SUMMARY MINUTES

CENTER FOR DEVICES AND RADIOLOGICAL HEALTH

RADIOLOGICAL DEVICES ADVISORY COMMITTEE

OPEN SESSION

March 5, 2008

Hilton Washington DC North
Gaithersburg, MD

Radiological Devices Panel:
March 5, 2008
Attendees:

Acting Chairperson:
Leonard M. Glassman, M.D.
Washington Radiology Associates, P.C.

Voting Members:
John D. Bourland, Ph.D.
Wake Forest University

Carl J. D'Orsi, M.D.
Emory University Hospital

Bharat B. Mittal, M.D.
Northwestern Memorial Hospital

Marvin C. Ziskin, M.D.
Temple University School of Medicine

Temporary Voting Members:
Craig K. Abbey, Ph.D.
University of California, Santa Barbara

Donald A. Berry, Ph.D.
The University of Texas, Houston

John A. Carrino, M.D., M.P.H.
Johns Hopkins University School of Medicine

Lori E. Dodd, Ph.D.
National Cancer Institute/NIH

Brian S. Garra, M.D.
Fletcher Allen Healthcare

David Kim, M.D.
University of Wisconsin Medical School

Marilyn Leitch, M.D.
University of Texas Southwestern Medical Center

Otto Lin, M.D.
Virginia Mason Medical Center

Robert D. Rosenberg, M.D.

University of New Mexico Health Sciences Center

Berkman Sahiner, Ph.D.
University of Michigan Medical Center

Kenneth J. Steier, D.O., M.P.H., M.H.A.
Nassau University Medical Center

Daniel R. Swerdlow, M.D.
Georgetown University Hospital

Georgia D. Tourassi, Ph.D.
Duke University Medical Center

A. Christine Watt, M.B., Ch.B.
St. John Hospital

Roy K.H. Wong, M.D.
Walter Reed Army Medical Center

Non-Voting Consumer Representative:
Nancy Finken, M.F. A.
McLean, VA.

Non-Voting Industry Representative:
David K. Spindell, M.D.
Abbott Laboratories

Executive Secretary:
Nancy Wersto

CALL TO ORDER

Chairman Glassman called the meeting to order at 8:30 a.m. and noted the presence of a quorum. Executive Secretary Wersto read the conflict of interest statement. All Panel members were found in COI compliance, and a waiver was issued to Dr. Carrino. Chairman Glassman had the Panel members introduce themselves. He then opened the floor to any audience members wanting make a public comment on clinical use of CAD. Seeing none, he called on Dr. Nick Petrick.

Dr. Petrick presented on ROC. He said that it is important to look at the clinical decisions that are made as well as how the technology may impact clinical practice. It is possible to obtain both the rating as well as the action item in the same study, so it is not necessary to choose between sensitivity/specificity and ROC types of measures. He gave the example of the Jiang study, in which both ROC curves and sensitivity/specificity operating points were studied. The study can get data on the ROC curve as well as individual operating points for a multi-reader, multi-case (MRMC) study design, and the study can be prospective or retrospective.

Dr. Dodd pointed out that C-RAD would not work with a ROC analysis, since there are too few points. She said correlation must be considered when there are multiple lesions or sites per patient. The labeling should clearly define how truth is defined.

PANEL DISCUSSION: COLON CAD (CONTINUED):

Because the questions on colon CAD had not been completed the previous day, the Panel moved to the discussion questions.

C5. Please discuss whether there are other types of performance testing you believe should be considered in the clinical evaluation of colon CAD devices.

Panel consensus was that stand-alone and reader studies are the tests to consider.

C6. Please provide comments on the practice of using an enriched dataset for the clinical evaluation testing discussed in questions C3-C5.

- **a. If you believe that the enriched dataset may be used for these evaluations, please discuss what you believe to be the appropriate clinical and mammographic characteristics (or range of characteristics) for that database.**

Please consider such items as:

- **i. Proportion of patients having polyps;**
- **ii. Proportion of patients having multiple polyps;**
- **iii. Polyp size.**

b. If you believe that enrichment is inappropriate, please provide your reasons and whether there would be an alternative method of assessing these devices.

Panel consensus was that enrichment and stress testing are appropriate for the stand-alone and the reader study, but in the reader studies, enrichment should be at a lower level to better reflect clinical practice. The Panel did not specify the proportion of polyps, but it said multiple polyps would not be important. Polyps in all areas of the colon, including flexures, should be included, and polyps between 6 and 10 mm and greater than 10 mm should be included in an adequate test. .

C7. FDA does not specify indications for use, but reviews indications for use that are requested by companies. What are the Panel's views regarding second reader versus concurrent reading using a CAD device? Specifically,

a. How are colon CADs used clinically?

b. Are second reader and concurrent reading modes both clinically relevant options for use in practice? If not, which paradigm(s) are appropriate for colon CAD devices?

c. Do you believe users understand that if a device is labeled as a second reader, they (i.e. the physician) should always read the radiological image completely before turning on the CAD?

Chairman Glassman opened the floor to public comment. **Roel Truyen** of Philips Healthcare addressed the Panel on clinical use. He said both concurrent and second read paradigms should be available, depending on the manufacturer's claims, but that devices should not be tested for use outside of the indication. **C. Carl Jaffe, M.D.**, from NCI, pointed out that the ACRIN trial has been completed. It has 2600 cases in it and will be published soon. Some secondary endpoints in the study addressed first read and second read. **Maha Sallam, Ph.D.,** from iCAD said it is overly burdensome for the industry to have to test for methodologies other than the intended use. **Dr. Berry** asked how the ACRIN trial tested for both concurrent and second read, since that would imply a different study design. **Dr. Jaffe** said reading paradigms were secondary aims of the study and that the protocols are available.

Panel consensus was that colon CAD is used both sequentially and concurrently. Panel consensus was to make no recommendations about the equivalence of second read of concurrent read in the absence of comparative data. Industry members can apply for their desired indication and provide data. Both second reader and concurrent reading modes are clinically relevant, but there are insufficient data to show equivalence or superiority. Even if physicians understand that a second reader device requires that the image be read before the CAD device is used, that does not mean the physician will follow the protocol. Labeling and training could support compliance. Dr. Tourassi asked what is to be done if a sponsor produces a second reader, then wants to clear it for concurrent read. Chairman Glassman said it would depend on the state of the scientific evidence and literature supporting equivalence. Ms. Brogdon asked if there should be further labeling discouraging or giving more information for off label use. Chairman Glassman doubted

the benefit of stronger labeling without scientific evidence showing inferiority. Dr. Abbey suggested tracking off-label use through post-approval studies. Ms. Brogdon said that could not be required. Dr. Berry called the Panel's attention to a paper in JAMA about flat growths and colon cancer. .

FDA PRESENTATION: LUNG CAD DEVICES

Sophie Paquerault, Ph.D. gave the FDA presentation on lung devices. Chest x-ray is the most commonly performed radiographic exam and is routinely performed to screen for pulmonary metastases, monitor cancer therapy, or to determine the cause of chest pain, shortness of breath, fever, or trauma. Chest x-rays are commonly comprised of the posteroanterior view and a lateral view. The most common test to clarify chest X-ray findings is CT examination. CT imaging detects smaller and more subtle findings and allows imaging guidance for biopsy. Chest CT imaging uses X-rays and computer processing to produce hundreds of cross-sectional images of the chest with detailed anatomic and pathologic information.

Screening high-risk patients for lung cancer is a topic of discussion in the medical field, but it is questionable whether early intervention in high-risk patients is sufficiently effective to justify screening large asymptomatic populations. The International Early Lung Cancer Action Project (I-ELCAP) is an ongoing clinical trial to justify screening high-risk patients using CT. The National Lung Screening Trial (NLST) is an ongoing clinical trial to determine whether chest x-ray or CT is better at reducing deaths from lung cancer and to examine the risks and benefits of CT compared to chest x-rays. Published studies show a variation in lung cancer detection with CT of 67 percent to 100 percent sensitivity and 50 to 95 percent specificity. From 25 percent to 50 percent of patients undergoing chest CT will have a lung abnormality, malignant or benign. A positive CT result typically results in additional radiation dose, invasive testing, and increased anxiety.

Early lung cancer typically appears as a nodule or mass-like opacity on x-ray or CT. On chest x-ray, a nodule is defined as a spherical opacity 3cm in diameter or less and surrounded by lung parenchyma. On CT, a nodule is defined as a round opacity, at least moderately well-marginated, and no greater than 3cm in diameter. It is typically described as completely solid, partly solid, or non-solid. It is difficult to distinguish benign from malignant nodules on imaging, and a solitary nodule could mean many different diagnoses. Nodule management is directed by criteria on nodule size and characteristics. X-ray prints with non-calcified nodules are referred to CT. On CT, nodules less than or equal to 4 mm are generally disregarded for low-risk patients. Follow-up CT is recommended for nodules greater than 4 mm. Biopsy is recommended for nodules that grow over time or exceed 8 mm.

X-ray and CT are used to diagnose various diseases and to evaluate many structures, so the radiologist relies on history and symptoms. Searching for nodules on an x-ray is complex. On a CT, searching for lung nodules is simpler, but it is burdensome due to the number of images. Patients with a single nodule seldom have symptoms, so the detection is in the context of searching for other findings. Detection varies greatly due to physician experience, nodule size, location, and the technique. Radiologists fail to detect lung nodules in chest x-rays in up to 50 percent of cases in which nodules are

visible in retrospect. Sensitivity and specificity of pulmonary nodule detection using low dose CT is comparable to conventional CT but varies by nodule size, location, and vessel proximity. The potential for detection improvements is greatest in small nodules, nodules in the hilar or central lung, and nodules adjacent to blood vessels.

Unlike mammography and CTC CAD devices, lung CT and x-ray CAD devices are intended to detect only one of the numerous diseases that may be present. One abnormality chest CAD devices can be designed to detect is solid pulmonary nodules. Like other CAD devices, it prompts to areas of the lung for physician consideration. Because many normal structures look like nodules in chest images, achieving a low number of false positives is a challenge. CAD devices have been reported with up to 5 false positives per view for x-ray and up to 10 per patient on chest CT. CAD can affect physician interpretation of images. It is important that the radiologist is not distracted from other important findings. Even when performed for high-risk patients, x-rays and chest CT exams should always be fully evaluated for findings unrelated to lung cancer. Studies have shown that CAD prompts may induce less visual search for abnormalities unrelated to the findings the device is designed to detect and that even CAD prompts that point to the target lesion without false positives fail to counteract satisfaction of search in chest x-ray.

Ground truth can be established as whether or not the patient has a chest abnormality, whether or not the patient has pulmonary nodules, whether or not the patient has cancer nodules, the location and extent of each nodule, and the nodule description. For ground truth by nodule type, pathology is necessary to determine the nodule type, and not all nodules undergo biopsy. For nodule location, an expert panel is used.

Standalone performance testing is derived by comparing the location of a CAD mark to the nodule location, as determined by an expert panel, and determining whether the CAD mark overlaps ground truth. The results of standalone testing are highly dependent upon nodule size, nodule location, CT protocols or type of x-ray acquisition technique, and co-morbidities that affect chest imaging. Overall standalone measures include sensitivity and number of false positives per scan and plots of the full FROC curve. Stratified standalone performance measures provide an understanding of the benefits and drawbacks of CAD systems.

Since most chest examinations are for symptoms unrelated to cancer, the device, if it is to be used in the general population, must encompass a full interpretation of all imaging findings. A reader study could be designed to simulate a field test. Following the least burdensome approach, a retrospective study reader may replace field tests with MRMC studies. Such testing may not reflect real-life assessment and may introduce bias.

Chest CAD may fit into practice as a second or concurrent reader. Concurrent CAD may reduce reading time. Reader detection accuracy increases with both second reader and concurrent reader use with chest x-ray CAD. It may be that all reading paradigms should be tested to account for user needs and software controls. The testing dataset for general population use should include the majority of cases with no nodules and other test diseases and be enriched with cases having different numbers, sizes, and morphologies of nodules. For a specific intended use, the testing database should be enriched with cases having pulmonary metastases and cases with a low number of nodules of various size and morphology. Stress testing may include a majority of very

difficult cases, including smaller nodules, nodules that are partly solid or have a ground glass opacity, and nodules in difficult locations. The study endpoints depend on whether the device is for general or specific use. It is difficult to determine if lesion, region, or patient-based analysis should be used as the primary endpoint for chest CAD devices. The clinical action following chest x-ray or CT may relate to many chest diseases and is location-specific. It may not be possible to adequately extrapolate from only one endpoint and provide clinically significant results using the device. Co-primary endpoints may be established.

Dr. Berry asked if it is possible to lower the false positive rate. **Dr. Paquerault** said use as a second reader can increase the false positive rate. Reduction would depend on the device and reader testing. **Dr. Tourassi** asked about controls. **Dr. Paquerault** said the software doesn't have controls to ensure that the device is used as a second reader.

OPEN PUBLIC HEARING: LUNG CAD

Dr. Eric Silfen of Philips Research said CAD systems are tools for clinical decision making that indications and contraindications. The purpose of standalone testing is to create expert systems. Reader performance testing addresses clinical use of the expert systems. He stressed that the two are distinct ways of determining the value of a system. The performance characteristics of the devices will vary, depending upon their field of use, and screening and diagnosis are very different uses. It is important that physicians know the sensitivity of the test. His time expired before he finished his presentation.

Joe Gardill from Healthcare Reimbursement RX referred the Panel to a study published in the Biomedical Imaging and Intervention Journal by Taylor, et al, comparing second read and concurrent read in CTC CAD. He said that cost is an important issue in adoption of technologies and that protocols affect cost. The valuation necessary to drive adoption of CT or MRI CAD as a second reader would be an economic burden on the payers. This would lead to wide scale off-label use as a concurrent reader. Were the devices tested and labeled as concurrent readers, the devices would be used as indicated.

Alok Gupta, Ph.D., from Siemens Healthcare commented that CT CAD devices have matured. Although some exams have up to 10 false positives, most exams have few false positives. Siemens' PMA studies showed an average of around 2 false positives.

Larry Clark from NCI asked if a consensus-formed reference database could be recognized as the first step in the process of approval of CAD tools.
David Naidich, M.D., a consultant for Siemens, said mammography CAD, colon CAD, and lung CAD are distinct. He said most cases do not require CAD, and there should not be a protocol requiring CAD use where it is not needed. **Dr. Carrino** asked about pulmonary CAD reader paradigms. **Dr. Naidich** said that the device is optimally directed toward cases in which a nodule is suspected and not found.

PANEL DISCUSSION

Chairman Glassman said lung biopsies are more serious than breast biopsies, so false positives are critical. He further commented that a concurrent reader paradigm created the possibility of the reader being distracted by CAD markings. **Dr. Steier** said that lung cancer screening is not as common as breast and colon cancer screening. The false positives are a concern. However, more nodules will be found with the device, and there is data supporting both second reader and concurrent reader. .

L1. Establishing ground truth (i.e. whether disease is present and, if so, its location and extent) is crucial for the evaluation of performance of any CAD device. Please provide your recommendations for defining ground truth for lung CAD devices.

Panel consensus was that ground truth should be defined by an expert panel, both for detection and diagnosis, using the full knowledge that they have, including pathology, follow-up CT scan for an abnormal chest X-ray, PET scan or follow-up CT scan for an abnormal CT scan. A full reading of the study for non-nodule ancillary findings will not be necessary. Dr. D'Orsi asked that the FDA provide the Panel with the indications of approved lung and chest CAD devices. Dr. Petrick presented on the indications. For x-ray and CT, the devices identify features associated with nodules and are used for second reading. They are detection devices and are not cleared for screening. .

L2. Please discuss the role of standalone performance testing in the clinical evaluation of lung CAD devices.

a. If you believe standalone testing should be requested in the evaluation of these devices, please provide your recommendations or comments on whether certain substrata (e.g. nodule size, shape, pathology, location; comorbidities; CT dose and imaging protocol; or others) should be considered in device testing and labeling.

b. If you believe that there are specific situations where standalone performance testing may not be important, please comment on what those might be.

Panel consensus was that standalone testing is necessary. There are substrata to be considered in device testing, and which substratum depends upon the intended use of the device. The substrata include nodule size (4 to 30 mm for CT, 10 to 30 mm for x-ray), nodule shape, nodule density, and co-morbidities such as air-space consolidation, mild to moderate interstitial lung disease, and emphysema. The Agency will have to develop criteria based on the indication. If the CAD package includes growth and management, phantom testing will be part of validating those functions.

L3. Please discuss the role of reader performance testing in the clinical evaluation of lung CAD devices.

a. If you believe reader performance testing should be considered in the evaluation of these devices, please provide your comments or recommendations on:

i. The appropriate primary endpoints and corresponding clinically significant effect size(s). Please specifically comment on ROC analyses;

ii. The merits of per lesion, per region and/or per patient endpoints in the assessment of endpoints;

iii. Whether reading time should be assessed, and if so, how.

b. If you believe that there are specific situations where reader performance testing may not be necessary, please comment on what those might be.

Panel consensus was that performance testing is critical for reader performance testing. The Panel had no firm recommendations on primary endpoints and clinically significant effect sizes. Endpoints will depend in part on the intended use. ROC analysis is important for diagnostic use. For detection, JROC or FROC should be used. The study should be powered to give information for those analyses. Per lesion information will be important to evaluate for false positives and false negatives. Reading time was not considered critical to the prospective users of the device.

L4. Please discuss whether there are other types of performance testing you believe should be considered in the evaluation of lung CAD devices.

Panel consensus was that there are none at this time.

L5. The prevalence of lung cancer cases in the population having chest X-rays and chest CT is relatively low. Please provide comments on the practice of using an enriched dataset for the clinical evaluation testing discussed in L2-L4.

a. If you believe that an enriched dataset may be used for these evaluations, discuss what you believe to be the appropriate clinical, imaging and pathological characteristics (or range of characteristics) for that database. Please consider items such as:

i. Numbers of patients with no nodules, single nodules, or multiple nodules;

ii. Range of nodule sizes.

b. If you believe that enrichment is inappropriate, please provide your reasons and whether there would be an alternative method of assessing these devices in light of the low prevalence of disease.

Panel consensus was that enrichment is appropriate. For standalone testing, a high prevalence of abnormals with nodules from 4 to 30 millimeters should be included, as well as a sufficient number of normals and benigns, including nodules smaller than 4 millimeters, scarring, sequestrations, and other things that could easily be confused with a nodule. The nodule types should come from multiple pathologies, including carcinoma and infection, sarcoid, septic potentially, and metastatic multiple nodules. Dr. Garra expressed concern about the scoring; pointing out that a 2 mm nodule detected by CAD would be scored as a false positive, even though the finding was accurate. Dr. Berry added that there should be improvements on the ROC analysis to take care of it. .

L6. FDA does not specify indications for use, but reviews indications for use that are requested by companies. What are the Panel's views regarding second reader versus concurrent reading of a CAD device? Specifically,

a. How are lung CADs used clinically?

b. Are second reader and concurrent reading modes both relevant options for use in practice? If not, which paradigm(s) are appropriate for lung CAD devices?

c. Do we believe that users understand that if a device is labeled as a second reader, they (i.e., the physician) should always read the radiological image completely before turning on the CAD?

Panel consensus was that CAD is used both sequentially and concurrently and that both options, based on the available science, are reasonable. Even if physicians understand the labeling, they may not follow it. Further training of users may help with compliance. Proper labeling should match the current scientific data.

L7. Chest x-ray and chest CT are done for many important reasons other than looking for lung nodules. Can the use of CAD affect the diagnosis for these other conditions? Can the presence of other conditions alter the effectiveness of the CAD function or the risk-benefit profile of the lung CAD device? If the answer to either of these questions is "yes," then are there specific conditions that should be represented by patients in the test database?

Panel consensus was that CAD can affect the diagnosis of other conditions. The presence of other conditions alters the effectiveness of the lung device. The effectiveness of CAD would be affected by the presence of other disease and trauma, so diseases and trauma should be represented in the test database, but the database should not be enriched for the subsets. Those diseases include sarcoid, septic emboli, pneumonia, scarring, air space consolidation, interstitial disease, emphysema.

FDA PRESENTATION: FUTURE ISSUES WITH CAD

Anastacia M. Bilek, Ph.D. said that the remainder of the discussion was to focus on CAD in general applicability of the prior discussion to other types of CAD, insight into future CAD or CAD-like devices. She asked the Panel to discuss the extent of information manufacturers should provide regarding their algorithm, its training, and its stability. For reader study designs including multiple readings of the same cases, she asked the Panel to discuss reducing the bias created by recall.

In reader testing, the control has typically been an unaided single reader. Alternative controls include unaided double reading by a single reader, unaided double reading by two readers, and reading aided by a sham CAD. She asked that they discuss alternative controls.

She asked them to discuss the reuse of a test dataset in the evaluation of subsequent algorithm revisions. She said the ideal approach is to develop a CAD algorithm, collect test cases, and then perform standalone and reader performance testing,

keeping testing isolated from training. However, companies may want to use the same test cases or an expanded version of the same dataset to compare performance.

Use of small enriched datasets leads to study populations that do not match the target populations. She asked the Panel to discuss standardizing the statistical analysis according to a designated standard distribution of clinical variables.

She asked the Panel to discuss the methods of demonstrating safety and effectiveness and their applicability to other CAD devices. There is the possibility of applying the type of functions radiologists find useful, including simple display functions and complex evaluation tools or computer prompting tools. Though the meeting focused on CAD detection devices, diagnostic devices are on the horizon, either in physician-identified candidates or computer-identified candidates. She asked the Panel to discuss the applicability of the evaluation methods discussed during the meeting to such future computer-based technologies.

Future CAD detection and diagnosis devices will search for cancer in other parts of the body or use different imaging modalities. CAD could guide biopsy or identify non-cancerous abnormalities. Potential CAD developments may monitor responses to therapy or provide diagnostic assessments.

OPEN PUBLIC HEARING

Dr. Akira Hasegawa of Fujifilm Medical Systems presented on CAD evaluation by ROC. The CAD under discussion was second read. One of requirements for clinical endpoint of CAD approval is to demonstrate a statistically significant improvement of the reader ROC or FROC curve by using CAD. He questioned the logic of using the ROC or FROC to evaluate the effectiveness of CAD for second read. Second read CAD helps readers reduce oversights, not cognitive errors, so if the readers do not make oversights, CAD has no effect. Failure to show a statistically significant improvement can be due to lack of reader error, and reader oversight is a random event. Therefore, lack of statistically significant improvement of ROC or FROC does not mean that CAD did not work as expected. .

Roel Truyen of Philips Healthcare and **Stephen Slavens** of GE Healthcare presented jointly on CAD submission data requirements. Mr. Truyen said that both companies are MITA members. He said submission data should provide scientific evidence for the claims made of the device. The type of study should rely on the claims, and state of the art methodologies should be applied. MITA supports the use of standardized methodologies, but until they are developed, sponsors may select the least burdensome approach, and the choice of methodology should lie with the sponsor. He said that extending the study beyond the claims is unnecessary. He cited colon CAD as an example. His time expired before he completed his presentation.

Mr. Slavens noted that the Panel had supported the use of registry data in device training and submission studies. Current requirements include informed consent, and exceptions for de-identified images in repositories are not allowed. Locating the patients is nearly impossible and overly burdensome. Retrospective studies do not impact diagnosis or treatment. MITA proposed that FDA apply the principles in "Guidance on Informed Consent for In Vitro Diagnostic Devices Studies Using Leftover Human

Specimens that are Not Individually Identifiable" to de-identified data and permit IRB discretion to waive informed consent for de-identified retrospective cases. FDA should also inform the sponsors on the required confidentiality and informed consent procedures and records.

Pat Milbank, a consultant for Medipattern, spoke for MITA. She said that labeling or warnings on off-label use may be required but that it is inappropriate to require that sponsors study their products for off-label use. She asked the Panel to clarify that off-label studies are not required for approval of their devices. Second, she requested that the Panel consider whether the effectiveness of washout studies is appropriate for use in CAD reader models. She said that requirements for multiple arm studies with various washout periods do not satisfy the least burdensome requirement. The standard study design recognized in the literature for most CAD products involves real-time, sequential readings conducted in a simulated clinical use environment. Her time expired before she finished her presentation.

Maryellen Giger, Ph.D., FAAPM, of the University of Chicago appeared on her own behalf and disclosed financial relationships with NIH, the Army, and R2/Hologic. The potential of CAD is expanding 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. Computer image analysis is becoming an integrated step in the diagnostic decision-making process. In computer-aided diagnosis, the computer's output is used to characterize a lesion and potentially indicate a computer-generated probability of malignancy. Systems are beginning to incorporate known databases and indices of similarity. CAD not only reduces disagreement among radiologists but increases efficiency of the reading. She asked the Panel to consider the potential for computer aided diagnosis as a concurrent read. The final decision would be made by the radiologist who interprets the tools, so standalone performance must be separated from user performance. She spoke on a potential technology assessment institute that could be tasked with performance assessment of new or improved CAD devices. The tests would be standardized and standalone, using datasets from the Institute's database.

Steven Vastagh of MITA said industry is offering to work with FDA on evaluating CAD issues and developing a guidance document and resolution.

Dr. Carrino asked about alternative methodologies to ROC and the washout period, which is an established tool. **Mr. Truyen** said an independent read is not needed for a second reader paradigm. A washout period may be needed for a concurrent or first reader paradigm, but the length of the period is not determined. He said that standalone testing using phantoms can be sufficient alternatives to ROC. **Dr. Steier** asked about training, quality assurance, and separating device performance from user performance. **Dr. Giger** said training should be run by academics to prevent bias and should be integrated into residency. QA is needed for CADx devices. She said that both the tool and the user are part of device use, so testing differences in radiologists' abilities will affect the test results. **Dr. Mittal** asked about the technology institute. Dr. Giger said it would be a

private nonprofit that would collect data and assess CAD. **Dr. Abbey** asked who would be responsible in the case of wide off label use. **Ms. Milbank** said that FDA cannot require studies on off-label use. However, as uses evolve, studies will have to be done.

PANEL DISCUSSION: GENERAL METHODOLOGIES

Chairman Glassman opened the floor for discussion. **Dr. Berry** said, in reference to Dr, Hasegawa's presentation that statistical significance is necessary in order to give proof of effectiveness and that .05 is a weak criterion. **Dr. Carrino** said that IRB waivers for de-identified data are appropriate. **Ms. Brogdon** said that the de-identified data creates problems for FDA's obligation to audit data. FDA encourages companies to discuss informed consent with FDA early on. **Dr. Steier** said that competency, training, and off-label use are important and must be fleshed out. **Dr. Bourland** said that virtual phantoms from manipulation of digital image can be used in standalone testing. As a future device, he suggested a device incorporating imaging, diagnosis, and treatment, such as a lung CAD with an ionizing radiation or ablative capability. He wondered about the appropriateness of CAD use in underserved populations. **Chairman Glassman** said that CAD is currently second read, though that may change as the data changes. **Dr. Rosenberg** said that the dividing line between CADe and CADx is an interesting question. **Ms. Finken** raised the issue of accessibility of the equipment to the underserved, uninsured, and rural populations as a future issue.

G1. 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?

Panel consensus was 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 to evaluate the device and subsequent algorithm updates.

G2. What may be appropriate constraints on the reuse of test data in order to balance data integrity and data collection for CAD assessment?

The Panel had severe concerns about reuse of test data. Optimally, a new test set should be obtained. However, there will be circumstances in which that would be unnecessary or inordinately burdensome to do so. In those circumstances, FDA may accept a partial use of the former test data.

G3. 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?

Dr. Tourassi said that sequential reading should be sufficient for the second reader paradigm. When there is not sequential reading and the paired design is used, there

should be a washout period (30 days is the standard supported by the Panel). The Panel did not have suggestions for improving paired design studies. Dr. Gwise asked about the possibility of bias in a sequential read and possible controls. Dr. Abbey said that there could be a washout period between the unassisted and the CAD read but that might be burdensome. Dr. Tourassi said that whether or not there will be CAD after the first read can be randomized. Dr. Spindell pointed out that CAD use is not random in clinical use. Chairman Glassman said the issue depends upon the significance of the bias. If the bias is insignificant, the burden is excessive. Dr. Berry said the bias is unknown, since there is no data. It could be of any magnitude and in either direction. Dr. Ziskin said the bias will reflect reader behavior in a clinical situation, so it can be ignored. Chairman Glassman suggested a trial to look for bias. Dr. D'Orsi suggested a time limit for the read. The committee did not reach a conclusion on that question.

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

Panel consensus was that the appropriate control group is a group of radiologists who practice as the standard of care without CAD. The control group should have 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.

G5. 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.

Panel consensus was that there are many problems with weighted analysis and that they are not in favor if it as a statistical test. The initial group of patients should be properly accrued so that weighted analysis will not be necessary, though that approach may be more burdensome in some cases.

PANEL DISCUSSION: FUTURE ISSUES WITH CAD

F1. 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 (CADx) devices, compared to the types of testing you have discussed in this meeting?

Panel consensus was that there would be a need for pathologic correlation in a higher percentage of cases. There would be other surrogates such as time of stability or negative PET scan, but an enriched dataset with a number of cancers would need a lot of pathological proof. The colon does not lend itself to CADx over CADe, since the task is merely to identify soft tissue polyps. Chairman Glassman said there should be a greater emphasis on pathology and follow-up. He urged an examination of clinical guidelines and pathways for head trauma and ankle trauma diagnoses. He noted that the bar would be much higher for diagnosis than for detection.

F2. What emerging CAD areas should FDA be aware of? Do you have comments on the types of testing needed for other possible CAD devices--present and future--compared to the testing you have discussed in this meeting.

Panel consensus was that potential emerging areas of CAD are studies that have kinetic significance, studies to check easy cases in mammography, the fatty breast cases, liver cancer, tomosynthesis, breast ultrasound, renal carcinoma, pulmonary embolism, and things involving temporal change. The Panel members preferred to not suggest different testing methods or statistical analysis without knowing the potential devices' proposed indications.

F3. Do you have any comments on the levels of testing for the different types of computer-based technologies compared to testing that you have discussed in this meeting?

Panel consensus was that there are many types of near-CAD technologies coming down the road, some of which, based on their improvement of image quality, will need human eye testing. Other may only need phantom testing or standalone testing. The testing required will be device and function-specific.

ADJOURNMENT

The agenda completed, Chairman Glassman thanked the Panel and the FDA and adjourned the meeting at 5:03 p.m.

I certify that I attended this meeting on March 5, 2008 and that these minutes accurately reflect what transpired.

_____

Nancy Wersto
Executive Secretary

I approve the minutes of this meeting as recorded in this summary.

_____

Leonard M. Glassman, M.D.
Acting Chairman

Summary Prepared by

Eric M. Hendrixson
Neal R. Gross & Company
1323 Rhode Island Ave, NW
Washington, DC 20005
(202) 234-4433
Wednesday, March 26, 2008