirm the 3-dimensionality of the finding. Comparison should always be made to prior mammograms when these are available.

With regard to reporting

Formatted: Line spacing: single

Formatted: Right: 0.25"

mammographic examinations, mammographic characteristics and findings are typically recorded according to the American College of Radiology, that's the ACR, Breast Imaging Reporting and Data System, that's the BI-RADS system. BI-RADS is meant to standardize the language and descriptions used in mammography reports.

With regard to reporting location, a finding should always be triangulated so that it's three-dimensional location within the breast is known. The ACR BI-RADS Atlas recommends using a clock face for each breast and divides the breast into anterior, middle and posterior thirds on the CC and MLO views.

The BI-RADS system has final assessment categories. These were developed and standardized into seven categories that correspond to the reporting requirements in the MQSA.

Category 0 means that a patient needs additional imaging evaluation and/or you

Formatted: Line spacing: single

Formatted: Right: 0.25"

need to obtain prior mammograms. Category 1 is definitely negative. Category 2 is definitely benign findings. Category 3 is when a finding is probably benign. An initial short-term follow-up is suggested.

But it's very important to note that according to the ACR BI-RADS Atlas, a finding placed in Category 3 should have less than a two percent chance of malignancy. The BI-RADS Atlas also states that it is inadvisable to render such an assessment that is Category 3 when interpreting a screening examination.

And, this point is important when you are reading literature on reader performance with mammography CAD devices. You have to always consider how many patients are put into this category, and how it is used in the final analysis.

Category 4 is for suspicious abnormalities when biopsy should be considered. These are sometimes broken up

Formatted: Line spacing: single

Formatted: Right: 0.25"

into 4A, B and C, based upon low, intermediate and moderate concern. Lesions put into the 4C Category should be or are expected to be a malignancy.

Category 5 are highly suggestive of malignancy, and these are lesions that should have a probability greater than or equal to 95 percent of being cancer. Category 6 is for lesions that have a known biopsy proof of malignancy.

It's important to look at the performance measures of mammographers. There is great variability in the published literature with regard to sensitivity that ranges from 60 to 100 percent, as well as specificity that ranges from 35 to 98 percent. Using the data in the Breast Cancer Surveillance Consortium, you see a sensitivity of about 79 percent and a specificity of about 90 percent for that very large number of screening mammograms.

Mammographic sensitivity is lowest

Formatted: Line spacing: single

Formatted: Right: 0.25"

in patients with dense breasts and for small masses. Given that greater than 99 percent of patients who undergo screening mammography do not have a cancer, the task of the mammographer is essentially to find a needle in a haystack. And a practicing radiologist may, therefore, perform detection and analysis very rapidly, that is both tasked to perform almost simultaneously when reading mammograms in a clinical setting.

Cancers that are visible on mammograms may draw the radiologist's attention, and they may be dismissed with or without formal description in the radiology report. But, if a finding draws the radiologist's attention and is missed, this does not constitute an error of detection, but is more likely to be considered an error of analysis.

Any device designed to reduce radiologists' errors should obviously focus on the types of cancers that radiologists tend to

Formatted: Line spacing: single

Formatted: Right: 0.25"

miss. But, it also might be beneficial to detect cancers that are missed simply because the radiologist is asleep at the wheel. Either way, increased detection should always be weighed against increased false positives.

Regarding false positive mammograms, these greatly outnumber actual breast cancers found. Approximately, 10 percent of patients who undergo screening mammography will be recalled for diagnostic mammography. At the same time, approximately, 0.4 percent of patients who undergo screening mammography actually have a breast cancer.

Approximately 50 percent of all women had at least one false positive mammogram over 10 years of screening. False positive mammograms can cause increased dose exposure, biopsy, complications associated with biopsy, and unnecessary anxiety.

What about false negative mammograms? Approximately, 20 percent of cancers are missed on screening mammography.

Formatted: Line spacing: single

Formatted: Right: 0.25"

It's important to look at what are these 20 percent of cancers. About 10 percent are thought to actually be visible on the mammogram and, approximately, 10 percent are thought to be not visible.

Therefore, there is obviously, room for improvement to capture approximately up to an additional 10 percent, perhaps more, of cancers that are otherwise visible, but go undetected or misclassified on screening mammography.

Of the 10 percent of missed cancers that are visible on mammography and are missed, approximately five percent are errors of detection, and approximately five percent are errors of analysis. And I would emphasize these are approximations.

Compared with cancers that are not missed, of the 10 percent of cancers that are missed and are visible, most are masses, architectural distortions or focal asymmetries. They tend to be smaller in size,

Formatted: Line spacing: single

Formatted: Right: 0.25"

and they tend to be present in patients with denser breasts.

How can we reduce radiologist errors when interpreting mammograms? One method that has been tried is double reading. Double reading of screening mammograms, that is reading by two radiologists, has been advocated as a way to increase radiologist detection of cancers.

And there are a lot of clinical studies on this that have shown that double reading of mammograms improves radiologist detection by about five to 15 percent. But typically, with an associated increase in recall rate of about five to 10 percent, unless consensus double reading is used with a recall rate that is much lower.

Double reading of screening mammograms can capture both errors of detection and errors of analysis. And the published literature does show that double reading can capture a substantial portion of

Formatted: Line spacing: single

Formatted: Right: 0.25"

the approximate 10 percent of cancers that are visible but currently go undiagnosed.

So where did mammography CAD devices come from? Well, CAD devices have been developed as a potential replacement for a double reader.

The intended use of current commercially available mammography CAD devices is to reduce errors of detection. That is, current commercially available mammography CAD devices attempt to capture the approximately five percent of cancers that are visible but are not detected.

The potential for improved detection, as I noted earlier, should always be weighed against the potential for false positives interpretations by radiologists using CAD devices. And just to give some context, the screening cancer incidence is approximately four per 1,000.

Current commercially available CAD devices place at least, approximately, two

Formatted: Line spacing: single

Formatted: Right: 0.25"

marks per patient, even assuming 100 percent sensitivity for the CAD device and assuming that marks are placed on both the CC and ML view for each cancer. There will still be about 249 false positive marks for every true positive mark.

It's therefore important to measure how easy or difficult it is for radiologists to dismiss these false positive CAD marks. It's also important to measure the effect of false positive marks on potentially distracting the radiologist from other findings that may not be marked.

I just want to review mammography CAD devices that have been approved by the FDA. There are four mammography CAD systems that have been approved through the PMA or Pre-Market Approval application process. The first approval order was issued in 1998. The first-of-the-kind device was the subject of a Radiological Devices Advisory Panel meeting that was held on May 11, 1998, and the

Formatted: Line spacing: single

Formatted: Right: 0.25"

transcript and other information are all available on-line.

All devices were first approved for use with digitized versions of screened filmed mammograms -- excuse me, of screen film mammograms that were obtained for screening purposes.

Further PMA supplements have been approved over time to expand the use of CAD devices to operate on digitized diagnostic screen film mammograms, at least the CC and MLO views, as well as mammograms obtained on full field digital mammography or FFDM devices. PMA supplements have also been approved for modified software versions of CAD algorithms.

The labeling of currently approved mammography CAD devices have an indications for use or IFU that is similar to the following: They are intended to identify and mark regions of interest on routine screening, as well as the CC and MLO views of diagnostic

Formatted: Line spacing: single

Formatted: Right: 0.25"

mammograms, in order to bring these regions of interest to the attention of the radiologist after the initial reading has been completed and to assist the radiologist in minimizing observational oversights by identifying areas on the original mammogram that may warrant a second review.

What about the data and information that was provided for original approval of mammography CAD devices? At the time of original approval in 1998, there was limited experience with use of mammography CAD devices by radiologists in clinical practice.

The data that served as the basis for approval for currently approved mammography CAD devices typically included four basic components: The first component was standalone performance on missed cancers. And missed breast cancers were identified by obtaining prior mammograms from patients with newly diagnosed cancer, that is, patients with interval cancers; and determining if the

Formatted: Line spacing: single

Formatted: Right: 0.25"

cancers were actually visible in retrospect and should have led to a clinical action.

Standalone performance on these missed cancers was considered a surrogate for the ability of the device to detect difficult findings. And it was used to estimate the maximum potential reduction of detection errors if radiologists actually used the device in practice.

The second component was standalone performance on cancers detected at routine screening mammography. And this was considered a measure for the ability of the device to detect more obvious and intermediate level of difficulty findings.

The third component was standalone performance on normal screening mammograms to determine the rate of false positive CAD marks on normal cases. The fourth component was screening exams, and these databases may have or may have not been enriched with some cancers, and these were used to determine the

Formatted: Line spacing: single

potential increase in recall rate resulting from use of the CAD devices.

What has been learned since the original approval of these devices? Well, there is a large body of literature out there on standalone performance of mammography CAD devices. There is also a large body of literature on the subject of reader performance testing of mammography CAD devices where radiologist's performance is measured both without and with use of the CAD device.

And these reader studies have employed two different general designs, some of which has been discussed previously: Retrospective clinical performance testing and prospective clinical performance testing.

The retrospective clinical studies use radiologist interpretations that are not part of actual clinical practice. The prospective studies use radiologist interpretations that are part of an actual clinical practice. And the prospective

Formatted: Right: 0.25"

studies include two basic designs: The sequential design and the historical control design.

With the sequential design, the images are presented to the radiologist without CAD information. It requires interpretation and then presents the same images with CAD markings and allows the radiologist to modify the assessment.

With the historical control design, you compare radiologist performance over a period of time without CAD devices to radiologist performance over a period of time after introduction of CAD devices. And it, obviously, uses different patients.

Here are some of the key points from the published literature which are contained in more detail in the Panel briefing document. Standalone testing has shown very high sensitivity to mark calcifications but much lower sensitivity to mark masses, architectural distortions, or focal

Formatted: Line spacing: single

Formatted: Right: 0.25"

asymmetries. Standalone testing has also shown a false positive mark rate of between two and four marks per patient.

Reader performance testing has shown conflicting results for detection of invasive cancers. Reader performance testing has also shown a trend toward CAD improving radiologist's detection of calcifications, especially DCIS. And reader performance testing has shown an increase in recall rate when using CAD devices in some studies, and some of these are statistically significant increases.

Now, I want to discuss some of the clinical testing issues specific to mammography CAD devices that will be the subject of the questions that we have for the Panel. And I tried to put on the top of the slide the question numbers that correspond to some of the comments and some of the questions in the slides.

With regard to ground truth, this

Formatted: Line spacing: single

Formatted: Right: 0.25"

has been touched on previously, but this is crucial for both standalone and reader performance testing. Ground truth includes whether or not the patient has a breast finding, whether or not the patient has one or more benign or malignant findings, the precise location and extent of each finding on each view, and the BI-RADS descriptors and final assessment of each finding.

Ground truth for cancer is determined by biopsy or surgery. Ground truth for benign findings is determined by biopsy/ surgery or by one year follow-up mammogram. And ground truth for normal is determined by a one year follow-up mammogram.

Ground truth for the location and extent, that is the lesion boundary of a finding, is determined typically by a panel of expert radiologists and can be annotated either manually or digitally on an image.

What about standalone performance testing? Standalone performance is highly

Formatted: Line spacing: single

Formatted: Right: 0.25"

dependent on case selection, including the following factors: the precise mammographic characteristics in the case set, including finding size; the pathologic type of lesions; the number of masses versus microcalcifications; and the distribution of breast densities.

Standalone performance testing also depends on the precise method of ground truth determination for the location and extent of disease; the precise scoring metric, for example, per lesion, per CAD mark, per patient versus per view; and the precise scoring methodology.

For example, using overlap criteria on the actual CAD mark or the actual region identified by the CAD system itself; that is, the algorithm segmentation. And the algorithm segmentation may not be displayed to the user typically.

Standalone performance can be done using a larger database than used for reader

Formatted: Line spacing: single

Formatted: Right: 0.25"

performance testing, and that is certainly a potential advantage. And using a larger database may allow meaningful stratified analysis on clinical mammographic and pathologic subgroups, which may be important information, because it may influence user confidence in the ability of the device to detect findings in each of the important subgroups.

You can also do -- use stratified measures and this may include mammographic finding types, such as masses, microcalcifications, architectural distortions and focal asymmetries, the different pathologic types. You could do it by size. You could do it by breast composition. In particular, it might be important to look at the number of small masses, that is, masses less than 10 millimeters in size, in particular in patients with dense breasts since these are among the most difficult findings to detect.

Formatted: Line spacing: single

Formatted: Right: 0.25"

Both overall and stratified standalone performance can be reported on a per lesion, per view, per breast or per patient basis. And one of the things that we would like input from the Panel is given that clinical actions following mammography are finding-specific, what are the advantages and disadvantages of each of the above reporting measures?

Without standardized methodologies for case selection, ground truth, scoring metric, scoring methodology and reporting, it's important to keep in mind that it may be difficult, it may be invalid to compare performance between devices from different manufacturers or even different versions of a device from the same manufacturer. And that's another question that we do want some input from the Panel.

It may also be difficult to account for differences in detection location and number of both true positives and false

Formatted: Line spacing: single

Formatted: Right: 0.25"

positives in such comparisons.

Now, I would like to discuss reader performance testing. While standalone performance testing indicates how well the device marks locations of interest, in the absence of radiologist interaction, it does not measure the safety or effectiveness of the device for its intended uses and conditions of use by a reader.

There are several types of reader performance tests that are designed to determine the impact of a CAD device on reader performance. Some of these have been discussed previously.

The prevalence of breast cancer in a screening population, as I noted previously, is, approximately, 0.4 percent. Following the Least Burdensome approach, reader performance testing may be accomplished using a so-called enriched dataset when you enrich the dataset with a significantly greater percentage of patients with breast cancer. So, you have a

Formatted: Line spacing: single

Formatted: Right: 0.25"

final population of patients with a prevalence of cancer that is much higher than that in a real screening population. And this is the population of patients that you use for your testing.

Ignoring prevalence, if the cancer and non-cancer cases that are used for enrichment, otherwise have clinical mammographic and pathologic characteristics that are typically seen in a screening population, such testing simulates the so-called field test, where a field test is clinical assessment of a system in real practice, you know, real life, real time.

However, the much higher prevalence in enriched datasets can introduce bias, as Dr. Gwise has previously discussed. And a question that we're interested in looking at is can reader performance testing using these enriched datasets give an estimated measure for device effectiveness as would be seen in clinical practice?

Formatted: Line spacing: single

Formatted: Right: 0.25"

And this is also where, as I think Dr. Gwise mentioned earlier, we're interested in what effect size would be needed in order to achieve this.

When enrichment is performed with only difficult cases, the testing is usually referred to as a stress test. In this situation, the test dataset is enriched with primarily or only difficult cases that challenge the readers. Stress testing that is done with only difficult cases will not capture information about the effect of CAD on cases that radiologists don't tend to miss. And this may lead to an incomplete assessment.

Another thing we are interested in getting input from the Panel is is stress testing alone sufficient to measure safety and effectiveness?

What about study endpoints for reader performance testing? Both overall and stratified reader performance, as I mentioned previously, can be reported per lesion, per

Formatted: Line spacing: single

Formatted: Right: 0.25"

view, per breast or per patient. And again, given that the clinical action following mammography are finding-specific, this is the same question we had for standalone performance, how critical is it to account for reader location accuracy when you are looking at reader performance testing?

What about stratified analysis? Meaningful stratified analysis may include breast density, finding size, finding type, or histologic type. And again, this is something we would like input from the Panel.

And now, I just want to touch on, this is the very end of my talk, some other issues for mammography CAD devices. Screen film and Full Field Digital Mammography (FFDM) devices have different spatial and contrast resolution and different noise characteristics.

Full Field Digital Mammography systems also vary between one another in spatial and contrast resolution because of

Formatted: Line spacing: single

differences of the types of solid state detectors, the pixel sizes, and the quantum and electronic noises. Different FFDM manufacturers may use different technologies and different image processing algorithms and techniques.

Therefore, the standalone and reader performance testing of CAD devices on screen film images may differ from testing results on FFDM images. This may also apply to testing on different FFDM devices. And again, a question that we would like input from the Panel is is there a reason why testing of CAD devices on a new or modified image input, whether it be a new digitizer for film or a new FFDM device, should that testing be any different than the standalone and reader performance testing already discussed?

Mammography CAD can be clinically implemented as a second reader or as a concurrent reader or perhaps other reader paradigms. Second reading can only increase

Formatted: Right: 0.25"

reading time, while concurrent reading may reduce reading time.

Another question we have for the Panel is how are mammography devices currently used in clinical practice? And then the last bulleted point there, mammography CAD has been reported with very high sensitivity for calcifications. Is there a clinical role for mammography CAD as a concurrent reader for calcifications?

And I'll end the talk there.

CHAIRMAN GLASSMAN: Thank you very much, Dr. Smith. Does the Panel have any questions for Dr. Smith? Dr. Berry, yes?

DR. BERRY: So, Dr. Smith, it's easy to see that CADs could improve sensitivity. Are there any CADs that the settings of the algorithms that improve specificity as well?

DR. SMITH: As far as approved indications for use, there is no approved indication for use that would deal with that

Formatted: Line spacing: single

Formatted: Right: 0.25"

issue. But that certainly would be a possible implementation.

DR. BERRY: So I don't understand. Are devices that have been approved, do we know what the sensitivity and specificity are under specific circumstances?

DR. SMITH: I would have to ask you if -- there are specific -- four different specific devices. Each device has its own performance testing and data that is provided in a submission. The current CAD device submissions have, essentially, not exclusively, standalone performance testing for those measures and testing with readers in order to measure recall rates. But there is no data available in current submissions for reader performance testing that would give you the measures that you are asking about.

DR. BERRY: Okay. I'm still somewhat confused, but maybe I'll be enlightened as the time goes along.

DR. SMITH: Well, I guess, if you

Formatted: Line spacing: single

Formatted: Right: 0.25"

are asking about the published literature, that's different than what's in FDA submissions.

DR. BERRY: So I'm trying to get a feeling for is it sensitivity/specificity that drives the clearance process? And are there devices that show both, you know, improvements in both? Is it necessary -- are there devices that have improved on the basis of let's say a statistically significant improvement in sensitivity without loss of specificity?

DR. SMITH: All currently approved devices, essentially, use the four components that I described in the slide. It's strictly standalone performance testing and reader testing only to look at recall rates. So there is no such data that has been used to approve current devices.

DR. BERRY: Another question. So just a follow-up. Is there any -- what is the range of improvement in sensitivity or specificity that one might see in these

Formatted: Line spacing: single

devices?

DR. SMITH: Are you talking about with a reader using the device?

DR. BERRY: Yes, yes.

DR. SMITH: There is a lot of literature out there on that. And there has been some conflicting results, depending on the methodology used for the studies. There have been basic -- two basic reader studies, the sequential study, where readers read without and then the same reader with CAD. Those studies have not shown statistically significant improvements, but the numbers of cancers have typically been quite small.

The studies that have been published have typically shown a trend toward improved detection of cancers that manifest as microcalcifications. But again, the numbers of cancers are quite small, so they have not been able to demonstrate statistically significant results.

DR. BERRY: So there is currently

Formatted: Right: 0.25"

no benchmark then for sensitivity or specificity for these products?

DR. SMITH: Well, that's one of the reasons that we have the Panel here is to give us some advice on this.

DR. BERRY: Thank you.

CHAIRMAN GLASSMAN: Dr. D'Orsi next.

DR. D'ORSI: Hi. Is it correct to assume that in the studies received by the FDA, there is no statistically significant difference in the data presented to the FDA for AUC?

DR. SMITH: We have not gotten AUC data with these submissions. Again, the data that has been used in the currently approved devices is standalone performance testing.

DR. D'ORSI: So there is no data on that part?

DR. SMITH: That is correct.

CHAIRMAN GLASSMAN: The last question.

Formatted: Line spacing: single

Formatted: Right: 0.25"

DR. BOURLAND: I have a question about the BCSC database. Who is -- has accessibility to that database? For instance, manufacturers. Are those images fully characterized, so to speak, available in DICOM form? And I'm sorry, the last part of the question is what's the image quality assumed for the images that, for instance, are inputs to CAD? For instance, are those MQSA compliant, so to speak, within the range or outliers allowed?

DR. SMITH: This is a database that, I think, is largely funded by the National Cancer Institute. It currently uses seven registries that link mammography and pathologic findings. The database is available -- the characteristics of the patients and numerous other features of the database are publicly available on-line. I do not know about the availability of the actual images.

They are all obtained from MQSA-

Formatted: Line spacing: single

Formatted: Right: 0.25"

certified facilities. But I don't know if the actual images themselves are available. I presume most of the images that are in that database are screen film. I don't know if they have ever been digitized and available.

CHAIRMAN GLASSMAN: I want to move on. We may have other questions, but we do have a chance later on in our general discussion to come back to these issues. So I would like to now proceed with the first of two Open Public Hearing sessions for today's meeting. The second Open Public Hearing session will follow the FDA presentation on colon CADs this afternoon.

Ms. Wersto will now read a statement prepared for Open Public Hearings.

EXEC. SEC. WERSTO: Both the Food and Drug Administration, FDA, and the public believe in a transparent process for information gathering and decision making. To ensure such transparency at the Open Public Hearing session of the Advisory Committee

Formatted: Line spacing: single

Formatted: Right: 0.25"

meeting, FDA believes that it is important to understand the context of an individual's presentation.

For this reason, FDA encourages you, the Open Public Hearing speaker, at the beginning of your written or oral statement to advise the Committee of any financial relationship that you may have with the sponsor, their products, and if known, any of their direct competitors.

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

Likewise, FDA encourages you at the beginning of your statement to advise the Committee if you do not have any financial relationships. If you choose not to address this issue of financial relationships at the beginning of your statement, it will not preclude you from speaking. Thank you.

Formatted: Line spacing: single

Formatted: Right: 0.25"

CHAIRMAN GLASSMAN: I would like to remind the public observers, at this meeting, that while this portion of the meeting is open to public observation, public attendees may not participate, except at the specific request of the Chair.

I would like to ask, at this time, that persons addressing the Panel come forward to the microphone and speak clearly, so the transcriptionist can make an accurate record of what you say. Please provide an electronic copy of your talk to the Executive Secretary for use by the transcriptionist to help provide this accurate record.

Prior to the meeting, we received formal requests to speak during today's Open Public Hearing session. Our first speaker is Heang-Ping Chan, American Association of Physicists in Medicine. And you will have five minutes. It is my understanding that after four minutes, you will get an orange light there, so that you have a minute to sum

Formatted: Line spacing: single

Formatted: Right: 0.25"

up. Thank you.

DR. CHAN: Good morning. I'm here to present a statement on CAD on behalf of the American Association of Physicists in Medicine or AAPM. The statement was drafted by the AAPM CAD Subcommittee. Here is a list of the subcommittee members. I'm the Chair of the subcommittee. Note that four of the members listed in the footnote did not propose or modify this statement.

Here is our statement. AAPM recognizes that CAD, including Computer-Aided Detection, Computer-Aided Diagnosis and more broadly Computer-Assisted Image Analysis, will be an indispensable part of diagnostic medicine in the near future.

Although commercialization of CAD systems is the bridge to clinical use, CAD research in academia has been and will continue to be a driving force for this progress.

Although CAD systems have been

Formatted: Line spacing: single

Formatted: Right: 0.25"

commercialized in several areas, the reported performances are excellent in some areas, for example, microcalcification detection on mammograms, but modest in others. For example, breast mass detection, lung nodule detection, indicating that CAD is promising, but the full potential of CAD has not yet been realized. Thus, further research and development of CAD should be strongly encouraged.

Continued funding support for research and development of CAD technologies will be vital for improvement in current CAD applications and development of CAD in new areas.

FDA-approval of a CAD system should be conditioned upon appropriate post-FDA-approval prospective evaluations of the CAD system in clinical practice. The majority of prospective clinical trials today indicated the promise of CAD in screening mammography, despite some negative reports.

Formatted: Line spacing: single

Formatted: Right: 0.25"

Standardization of study design and proper evaluation technologies are clearly important issues that need to be addressed.

Standardization of CAD evaluation technologies includes, but is not limited to, the following: clear definition of task, patient population, reader training for reading with CAD, without and with CAD reading design, definition of truth, data analysis methods, identification of biases and variances, and endpoint for assessment of the success or failure.

Journals should offer fair opportunities for publishing rebuttal or critical review of published studies. Quality assurance procedures should be established for CAD systems implemented in clinical use.

Radiologists should obtain training on the proper interpretation of the information provided by a specific CAD system before using it clinically. Procedures should be established to ensure that a CAD system is

Formatted: Line spacing: single

Formatted: Right: 0.25"

used as labeled.

Although an accurate CAD system can potentially be used as a first reader, such use should be first proven by properly designed prospective studies to evaluate its efficacy and approved by FDA. Vendors should be prohibited from advocating off-label use of a CAD system without proof and FDA-approval.

Thank you for your attention.

CHAIRMAN GLASSMAN: Thank you, Dr. Chan. Next is Stephen Vastagh from NEMA.

MR. VASTAGH: Distinguished Chair, Members of the Panel, Madam Secretary, FDA staff, good morning. My name is Stephen Vastagh. I am MITA staff liaison to the CAD manufacturers. MITA would like to thank FDA for holding this special Panel meeting on CAD.

I am employed by MITA, which is a membership organization of the manufacturers. Thus, I am and the organization is funded by the dues of the manufacturers.

MITA has represented the

Formatted: Line spacing: single

Formatted: Right: 0.25"

manufacturers of medical imaging and therapy systems for over 25 years. MITA has been on the leading edge of technical issues. We developed the early x-ray standards under the leadership of a former director of the Office of Device Evaluation.

Currently, we develop, maintain and publish DICOM, the world's foremost standard on imaging communication. Also, FDA experts were along side MITA member experts in international standardization of medical devices.

Today, we are here on behalf of the CAD group of MITA. In the interest of time, I will not review this slide, which was our organization and the CAD group in our organization.

Our goal is to provide the MITA members' views. Today, speakers from Fujifilm, Hologic and tomorrow, from Phillips Healthcare, GE Healthcare, and Medipattern will present these views. Our presenters have

Formatted: Line spacing: single

Formatted: Right: 0.25"

significant experience in the CAD field. These experts personally and their companies institutionally are just as much compelled to produce safe and effective CAD products as FDA is compelled to review whether they are safe and effective. The companies have a vested interest that all their products be safe and effective.

The scope of the CAD product is defined by the claims the manufacturer makes. These claims are interpreted into manufacturers operating instructions to define how the CAD product is to be used. MITA hopes that one of the outcomes of this Panel meeting is a common understanding of how CAD products are to be used, and the relative roles of the physicians and the CAD product.

We hope this will give confidence to evaluate CAD products to the extent of their claims and not require studies regarding possible off-label uses. It is likely that CAD technology is on the verge of increased

Formatted: Line spacing: single

Formatted: Right: 0.25"

growth and diversification.

New technologies are developed to accomplish similar functions; therefore, it is important to evaluate the products on the basis of their performance and not on the basis of technology.

In view of the advance of CAD technology, it is easy to reach a conclusion that CAD should improve the performance of radiologists. However, the manufacturers of CAD products do not intend such function, unless they explicitly claim so.

Given that physicians make decisions, which will hopefully be made clear during these two days, manufacturers should not need to demonstrate improvement of radiologist performance unless claimed. Indeed, this is not a requirement for other radiology devices.

Our presentations today and tomorrow conclude with several recommendations:

Formatted: Line spacing: single

Formatted: Right: 0.25"

FDA to look at the state of the art of CAD, consult with physicians and assess the risk of CAD product submissions in light of clinical experience.

Secondly, issue official guidance concerning the separation between Class II and Class III CAD devices and requirements. Use the role and involvement of the physicians as the guide to evaluate risk. Use a collaborative guidance development process that allows industry to provide input to the process.

Recognize that CAD guidance development generally takes two to three years, as has been mentioned previously. Also, recognize that industry is in great need of a uniform and transparent interim solution or an extraordinary expedited process for the guidance development.

These recommended actions will help introduce innovations and more advanced technologies to patient care. With these

Formatted: Line spacing: single

Formatted: Right: 0.25"

innovations in clinical practice, we can expect new clinical studies to advance this important science. These actions will also help to differentiate CAD from products that enhance and measure images, but are not CAD.

MITA looks forward to working with the FDA on this important issue. Thank you for your attention.

CHAIRMAN GLASSMAN: Thank you very much. Next, Robert Nishikawa from the University of Chicago.

DR. NISHIKAWA: Thank you very much, and I thank you for the opportunity to come and present some information here. This slide just describes my activities with Fujifilm Carestream and Hologic, but I'm here representing myself and the views I'm going to express here are those -- not necessarily those of my colleagues, collaborators or employer.

So evaluation of -- clinical evaluation is difficult period. It's

Formatted: Line spacing: single

Formatted: Right: 0.25"

extremely difficult if cancer prevalence is low. There is no perfect evaluation methodology, some compromises need to be made and this makes it difficult for deciding on FDA-clearance and approval process.

Part of the problem, is there is a view that the clinical data on CAD in screening mammography is conflicting. What I want to show here is that this is not true. The clinical data actually presents a consistent view that there is a benefit for using CAD.

So there are 10 clinical studies. I grouped them into two categories. One I'm going to call longitudinal, which I think is described as historical controls, and cross-sectional, which is the sequential study.

If you look at the longitudinal studies, you see the increase in the cancer detection rates around one to two percent, except for this one outlier here. The cross-sectional studies present an increase whenever

Formatted: Line spacing: single

Formatted: Right: 0.25"

cancer is detected at around 10 percent.

Oh, I see one more thing. And also, if you look at the Gromet Study, that actually measures sensitivity. It's about 8.2 percent, which is consistent with these numbers. And the Fenton sensitivity increases about 8.5 percent when you correct for a bias. And it's also consistent with these studies.

So we developed in our lab a simulation model to look at the effects of these two different evaluation methods. We start with, approximately, 50,000 women. We screen them every year and all cancers grow at the same rate.

So as one would expect, when screening starts, there is a prevalence effect, the cancer detection rate is high, but it goes down to some steady state level. For the same reason, we introduce CAD. There is a prevalence effect, and then you hit the steady state level.

But if you compare this curve to

Formatted: Line spacing: single

Formatted: Right: 0.25"

this curve, the difference is extremely small. So if you're measuring changes in cancer detection rate, it's very difficult to do. In fact, if the interval cancer rate was 0, these two lines would be collinear.

In a sequential reading method, you are trying to measure this difference here and it's actually quite doable.

Let's see, the previous example was with minimizing all of the sources of variability. You now add that cancers grow at different rates, so different numbers of cancers are present in the population each screening year.

This does not take into account variability. The radiologist will make this graph even more noisy. So now, what we are trying to do is measure some -- in longitudinal studies, some time point over here versus some time point over here. And you can see depending on how the data is -- falls, you may get -- measure an increase, you

Formatted: Line spacing: single

Formatted: Right: 0.25"

may measure no change, or you may measure a decrease in cancer detection rate.

But if you look at the difference between the yellow line and the white line, it is fairly consistent, even though the lines are going up and down. So, that in the cross-sectional methods or sequential method, you are able to measure this difference, but in the longitudinal or historical controls, you can't measure a difference.

The reason here is fundamental. See the -- the goal of CAD is not to find more cancers. It's to find cancers earlier. So if you measure CAD at detection rate, you are making the assumption you are going to find more cancers, which is not true.

And so using the longitudinal method, I believe, is flawed. Measuring this method, in fact, you're looking at how you can decrease the false negative rate, which is exactly what our CAD is trying to do. So this is a -- measures the effectiveness of CAD more

Formatted: Line spacing: single

Formatted: Right: 0.25"

directly than measuring the longitudinal -- than measuring a cancer detection rate and a historical control method.

There is one publication that says that, basically, CAD is bad to use, and I just want to address this quickly. There was actually six different results that I summarized here from the study. I can't discuss the reasons I have for all these points, but I will point out, for example, the sensitivity increase is not statistically significant.

They quote a value of 4.5. This is a biased estimate of the increase. I estimate it's close to 8.5 percent. I don't know. Sorry, 8.2 percent. I don't know if that's statistically significant, but it's substantially bigger than 4.5. I just addressed the cancer detection rate to increase. That's the wrong endpoint to use.

Many people have commented that the high recall rate is because the readers were -

Formatted: Line spacing: single

Formatted: Right: 0.25"

- the radiologists were inexperienced.

CHAIRMAN GLASSMAN: Excuse me, Dr. Nishikawa, your time is up.

DR. NISHIKAWA: Okay. Okay.

CHAIRMAN GLASSMAN: Thank you. Next is Dr. Carl Jaffe from the National Cancer Institute.

DR. JAFFE: Thank you. I'm Dr. Carl Jaffe, Branch Chief of Cancer Imaging Program, Division of Cancer Treatment and Diagnosis, NCI. I have no financial disclosures to report, but I am speaking as an individual, rather than for my colleagues or my division.

The imaging test, I think we have already had outlined, which is primarily the issue of detection and classification. We have already determined that expert observers are good, but inconsistent. That effective technology may reassure the best and conceivably could lift the performance of average performers. Man-machine systems, like

Formatted: Line spacing: single

Formatted: Right: 0.25"

any graphic user interface, to require run-in. The amount of time and the amount of experience of that is still to be determined. ROC analysis depicting the full range of observer talents is very expensive and often impractical.

We have already seen this graph from Craig Dean's work on mammographers. There is 110 mammographers nationwide published in the 1990s. And it shows you that a 10 percent false positive rate that the sensitivity for observers ranges from 50 to about 95 percent.

This tells you that people who might volunteer for a reader study may already be in the higher level. They may be a bias sample set, because the actual clinical practitioners are often falling well within a range that is much below what might be occurring in a confined reader environment, that is in a testable environment.

The imaging test needed for

Formatted: Line spacing: single

Formatted: Right: 0.25"

improving critically important are for therapy assessment. And quantitative monitoring of change over time turns out to be as yet not discussed in this particular meeting, but here performance is mediocre to poor on the observer's part. And hence, compromise work-arounds like the RECIST technique that was used, which are thought to be quantitative are, in fact, semi-quantitative.

We can see that easily when you look at a progress over time between two images on the same patient and the observers were unable to measure this consistently with a range of over 114 percent variations on that. We can also tell that ground truth is almost unknowable in many of the situations, particularly with lung, except in phantoms or simulations.

Quantitative solutions are an engineering exercise. They are not hard science. So they must evolve over time and must be proven by clinical use. One benchmark

Formatted: Line spacing: single

Formatted: Right: 0.25"

is the comparative human performance, and that is an effort that is being made by our division under what is called the Lung Image Database Consortium.

Validation is a clinical field exercise. Clinical correlates and outcome data are absolutely needed and, hence, the need for big public databases that are available for every developer of CAD systems to test against.

You can see here, for instance, it may be a little difficult to see the outline of the observers, but a large dataset now has become available in which the variation of observers for the measuring of something that is quite an obvious target become evident. And we have seen a little bit of that earlier in one of the presentations.

All of those, in a sense, are truths of a form. What we are doing at NCI is to try to help the -- and encourage the development of effective CAD by developing

Formatted: Line spacing: single

Formatted: Right: 0.25"

open public databases that contain the meta-data and the clinical data that can be correlated with it, so that new methods are tested against something that at least the performance becomes gradually evolutionarily known.

For instance, in this variable DICOM available dataset, the original first collection was actually the virtual colonoscopy set that was provided by the Department of Defense Study of 2003.

So what are the unresolved questions? It is how big must this referenced image dataset be? What are the training and testing components? And who possesses that dataset? Who maintains it? Who makes it available? And more importantly, who judges performance and by what metric? Thank you.

CHAIRMAN GLASSMAN: Thank you very much. Our next speaker is Steve Worrell from Riverain Medical.

MR. WORRELL: Good afternoon. I do

Formatted: Line spacing: single

Formatted: Right: 0.25"

not have slides today. I'll just be reading from a prepared statement. My name is Steve Worrell. I'm Vice President of Research at Riverain Medical.

To date, Riverain Medical is the only company to have demonstrated the safety and effectiveness of a Computer-Aided Detection device intended to identify and mark regions of interest on frontal chest radiographs. The device is intended for the use as an aid only after the physician has performed an initial interpretation of the x-ray.

As a company who has committed significant resources to maintain conformance with the strictest Category 3 regulatory requirements, we, obviously, have a strong belief that Class III Regulation is necessary to mitigate risk associated with today's chest CAD devices.

PMA processes insure that CAD devices are properly developed, tested,

Formatted: Line spacing: single

Formatted: Right: 0.25"

marketed, sold using approved indications of use and clinically validated performance claims. Chest CAD technology has evolved into a complex and integral part of the practice of medicine.

It's only through appropriate controls that chest CAD has become and will continue to be a trusted element of the clinical work flow. Chest CAD is used to aid physicians in discovering important findings that would otherwise go undetected.

Riverain believes in the importance of the PMA process and Class III Regulation of chest CAD technology. Chest CAD is appropriately classified as a Class III medical device. It is part of a generic type of device that is intended for a use which is substantial, important in preventing impairment to human health and its use can prevent a risk to patients.

For example, false positives can link to unnecessary CT exams. Currently CAD

Formatted: Line spacing: single

devices are regulated as both Class II and Class III. While this apparent inconsistency in FDA Regulation can be explained, the distinction between the regulatory classifications is not clear.

In fact, the intended use and design of Class II chest CAD devices are encroaching in areas that deserve Class III regulation. If the integrity in the PMA process is to be preserved, a clear distinction between Class II and Class III chest CAD devices must be delineated by the Agency.

As a Class III device, the Agency has subjected Riverain chest CAD to justifiably high degree of regulatory control. As a consequence, Riverain maintains high standards that affect all company activities from the development, the manufacturing, regulation of labeling, as well as promotion and advertising.

Because of the importance of Class

Formatted: Right: 0.25"

III Regulation, Riverain advocates requiring a reader study for initial PMA chest CAD applications.

Further modifications could be supported by standalone testing on independent clinical test sets that demonstrate the modified device is performing as good or better than the approved device. This approach is consistent with the Least Burdensome provisions of the FDA Modernization Act and provides reasonable assurance of safety and effectiveness.

While we appreciate the need to thoroughly evaluate devices prior to their release, we urge the Agency not to regulate CAD and present -- prevent superior products from being released to the market.

In summary, Riverain believes the following:

(1) The integrity of the PMA process should be preserved.

(2) A reader study consistent with

Formatted: Line spacing: single

Formatted: Right: 0.25"

the intended use should be conducted for the original PMA applications.

(3) Reasonable supporting evidence for modifications to algorithms consistent with the Least Burdensome provisions of the FDA Modernization Act should be required.

Thank you for taking the time to allow me to state my opinions.

CHAIRMAN GLASSMAN: Thank you very much. Our next speaker is Akira Hasegawa of Fujifilm Medical Systems USA.

DR. HASEGAWA: I'm Akira Hasegawa from Fujifilm. The title of my talk is Risk Assessment of CAD. As you know, there are many different type of CAD. The different CAD have different indication for use and IFU is different. Risks are also different. Please, note that standard -- the standard reading procedure may be depending on organs and modalities. Risk assessed by this comparison may vary depending on organ and the modalities.

Formatted: Line spacing: single

Formatted: Right: 0.25"

This slide shows the standard reading procedure without CAD. A case comes in, and the radiologist start searching process -- searching suspicious area. Once he identifies the suspicious area, he looks closer and makes decisions. And it is repeated until he goes through all area in the case.

Here, we would like to insight that radiologists are supposed to look at everything, evaluate all suspicious areas, and they should not overlook anything, even without CAD.

Type 1. This is the standard procedure without CAD. Again, even without CAD, radiologists should not overlook anything. However, radiologists are also human being, and he may do some perceptual oversight. To avoid such perceptual oversight, after the completion of this standard reading, radiologist can do optional second read.

Formatted: Line spacing: single

Formatted: Right: 0.25"

And there is a CAD to support this optional second read process. This CAD is normally called Computer-Aided Detection, but actually this CAD does not assist detection process, it's here, but assist optional second read process. Because this optional second process after the standard reading procedure, there is no effect to this standard reading process.

Type 1 CAD is for optional second read. It assist the users to do optional second read. It helps users to reduce potential oversight. The way it used as recommended by the manufacturer, this type of CAD does not affect users' initial reading. And it does not provide any diagnostic information to users.

Type 2. Again, this is the standard reading procedure without CAD. Again, radiologist should not overlook anything without CAD. However, radio -- every radiologist has -- sorry. Every radiologist's

Formatted: Line spacing: single

Formatted: Right: 0.25"

knowledge and experience are different. If some radiologist does not have enough knowledge and experience, he may make cognitive oversight in decision making process.

To avoid such a cognitive oversight, there is a CAD to assist radiologist to do decision making. Normally, this is called as Computer-Aided Diagnosis, because this CAD influences readers' decision making process, which is a part of the standard reading procedure.

Types of CAD is for interpretation. It assist the users to make an interpretation. It provides classification information or diagnostic information to users. It affect users' interpretation process. The risk, it may affect users' decision making ability negatively.

The Type 3. Again, this is the standard reading procedure without CAD. There is a CAD to assist radiologist to detect

Formatted: Line spacing: single

Formatted: Right: 0.25"

suspicious areas. This type of CAD is CAD for concurrent read, and it's normally called as concurrent CAD. Obviously, this CAD affect certain process in the standard reading procedure.

The Type 3 CAD for concurrent read it helps users to find the suspicious ROIs. It affects users' initial searching process. It does not provide diagnostic information to users. Risk, it cause users satisfaction of search. It may affect users searching negatively.

CHAIRMAN GLASSMAN: Excuse me, Dr. Hasegawa, but your time is up. Thank you.

DR. HASEGAWA: Oh, I thought we had -- I have a 10 minutes?

CHAIRMAN GLASSMAN: You had five minutes.

DR. HASEGAWA: Oh.

CHAIRMAN GLASSMAN: Does anyone on the Panel have questions for any of the speakers who have just presented? Dr. D'Orsi?

Formatted: Line spacing: single

Formatted: Right: 0.25"

DR. D'ORSI: Dr. Hasegawa, are you suggesting the use of CAD Type 1 and Type 2 on screening exams?

DR. HASEGAWA: Actually, I not really suggesting. It's -- my point is that it's a depending IFU from manufacturers.

DR. D'ORSI: Thank you very much. Could I do a question for Dr. Jaffe?

CHAIRMAN GLASSMAN: Go ahead.

DR. D'ORSI: Dr. Jaffe, the database for the ACRIN Study, would that be something that would be available to the FDA for a sort of general testing set that could be developed?

DR. JAFFE: ACRIN is a grantee of ours. We have -- we are working out a complex arrangement for the issue of their initiatives for data-sharing and ones that we believe are important for both parties, both NCI and the field in general. We have done that through the terms of award on the virtual colonoscopy trial, which has just finished, and there will

Formatted: Line spacing: single

Formatted: Right: 0.25"

be a substantial portion of that trial made available.

The problem with the issue of the DMIST trial is that it included devices as well as CAD, as well as what could be CAD components. When we get into the issue of devices, it gets rather complex, because there are multiple parties that have financial interest in that. So that has been more complex.

It is absolutely an area of great interest to us, because we feel this field really could develop much better if we had some open databases that can act as benchmarks. It puts more competition into the system. So you don't actually have to file with say the database and the operational database, but it allows you to actually develop your material, so you know how well you are doing against other competing procedures.

CHAIRMAN GLASSMAN: Thank you, Dr.

Formatted: Line spacing: single

Jaffe. Any other questions? If not, I want to move on to our public speakers about mammography CAD. First is Dr. Rachel Brem on behalf of iCAD Medical. She is also a grandparent with a new granddaughter, by the way.

DR. BREM: Thank you very much. I'm Rachel Brem. I have been involved in CAD research for, approximately, a decade and have been a Director with iCAD and am a Director, but primarily I'm a practicing radiologist and a Director of Breast Imaging at George Washington University and personally interpret over 10,000 mammograms a year.

What I would like to review today is the scientific integrity, and the rigor of the data presented for the original PMA submission, as well as subsequent PMA supplements that establish the safety and efficacy of mammography CAD and that the currently utilized paradigm of establishing the safety and efficacy with the PMA is

Formatted: Right: 0.25"

appropriate one as validated by the literature.

The scientific data which we presented with our original PMA submission resulted in a 20 percent improvement in breast cancer detection. And it was based on a large scientific body of data, which not only showed an improvement in breast cancer detection, but that breast cancer with CAD was diagnosed 15 months earlier than without CAD.

The clinical utilization study of nearly 4,000 patients demonstrated a significantly -- a statistically insignificant .5 percent increase in recall rate.

The data which we submitted for CAD standalone performance utilized a large number of breast cancers to speak to both the stability of the data using ground truth as well as the vast array of mammographic presentations of breast cancer and 153 normals to establish the false positive rate.

CAD -- A mammogram is a mammogram,

Formatted: Line spacing: single

Formatted: Right: 0.25"

regardless of how the image is obtained. In the original PMA submission, which the base -- is the basis for establishing the safety and efficacy of mammography CAD, requires confirmatory studies that the standalone performance of new image sources of mammography demonstrate comparable performance.

A mammogram is the same image, regardless of whether it is a screen mammo -- film screen mammogram or digitally obtained. And therefore, the confirmatory studies need only show comparable performance with the safety and efficacy based on the original PMA submission.

The literature is full of studies evaluating the improvement of breast cancer detection, and whether it be sequential studies that is on the same patient prior to and after the implementation of CAD, which has been reported in over 54,000 patients or historical controls, that is the cancer

Formatted: Line spacing: single

Formatted: Right: 0.25"

detection rate prior to and after the implementation of CAD, which has been reported in nearly 150,000 patients, have demonstrated improvement in CAD detection from 4.5 to 19.5 percent. And regardless of the study design, an approximate one percent increase in recall rate.

And there is an obligatory need to have an increase in recall rate in order to identify more cancers. And as a practicing physician, after an initial learning curve, I can tell you that the false positive rate becomes a minimal, if distractor at all. And the increase in recall rate to patients, the cost of the increase recall rate to patients is not only acceptable, but based on my patients, often welcome for the improved detection of breast cancer.

CAD works, and the literature supports that, whether it be in private practice or in the academic setting, with an experienced and novice mammographer, although

Formatted: Line spacing: single

Formatted: Right: 0.25"

the impact is greater with a novice mammographer, and in prospective and retrospective trials.

Also, fatty and dense breast mammography works, and this is important as dense breasts not only hinder interpretation of mammograms, but is a strong independent risk factor for the development of breast cancer.

And with regard to cancer size, CAD works not only in large cancers, but in small sub-centimeter cancers and even in cancers that are 5 millimeters or less. With regard to pathology, CAD works in the most common invasive ductal carcinoma and the most difficult to diagnose invasive lobular carcinoma.

And make no mistake, ductal carcinoma in situ is, in fact, breast cancer. And although it is not yet invasive, if we could diagnose all cancers in their in situ phase, then we would be able to, essentially,

Formatted: Line spacing: single

Formatted: Right: 0.25"

cure breast cancer. Once a woman has an invasive cancer, we can no longer assure her a cure from breast cancer. And therefore, the diagnosis of ductal carcinoma in situ is not only desirable, but laudable.

So in summary, mammography works in all clinical situations with various types of readers, pathology size, and pathology and size. And therefore, the currently utilized paradigm of establishing the safety and efficacy with the PMA submission and subsequent confirmatory studies to establish comparable performance of CAD is critical.

As a physician, as a woman, and as a breast cancer survivor, the critical technology of CAD is important for the optimal diagnosis of breast cancer. Thank you.

CHAIRMAN GLASSMAN: Thank you. Our next speaker is Mr. Julian Marshall from Hologic/R2. Is Mr. Marshall here? Yes?

MR. MARSHALL: Good morning. In my 14 years at R2, now Hologic, I have had

Formatted: Line spacing: single

countless opportunities to discuss CAD, particularly mammography CAD, with radiologists of all skill levels. The conversation has become easier over time, particularly in the U.S., where adoption has been rapid, and an amounting body of evidence has made it more and more apparent that mammography CAD is clinically useful.

Over the last 10 years, hundreds of papers have been written on mammography CAD. They have varied widely in study design, population, and controls. It seems that the days of tiny academic studies is largely over, replaced now by much larger studies with real clinical basis authored by practicing clinicians.

Yet, there is a surprising amount of debate as to what truly will establish the clinical efficacy of CAD. Early on, we were told that CAD would not be established until independent prospective studies were published in peer reviewed journals. Subsequently, we

Formatted: Right: 0.25"

became aware that it would be necessary for those studies to demonstrate increase in detection concordance with increase in workup rate. And just in the last few years, reduction in size, stage and age at detection.

By this time, a handful of prospective clinical studies have been published, most on the scale of 10 to 20,000 patients. All have demonstrated a cancer detection rate increase, some concordant with the increase in workup rate. Dr. Young's paper is an outlier because CAD was used as an adjunct to human double reading in a clinical environment where any single radiologist can trigger the recall.

In addition, please, note the dates in red on this pie chart. Mammography CAD was first approved in 1998, yet, it took three years before the first independent prospective study was published. Requiring such studies prior to approval of a device will dramatically delay putting these clinically

Formatted: Line spacing: single

Formatted: Right: 0.25"

beneficial low-risk products into the hands of clinicians.

When institutions implement CAD into their clinical work flow, the incidence of breast cancer doesn't suddenly increase. Yet, the prospective studies all show an increase in detection with use of CAD. It wasn't until Dr. Cupples' published a paper in 2005, that it became clear why this was.

Dr. Cupples was first to point out that the extra cancers detected were really future years' cancers detected earlier. His paper demonstrated a 164 percent increase in detection rate of invasive cancers less than 1 centimeter and found those cancers in women 5.3 years younger.

He wrote that in multi-variable analysis of invasive cancers early stage, State 1, was strongly associated with detection by CAD. Subsequent publications, particularly by Dr. Nishikawa have discussed this earlier detection effect in far more

Formatted: Line spacing: single

Formatted: Right: 0.25"

detail.

Mammography CAD is an ancillary source of information for the reading radiologist, who remains the interpreter of the images and the final arbiter. Should the CAD device fail or results not be available, the radiologist must still be able to finish the reading in a timely manner.

As in all healthy debates, mammography CAD papers and opinions exist that indicate evidence to the contrary, such as the heavily marketed paper by Fenton, et al. But for each of those more critical studies, there is also a lesser known body of published commentary pointing out the significant weaknesses in those studies.

There is clearly a mountain of evidence indicating that mammography CAD has a positive clinical impact, and we are pleased by the recent publication of the paper by Dr. Gromet, which included more than 118,000 cases read with CAD and which showed statistically

Formatted: Line spacing: single

Formatted: Right: 0.25"

significant performance improvement.

Fenton, et al, which is often used as evidence that CAD does not work only studied just over 31,000 cases with CAD, less than 27 percent of the number used by Gromet.

Please, read the Gromet paper, which the FDA has kindly provided to you.

In the last 10 years, the number of women have confided that my doctor told me he would not have found my breast cancer without your system. I don't understand why a CAD device that is already approved with the given sensitivity/specificity cannot be upgraded to a better algorithm with better performance using the same assessment of standalone performance. Thank you.

CHAIRMAN GLASSMAN: Thank you. I was just told that I downgraded you to a physician, Mr. Marshall. I apologize.

DR. MARSHALL: I can clarify that.

CHAIRMAN GLASSMAN: On the other hand, our next speaker is a physician, Dr.

Formatted: Line spacing: single

Formatted: Right: 0.25"

Robyn Birdwell from the Society of Breast Imaging.

DR. BIRDWELL: Thank you very much. I am sponsored today by the Society of Breast Imaging, but I come to you as a user of CAD since the late 1990s, as a researcher in the area and, in my opinion, CAD does make a difference in reducing false negatives. This is an opinion shared by other colleagues with a modest increase in recall rate.

Radiologists overlook cancers. We have heard many reasons for this. It happens whether it is film screen or digital mammograms. I turn to the DMIST study just to show you that despite training, experience, continuing medical education, review of audits, we miss cancers.

It's a high volume, low prevalence in the screening mammography world, four cancers in every 1,000 examinations. And you can see that having a tool to improve sensitivity is a good idea.

Formatted: Line spacing: single

Formatted: Right: 0.25"

We miss cancers because they can't be seen. We don't know what can help us for that, at this time. Visible and detected by the radiologist, but we have an error of cognition. We see it, but we don't interpret it correctly. This is not what CAD is presently designed to help with. It's visible, but simply overlooked by the radiologist. It's an observational oversight or an error in perception. This is where CAD is helpful in our today's practice.

The performance and benefit has been demonstrated to be equivalent, whether the image source is film screen, digital mammography or computed radiography. Again, the present benefit, the so-called CADe, is in making visible and overlooked cancers visible to the radiologist by reminding us to look again at this area.

How do we measure the performance in the clinical setting? We must, as always, focus on safety and efficacy. We assess

Formatted: Line spacing: single

Formatted: Right: 0.25"

safety by recalls. We assess safety by biopsy recommendations. Do we indeed recall more patients because of CAD? Yes. If you want to find more, you have to do more. Do we have more biopsy recommendations? Yes. If you want to find more, you need to do more.

Are we finding, however, additional cancers that otherwise would have been overlooked? Yes. We need to look at the prospective sequentially read clinical studies, you have heard much about most of these today, looking at the conglomerate of 53,000 number of screens from studies done both in the community as well as in the academic world.

We see an overall increase of cancers detected at 9.7 percent, an increase in biopsies, and a moderate increase, percentage increase in recalls. The numbers here vary possibly because of the patient populations, possibly because of the practice numbers, maybe because of the type of

Formatted: Line spacing: single

Formatted: Right: 0.25"

radiologist reading these studies, but overall, there is an increase.

This next slide, obviously, shows I don't have CAD on my computer. This should say historical studies in the first line. We have heard about these as well. The Gur Study, the Cupples, the very recent Gromet Study, and then the much touted and difficult to interpret Fenton Study.

But overall, we see again, at least if you look at the low volume readers in the Gur Study, an increase in cancer detection to 19.7 percent. The most recent study with 112,000 examination read pre- and, approximately, the same number read after initiation of CAD by Gromet, we see an increase of 11 percent sensitivity. An improvement in sensitivity of 11 percent.

And then the outlier of Fenton that was a survey type study and bears its own issues.

In conclusion, based on the USA

Formatted: Line spacing: single

peer reviewed literature of prospective, sequentially read or historically controlled clinical studies, the preponderance of evidence shows that currently FDA-approved CAD systems and algorithm improvements are efficacious, by increasing cancer detection, and improving radiologist sensitivity, and safe as looking at a modest increase in recall rates concordant in most studies with the increase in cancer detection.

The FDA review process for mammography CAD it works as intended for women and their caregivers. Thank you.

CHAIRMAN GLASSMAN: Thank you very much. Next, Dr. Gillian Newstead for the American College of Radiology.

DR. NEWSTEAD: Good morning. It's a pleasure to be here, and I would like to disclose that I do receive research as indicated here, research support from Phillips Medical, Bayer Health, and my spouse is a stockholder in Hologic.

Formatted: Right: 0.25"

Formatted: Line spacing: single

Formatted: Right: 0.25"

I would like to move a little bit from the mammography CAD into the broader aspects of integration of the digital enterprise for breast imaging, because it's now very difficult for the healthcare provider to assimilate clearly and effectively, all of the different information presenting to us and moving forward in breast cancer diagnosis and screening. There is a vast increase in the amount of information that is available to us now, and radiologists are basically drowning in large amounts of data.

Certainly, it is true with the recommendations for MRI screening in high risk women, in addition to mammography screening for women with no known risk factors. And moving into the diagnostic area of CAD as it relates to breast cancer diagnosis, we do need computer-assistance to view, manipulate and analyze large complex imaging datasets, beyond the mere interpretation and analysis of a two view screening mammogram.

Formatted: Line spacing: single

Formatted: Right: 0.25"

In addition, if radiologists are going to become very more involved in patient care management, in terms of making real diagnostic decisions that affect healthcare, we need information regarding the patient's medical records, clinical history, pathology information, et cetera, which will allow us to compose more detailed and management-type decisions that would be important.

This is how I'm reading right now at the University of Chicago. I have multiple, in our reading room -- we have 27 monitors, all of them related to a different kind of acquisition modality devices. We are juggling mammograms, ultrasound, magnetic resonance imaging, interventional studies, reports, all of this partly from symptomatic patients, partly from screen detected mammography patients. We need to review pathology to make management-type reports and integrate the clinical findings.

Computers are going to be essential

Formatted: Line spacing: single

Formatted: Right: 0.25"

for us in order to be able to do this. One example I'm just going to show would be the interpretation of breast MR images, where we need computers to assist us in demonstrating the kinetic and morphologic features of lesions that we detect at screening.

The MRI Lexicon for breast MR provides a very nice organization of reports of findings that we will deliver in our Lexicon reports for lesions that are detected. It has been shown that radiologists are aided by having computers depict the kinetic characteristic lesions, rather than visually estimating them during the reading process.

Computers can allow us to select certain areas of thresholds and display the enhancement characteristics of lesions that are exhibiting kinetic enhancement beyond a certain threshold, which is very helpful to us when interpreted in context with the morphology of the lesions.

Computers allow us to produce

Formatted: Line spacing: single

Formatted: Right: 0.25"

volume metric assessments. This is important in follow-up in the diagnostic setting, display the enhancement curve of the lesion, which you see here, and produce angiogenesis maps, which depict again the dynamic information appropriate to the particular lesion detected.

So in general, computers are really important in allowing us to progress from a purely visual analysis of a 2-dimensional film to a 3 and 4D analysis of advance technology as we move forward, and how we incorporate that will be important for this panel and others to evaluate these devices and learn and test them.

What does this mean for CAD devices? For example, an MR, well, the radiologist is using these devices as a tool really to assist us in diagnosis. We are using color mapping, angiogenesis mapping, and many other display methods as a tool for reaching an accurate diagnosis.

Formatted: Line spacing: single

Formatted: Right: 0.25"

This may help in depicting cancers, that may be overlooked by the radiologist, and morphologic assessment is going to be important.

So in conclusion, I think computers will increasingly become an active and important play -- play an important role in the diagnostic assessment and in the screening assessment of lesions as technology proceeds to give larger datasets some more advanced information than we currently have. Thank you very much.

CHAIRMAN GLASSMAN: Thank you very much. Our next speaker is Dr. Robert Nishikawa from the University of Chicago.

DR. NISHIKAWA: Thank you very much. I already disclosed my financial activities. So I showed this slide earlier. I condensed the numbers down into this slide. And the two, three important numbers are the 9.7 percent increase in sensitivity, 13 percent increase in recall rate and if you

Formatted: Line spacing: single

Formatted: Right: 0.25"

take the ratio of the pink to the green, you get 1.3.

So let's suppose that's exactly what the CAD does clinically, and we want to be able to measure that in a reader study. So those are the two points. And you can fit two curves to those ROC curves, and you want to compare the area under the curves.

Well, the area difference is only .025. And if you try to design a reader study to find that difference, it's a very huge study. These are based on Obuchowski's work, but she doesn't go down to that small of a difference. So I don't know exactly what the number of cases is, but I'm guessing even at 10 readers, it's probably a few thousand. I don't think that's a practical observer study to conduct.

So then, what is a reasonable requirement for the FDA to approve or approve CAD devices? Well, the goal of CAD is to reduce the number of missed cancers by

Formatted: Line spacing: single

Formatted: Right: 0.25"

radiologists. So then I think it's reasonable to expect CAD to work something like double reading with two independent readings. I'm sorry. Independent double reading with two radiologists, which in the literature shows, approximately, 10 percent increase in sensitivity and a comparable increase in recall rate, basically, what you find in the clinical studies to date.

So one may argue, that it's both sensitivity and recall rate increase are just shifting on the ROC curve. Well, that dotted green curve is a single curve through those two points. And in fact, if that was the true curve, you would be shifting on the curve. But that is not a valid ROC curve for people who understand ROC analysis, those are the A and B values and they are not consistent with anything in the literature.

Also, if you look at the curve, it says that radiologists, at the very bottom, recall a lot of women before they find any

Formatted: Line spacing: single

Formatted: Right: 0.25"

cancers, which is not found clinically.

Another argument is that reading more aggressively, which is sliding on the -- it really moves you on the ROC curve, is basically the radiologist changing their threshold for recalling a patient. But what CAD does, is help find cancers that were missed. No matter how aggressively you read the image, if you don't look at a certain part of the image where the cancer is, you will never find it. So a higher vigilance is what CAD is giving you. And higher vigilance will move you to a higher ROC curve.

I'm going to skip over that. So what's a reasonable comparable increase in recall rate? So here the yellow dot is a four times increase in recall rate compared to the increase in sensitivity. And you can fit an ROC curve, that pink curve, which is reasonable ROC curve for a radiologist, so you can go up to a four times increase in recall rate compared to sensitivity and still be

Formatted: Line spacing: single

Formatted: Right: 0.25"

considered operating on a different ROC curve.

So one reasonable endpoint might be that the increase in recall rate to the increase in sensitivity is less than 3, than that's consistent with reading on two separate ROC curves, which means you have an increase in performance. You won't be able to measure that in the study, because the differences are very small.

So let me just comment on a couple of other things that the Panel is considering. Inclusion of benign cases in observer study. I think this adds a complication to interpretation of CAD. For CAD to be effective, the radiologist needs to be confident that the system can find cancers.

Since benign lesions that go to biopsy appear malignant to radiologists, who think that it's a cancer, then the computer needs to point those out. But in scoring those, since they are a benign lesion, scores are false positive. So if the computer

Formatted: Line spacing: single

Formatted: Right: 0.25"

doesn't point to it, the radiologist loses confidence and doesn't use the computer effectively. But if it does point to it, you get penalized for pointing that out.

So this is a problem particularly in an enriched test set. If you had the normal prevalence, this wouldn't be a problem.

I also would like to comment on evaluating CAD for different digital systems. It's not reasonable to me to require PMA for each -- for a CAD on each different digital detector, since DMIST found no difference in radiologists' performance between different digital systems.

It's more reasonable to show CAD is effective on one digital system and show comparable performance on the other manufacturers systems. Thank you.

CHAIRMAN GLASSMAN: Thank you. The next speak is Dr. Mary Ellen Giger from the University of Chicago. And she is a real doctor. She is a Ph.D.

Formatted: Line spacing: single

Formatted: Right: 0.25"

DR. GIGER: Thank you. I'm from the University of Chicago. I have been involved in CAD research for almost a quarter of a century now. I'm representing myself. I have research report from NIH and the Army, and I am a shareholder and receive funding from R2/Hologic.

Well, I'm concerned about the timeliness and consistency of the translation of CAD developments to clinical use. And, as mentioned earlier, computers are increasingly being incorporated into our lives. And the progress of CAD depends on research funding, careful clinical studies, and training with the methods when introducing them into the clinical arena.

Let's stop for a second and look at the imaging chain. All these various components ultimately lead to the radiologist interpretation and decision. Many factors affect the radiologist's performance levels, including some such as shown here in the

Formatted: Line spacing: single

Formatted: Right: 0.25"

physical quality of the image.

Attempts to help in this area led to new detector systems along the imaging chain, as shown here where a screen film system versus Full Field Digital Mammography system was investigated.

Other areas along the imaging chain affect the ability -- may help the ability of the radiologist in terms of prior training and knowledge, and help overcome some interpretation conditions. These are illustrated in orange here along the imaging chain, which includes CAD, both in quantitative image analysis and image display, and how you ultimately output that CAD output to the radiologist.

And how does the radiologist use this? Well, the radiologist could read without. They can perform double reading, or they could read with CAD. So to get to the point of my talk, and I was so worried about time, I rushed through those probably too

Formatted: Line spacing: single

Formatted: Right: 0.25"

quickly.

I'm just giving maybe a new way to go in evaluating systems for the FDA. We know that double reading improves detection sensitivity with some increase in recall rate. This occurs at the intended goal of double reading is to improve sensitivity.

Currently, double reading by radiologists is accepted, even though it has never gone through any FDA-approval. From the literature, the increase in sensitivity and recall rate are similar for double reading and single read with CAD.

With academia and commercial CAD manufacturers, it might be useful to demonstrate equivalency between single read and CAD. Equivalency between single read with CAD and radiologists double read, similar to DMIST with the screen film mammography and Full Field Digital Mammography. I am aware that that study was looking for superiority, however, for the FDA requirement, my

Formatted: Line spacing: single

Formatted: Right: 0.25"

understanding is equivalency was accepted.

So the way to do this would be to determine a performance standard for CAD based on published data, require a CAD system to meet that standard to be included in a cooperative study, based on performance on some test set. Perform the cooperative study, similar to an ACRIN Study, but here the goal will be equivalency and not superiority. And allow only radiologists trained in CAD usage to participate in this one big observer study.

After equivalency is demonstrated via this multi-institutional study, as improvements are made to a CAD system, it will be necessary to demonstrate that the systems performance meets or exceeds the earlier specified performance standard.

One could use an independent technology assessment institute, which would be tasked with the performance assessment of all new and improved CAD devices. The institute would have a safe, sufficiently

Formatted: Line spacing: single

Formatted: Right: 0.25"

large database with appropriate distributions of cancer types to allow for random sampling of cases for a CAD system that is undergoing assessment.

Of course, one would have to worry about the potential different types of false positives from different systems. However, for this large database, it should be large enough so that one could randomly select a subset of cases that match the distribution, and this would help preserve the integrity of the test set. Thank you.

CHAIRMAN GLASSMAN: Thank you very much. Does anyone on the panel have any questions for our last set of presenters? Dr. D'Orsi?

DR. D'ORSI: Just one question for Dr. Nishikawa and I think Dr. Brem. Do you think that the use of CAD in secondarily obtained digital images, i.e., from film, is equal to the CAD operation from directly attained information from direct digital? Do

Formatted: Line spacing: single

Formatted: Right: 0.25"

you think that one requires a PMA? I agree with you that the others do not, but what about the difference between secondarily derived digital data and primarily derived digital data for CAD performance?

DR. NISHIKAWA: Are you talking about digitizing the screen film images?

DR. D'ORSI: Correct.

DR. NISHIKAWA: I think that the early overture is that they are probably equivalent, but I am not 100 percent convinced. But I don't think you necessarily then have to go through a PMA to establish efficacy.

DR. BREM: And I agree. I think that the efficacy and safety was established on digitized analog data and that, you know, studies that show comparable performance is what is needed for different image sources.

CHAIRMAN GLASSMAN: Dr. D'Orsi again.

DR. D'ORSI: One question for Dr.

Formatted: Line spacing: single

Formatted: Right: 0.25"

Nishikawa. How did you arrive at the four times increase recall to washout the increase in sensitivity? Was this in order to obtain a different ROC at a different level or was it something else operative?

DR. NISHIKAWA: Right. So what I tried to do was find -- so I have two points in the ROC curve, and I'm moving one to the right, increasing the recall rate, sensitivity is the same. And I kept doing that until the ROC curve became at least symmetric.

So if you look at the DMIST ROC curves, they are all skewed to the left, which is what you would expect. So in the earlier green curve I showed you, was skewed to the right, which is not a real curve. So once it became symmetric, I said that's a possible ROC curve for radiologists.

CHAIRMAN GLASSMAN: Question?

DR. BOURLAND: Yes. A question for Dr. Giger. Are you suggesting an institute, for instance, a radiological imaging center or

Formatted: Line spacing: single