action may be necessary.

Follow-up CT scan might be prudent at anything from three months to 12 months depending on the nature of the nodule and the radiologist level of suspicion. Other kinds of imaging studies such as a PET scan may be applicable, especially in larger nodules that are in the range of 8 to 10 millimeters. This may distinguish cancer from a benign nodule, Finally, we can consider biopsy, either transthoracic needle biopsy, bronchoscopy, or thoracoscopic resection.

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

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

nodules and I'll just cite two particular studies of interest that have addressed this issue of missed nodules and CT scans.

Dr. Hartman and others at the Mayo Clinic looked at over 1,000 screening CT scans and compared them with prior screening CT scans one year earlier to see how many nodules may have been overlooked. They found that as many as 24 percent of the prior prevalent scans had nodules that were not recorded at that time.

This might seem an astonishingly large number but this is consistent with some other studies. Now, a large number of these nodules were relatively small put more than one-third of them were about three millimeters and in the size range where they are likely to be considered actionable.

And, in fact, 6 percent of them had grown which would mean that they were highly suspicious for lung cancers so there seems little doubt that nodules are being missed even in excellent centers such as the mayo clinic in a study that was focusing specifically on the detection of nodules.

One other study performed by Gruden and others at Emory University looked at 25 patients with presumed lung metastases. These patients had soft tissue sarcomas and melanoma and they established truth by consensus which is a practical method using five readers. These nodules were three to nine millimeters in size and they were solid nodules. Two to nine solid nodules in each case by consensus.

They found that the miss rate for individual readers ranged from 20 percent to 39 percent of all of the nodules in this size range. This was in an observer test setting where the readers were focused on detecting nodules and presumably had no other task in mind so one would expect a relatively good performance in that situation.

So between these two studies we can see that there is a considerable problem with oversight errors in reading CT scans. Now we have a trend towards thinner CT sections with the newer multi-detector scanners. This allows improved ability to detect and characterize lesions. It does allow us to do a high quality off-axis reconstructions.

On the other hand, it does present us with more image data, more opportunities for error. In a chest CT scan performed with a multi-detector unit we may have anything from 18 to almost 300 images of the chest and the radiologist has to interpret those visually.

I think that the evidence that we've seen strongly suggest that traditional visual interpretation is no longer sufficiently reliable for detecting these very small and potentially dangerous common nodules.

At this point I would like to introduce Ronald Castellino, Chief Medical Officer for R2 Technology.

DR. CASTELLINO: Thank you. My name is Ron Castellino. I'm also a diagnostic radiologist but currently I'm the Chief Medical Officer of R2 technology.

At the outset I'd like to particularly emphasize the definition of computer-aided detection which is also called CAD as we will be using it in the presentation today. Computer-aided detection as we

use it refers to the availability of computer algorithms that automatically identify regions of interest on a medical image for the radiologist to evaluate.

It's purpose, of course, would be to decrease what I would term observational oversights. That is, findings that are present on the image but, in fact, are not seen by the radiologist. This is not a device to tease apart very unusual nodules that might not be present or barely present on the image. These nodules are actually clearly visible on the image.

The image check for CT CAD system specifically is designed to automatically detect regions of interest with features suggestive of solid pulmonary nodules on CT exams of the chest. It's important to remember that it is to be used as a supplemental review. That is, after the initial assessment has been made by the radiologist. It is not a first reader.

The radiologist, most importantly, remains responsible for the final interpretation of the

findings that the CAD marks may put on the image. That is, to determine if the mark is actually a true mark or if it is a false mark.

A brief review of the device description. The CT scan is performed in the standard fashion. The images or the data set is moved to increasingly types of work stations that radiologists review the images on and what is what we call a soft copy display. These images may be reviewed slice by slice but increasingly they are reviewed in some type of a melt-through or a cine mode to facilitate reviewing these hundreds of images that are generated.

By the same DICOM standard the data set can also go through a server computer. Various image analysis algorithms can be put into place. In this case, I point out segmentation. This type of information can also be transmitted to the work station to help the radiologist further analyze the images and this is an image checker CT work station which was cleared by the FDA in 2002. This is an existing product that has been cleared.

The same DICOM data set can also go

through an image checker CT CAD software system and provide on the work station CAD information as well. It is this specific piece of the product that is under review today by the panel.

I'll show you a few screen capture images of the front end of the work station on which the CAD marks are displayed. The view port on the right is familiar to radiologists. This is where we can see the axial images. I guess I can't use this thing. Thank you. We are a high-tech business as you can see.

There we go. On the large view port on the right we can see the axial image displayed to the radiologist which is viewed either singularly or, like I said, melt-through a cine mode. The smaller view port on the upper left is a three-dimensional reconstruction of the contents of the lung.

You can see the pulmonary vessels. In fact, a few nodules perhaps you can see there. And the horizontal lines simply indicates to the radiologist what level on the image the axial image is displayed. We see a nodule here quite clearly in the

right apex.

The radiologist then will move down the entire sequence of the lung in the lung windows looking for other abnormalities, nodules as well as a multitude of other features that the radiologist searches for sometimes seeing nodules and sometimes not seeing nodules.

When they completely review the entire study, which I'm giving to you in a very schematic fashion here, the radiologist then will activate with a mouse click the CAD button we call the R2 button. At that point in time the CAD process takes over and presents the following.

The circles indicate candidate nodules that the CAD system has identified shown to the radiologist on the three-dimensional display of the lungs, as well as brings the radiologist automatically to that specific site where the nodule is best seen by the CAD system.

In addition, out other view port on the lower left is shown. This is a three-dimensional reconstruction that can be rotated to separate the

nodule out from adjacent vasculature. I would like to emphasize that upon the CAD review the radiologist need not go through the entire data set once again but simply by moving and hitting one of these little buttons here with a mouse click which you can't read here. It automatically jumps the image. By the way, the size is automatically shown as well.

It automatically jumps the image to the next CAD detected nodule and the next and so forth. For example, this nodule, as I showed you and, for example, a nodule at the right base which is clearly a nodule but, in this case, had been overlooked by the radiologist on the set of images.

That is the CAD display on the work station. What does the CAD search for? It is specifically designed to search for solid lung nodules that are 4 mm. or greater in size and we find that further as follows. They should have an approximate spherical shape.

The margins can be smooth, lobulated or spiculated and should have soft tissue density which we define as having average density of minus 100

Hounsfield units or greater. Some of the typical CAD marks you've seen already. They circle the nodule. We consider this a true mark if it actually encompasses the size of the nodule sometimes quite small, moderate in size.

I would like to emphasize that also although we look for spherical nodules if, in fact, the nodule is adjacent to a plural surface where a portion of the sphere is obliterated by contract with the plural surface. The algorithm tries to find these as well.

Secondly, this image perhaps some of you can see, although it is easier for the radiologist and the CAD system to detect a nodule that is surrounded by completely normally aerated lung, if there is adjacent modest non-aerated lung as we see here in the appended edema, the CAD algorithm often is successful in teasing out the nodule as well.

There are a multitude of other parenchymal abnormalities within the lung tissue that the CAD algorithm does not search for. The radiologist must look for these but the CAD algorithm does not search

for. For example, linear strands which do not fit the criteria. I would like to point out importantly although this fits the criteria of being a spherical nodule, we call these ground glass opacities.

They are increasingly noted to be of importance, particularly for lung cancer screening programs that because of the Hounsfield density cutoff that we have, this type of nodule currently is not searched for with our set of algorithms.

All CAD systems have false marks. We see a few here such as this one here where a branching vessel exist. The CAD algorithm thought this was a nodule and marked it incorrectly. Plural tags are at times marked incorrectly. I can tell you that our experience internally as well as with users indicate that the vast majority of these false marks can be readily dismissed as you see here.

As an aside, we have found that a regulatory database a median of three false marks per exam. I would like to emphasize this is per exam. There is a median of 160 images per exam so we're talking about approximately one false positive mark

for every 50 to 55 individual images.

Now, the clinical study was designed around an ROC study as you've heard from Dr. Wagner. It was done in close collaboration and support with the people from the FDA. The ROC study in a large extent does measure -- a combined measure of efficacy of safety. There is some discussion about that and Dave Miller will fill you in on that as we see it, at least.

There are three parts. We've collected cases. I'll review that. These cases were sent to a reference truth panel and finally to the MRMC ROC study which you'll hear about from Dave Miller.

I would like to spend only a brief comment upon the target of nodules. You've heard from Dr. MacMahon that we are increasingly seeing smaller nodules on our CT scans and our clinical practice. We wanted to design a CAD system to help radiologist detect all solid nodules between 4 and 30 mm. That was the focus of our research effort.

And, as you are well aware, those in the clinical practice you will recognize that most lung

nodules most of the time are typically sampled by biopsy or thoracic resection if they are 8 or 10 mm. or so greater in size. There are obviously exceptions to this but, in general, they are.

The availability of a biopsy proven so-called gold standard to evaluate nodules in this smaller size range was just not available to us. We settled on a gold reference standard of a consensus on actionability as being the only practical standard that would capture all solid nodules of clinical concern in this size range. We are really focusing and trying to help the radiologist in the 4 to 8, 10 to 12 mm. range. The larger nodules, of course, radiologist will almost always see.

We collected cases from five centers. They contributed consecutive non-selected cases. We tried to make this as representative as possible. They were all in adults. They were performed for a variety of clinical indications. There were no screening studies in this group.

Cases with greater than 10 nodules were excluded. We felt that there were a multiplicity of

nodules. The issues of searching for nodule where the radiologist has already seen 8, 10, 12, 15 would be reported. The images, of course, have to reach certain technical parameters.

These cases were divided into two categories to begin with by report. The nodule-present cases had in the report the presence of one nodule or more described by the reviewing radiologist. These patients by definition had a history of biopsy proven documentary cancer either primary to the lung or in an extra thoracic site.

We did this to try to increase the likelihood that nodules in this group might have clinical significance because they were in patients with cancer but I would like to point out that the specific nodules themselves were not biopsy proven. The nodule absent cases, once again by report, no nodules were described within the context of the report. These patients could have a history of cancer or not.

The final truth was determined by the reference panel which you'll hear about from Mr.

Miller. Five sites contributed to the study. Three of these are community imaging centers, two are university centers. They were from the east coast, mid-west, west coast. There were 63 cases that had nodule-present by report, 88 nodule absent by report.

You can see the distribution between male and females were similar. The age range was similar in the two groups. There was a slight increase in median age in the nodule-present cases perhaps because they all had documented histories of cancer as compared to this group. The type of cancer in the nodule-present case, 38 percent had a documented primary lung cancer and 62 percent had documented extra-thoracic primaries.

Here are some of the parameters of the technical aspects of the case characteristics, the median number of slices you see here. There is a slight predominance of thinner slice sections in the nodule absent cases mainly because one of the centers was doing much thinner slices routinely and they contributed a larger amount of nodule absent cases. The CT vendor's use in these five sites were General

Electric or Toshiba.

I would like to ask Dave Miller to present the methods and the results of the study.

MR. MILLER: Thank you. My name is Dave Miller and I am currently the Director of Statistical Analysis at Ovation Research Group. At the time that this study was conducted I was the Director of Biostatistics at R2 Technology. R2 is paying for my time and travel. However, I do not have any financial interest in R2 Technology.

Just want to quickly go through an outline of what I'm going to discuss because I'll be up here for a little while. I'm going to go through some definitions that I'll be using during the talk. Then I'll talk about the reference truth panel. I'll talk about the ROC study design, our primary analysis. Then we did a large set of robustness analyses. Then finally the study conclusions.

So gold standard, and these are definitions that I'm going to use. They are not necessarily dictionary definitions of these but gold standard is something that I'll define as an objective

and definite measure of truth.

The reference truth is a truth standard for a subjective construct. It is a term that is fairly widely used and it's a term that I'll be using here as a standard that's used in lieu of an available gold standard. The kind of thing that reference truths are used for are things like actionability where actionability is something I'm defining as a subjective point-of-care decision which is really what we're targeting with actionable nodules.

Nodule also is a subjective definition. It's a subjective characterization of a lung abnormality. Finally, a panel is a group of radiologists with a given task. In this case, their task was to identify and characterize actionable nodules. Consensus is a term I'll use only for unanimous agreements. When you hear we use consensus, that means unanimous agreement as opposed to majority agreement.

Then, finally, a few study definitions. I'll run through these very quickly because you've got a very nice tutorial from Bob Wagner this morning.

The ROC curve is the receiver operating characteristics curve. AZ is the area under the ROC curve, the measure of interest in the study.

MRMC stands for multi-reader, multi-case. I'll use the term primary analysis for our protocol specified primary analysis and the term ANOVA-after-jackknife. The ANOVA there is analysis of variance and you've got a nice description of both the jackknife and the bootstrap earlier.

So under the reference truth panel the goal of the reference truth panel was to fully identify all nodules in the case sets. These are the cases that Ron described how they were collected. We wanted them to rate the actionability of any nodules that they found. Specifically we are defining actionable as a nodule that requires surveillance or intervention so it could be follow-up or it could be more of an intervention.

We define the reference truth so that we could use it in the ROC study. The method was to have a panel of three radiologists independent review the cases and we followed a two-path process to reduce

observational oversights.

The reference truth panel qualifications were that they needed to be board certified radiologist, that they had at least six months of reading thin slice which we defined as less than or equal to 3 mm. collimation CT of the chest, and they needed to have experience with reading soft copy.

A total of 11 panelists participated in at least one of the three-member panels that were convened. Just to be clear, we didn't have a single three-member panel because it just would have taken weeks for three people to review the set of cases that we had. We had a succession of panels and there were a total of 11 different panelists that participated in at least one of those panels.

Nobody participated in more than three and obviously nobody participated in less than one. This is how the panels worked. We brought the radiologists in and we put them in three different rooms. This is after a brief sort of training that we gave them prior to going to the three different rooms. They had three different work stations set up and they each

independently reviewed a set of cases. In a typical sessions we had about 20 cases reviewed.

After they had reviewed all of the cases for a given day, and this usually took maybe four or six hour or so, we took the computer files of all of their findings and these are findings of the exact locations and we brought them together to get the union of all findings so that redundant findings were captured and we knew every finding that any panelist had found.

This is a little hard to see up there but we also at this stage excluded nodules that were less than 4 mm. in size or greater than 30 mm. in size. Those were protocol exclusions and we had asked the radiologists not to spend too much time taking precise measurements as they were doing this.

After this there were 95 findings where three our of three of the panelists agreed that it was a consensus actionable nodule. I couldn't say consensus. Three out of three agreed and, thus, there was a consensus that it was an actionable nodule.

Now, there was also a large set where

there was disagreement. Either one out of three or two out of three of the radiologist had identified the finding and the other radiologist either had overlooked the finding or didn't feel that it was an actionable nodule. These went to a second pass.

The way the second pass worked is that after about half hour of prep or so they went back into their individual rooms so they didn't come together and talk about the cases. They each went back to their individual rooms and they had the locations of each of these disagreement findings identified for them. So the second pass went fairly quickly because they didn't need to go through the whole case. They were just looking at and being directed to specific spots and being asked to rate the actionability.

After this there were 47 additional nodules that went into our truth set of unanimous nodules. There was also a fair number that went into what we call the majority group, that two out of three felt that it was actionable, and a minority group that one out of three felt that it was actionable.

Our primary analysis focuses on consensus agreement but we did do some robustness analyses around the majority and minority. I'll be talking about that later but for now I'm focused on the unanimous nodules.

So as a result of this process the eight three-radiologists panels. I told you there was a series of panels. There were, in fact, eight of them. They identified 142 consensus nodules in 65 nodule present cases. You might notice that number 65 is slightly different than the 63 number that you saw earlier. That's because now our consensus panel is the definition of truth for this study.

You can see the size of these findings. The median size was 7.9 mm. and there were a lot of them that were in the 5, 6, 7 millimeter range. The remaining 86 cases were categorized as nodule absent by virtue of not having any of the unanimous nodules in them.

So moving onto the MRMC ROC study, the objective of this study per protocol was to demonstrate that review of CAD output improves

performance of radiologists reviewing MDCT with respect to their ability to accurately identify actionable nodules.

Our outcome measures were AzB. That is, the before CAD area under the curve, AzA, that is the after CAD, the area under the curve and, most importantly, Azdelta. This is basically the difference between the two curves. And the hypothesis in a formal statistical sense -- the null hypothesis was that the mean change in the area under the curve was zero and the alternative hypothesis, of course, is that Azdelta is greater than zero meaning the CAD did have a benefit.

The study was conducted in two phases. We first did a 32-patient study and then after doing that study we had some discussions with FDA and we outlined what would be the appropriate methodology to use for a second study, what the appropriate size for the second study would be based on the type of methodology that was suggested. So I'm going to be talking about that second 90-case study as the focus of this talk.

The reader qualifications for the ROC

study, so this is, again, new set of readers. Don't confuse them with reference truth panel. Completely different people. It would be wrong to have the same people. These people had reader qualifications that they be board-certified radiologists and have at least three months of reading MDCT of the chest.

The basics of the study is that we have 15 readers read all cases. We had 90 cases. Of the 90 cases 48 had at least one actionable nodule and 42 did not have any actionable nodules and that was based on a stratified random sample of our complete set of cases.

There were, of course, four quadrants per case by definition but the important point is that these quadrants, all four of them, were rated pre-CAD and then sequentially post-CAD. The ratings were finally evaluated against the reference truth so the ROC curves were drawn by comparing the ratings which were on a continuous scale to the reference truth established by the panel.

I want to clarify what the unit of analysis is because I know people have a tendency to

want to sort of track the numbers as they go through the slides and see where things add up so, just to be clear, nodules were the unit of analysis for the reference truth. The reference panel was supposed to identify every nodule.

Quadrants -- the quadrant truth was computed from the nodule truth. For instance, if there was a quadrant that had one actionable nodule and one non-actionable nodule, the quadrant was, nonetheless, considered nodule-present quadrant because it had at least one.

On the other hand, if there was a quadrant that had a minority nodule in it, in other words, a nodule that at least one person on the panel thought was a nodule but not unanimous, that was considered a nodule absent quadrant. Every quadrant counted in every analysis that we did.

Now, the reason that we went with this quadrant approach is that the LROC methods were not developed at the time that we embarked on this for multi-read, multi-case studies. I think they probably will be in time and they may even be right now but at

the time we began the study, they were not.

Bob Wagner described it a little bit as these being sort of competing fields that people that went with the ROI approach versus the people that go with the full localization. I think really there are two camps that are going after the same thing of trying to get some measure of localization added to the ROC method.

We felt that for this particular case where you might have a nodule that was quite large in one lung and then a smaller nodule in a contralateral, that that smaller nodule in some cases might be the really important one that actually drove the care. We felt that getting at localization in some way was important. We went with the quadrant approach.

The quadrants were rated by the ROC readers but then the case, not the quadrant, is the unit of analysis for the computation of the p-values and the confidence intervals based on the jackknife and the bootstrap. You heard these references mentioned earlier but Obuchowski specifically is the reference for using this region of interest or

quadrant approach. Carolyn Rutter is the person that developed the method of using the bootstrap to sample cases.

The reading environment for our study is that readers were trained on work station use and we really tried to create a reading environment that was as similar to their individual practices as possible.

So the usual work station controls were available to them. If any individual reader had a particular window or leveling preferences, they were allowed to modify that. We didn't have it in the protocol that they had to read a particular way that would take them out of their reading environment. They were allowed to practice on three cases with the trainer present. The ambient lighting was adjusted to the radiologist preference. There was no hard time limit.

The instructions given to the readers was to only search for 4 to 30 mm. actionable solid nodules, to rate each case post-CAD immediately after the pre-CAD rating so they had to go through the entire case pre-CAD and provide the ratings before the

computer would even allow them to turn on CAD and then provide the post-CAD ratings.

They were instructed to consider age, gender, and clinical indication. These were taken from the radiology report. We did not provide them with the full radiology report as that obviously would have provided too much information for them to be able to make up their own decisions.

So the basic study work flow here -- let's see which of these works. Yeah, this one works. When you saw the work station earlier, there was no blue line. The blue line is separating the upper quadrant from the lower quadrant. We didn't feel like we needed a line to separate left and right. The yellow line is indicating where they are in the exam.

As they were reading the case, they had the opportunity to bring up a pop-up menu to rate the quadrants at which point they would get this little cartoon of sorts with these slider bars. They would move the slider bars either all the way over -- you can't see. There's a little 100 there -- to indicate complete confidence that there was at least one

actionable solid nodule present in the quadrant, or zero to indicate complete confidence that there were none.

In this particular case you can see that the reader has gone through and given a pretty low confidence or, I should say, a high confidence that there are no nodules present in any of the quadrants.

Having done that they then have the opportunity to click this button up here and turn on CAD. It's a little bit hard to see here but there is a potential nodule. I'm not a radiologist. I won't tell you whether it is a nodule but it is located there in the upper right quadrant. Then they would have the opportunity to rate the case again.

In this case they might have changed their rating. In the other quadrant since there was only a mark in the upper right-hand quadrant, it's fairly unlikely that they would have changed any of their other ratings but they were allowed to.

So after doing this with our 15 readers who each read the 90 cases, both pre-CAD and post-CAD, were able to draw the ROC curves for each of the

individual readers. This is just an example of a single reader and so the area under the dash line is the pre-CAD Az and the area under the blue line is the post-CAD Az and then the area in between the lines is the Azdelta.

These are the 15 pairs of readings. I didn't produce this plot specifically to answer some of the questions that came up earlier this morning but I think it might answer some of them a little bit. Now, this is not the same plot that you saw earlier.

This has the pre-CAD area under the curve on the bottom and the post-CAD area under the curve going on the Y axis. So pre-CAD the range was from about .82 up to .96. That's the range of the 15 readers area under the curve. Post-CAD the low end was .86 to .96 so you can see a narrowing of the range post CAD with respect to Az.

In particular, these three readers who had -- I'm trying to look for a different word than worst -- had the worst pre-CAD Az performance of around .82 to .84 were the ones that improved the most, or were among those who improved the most. You might wonder

what about readers that did pretty well. Well, these two readers did very well pre-CAD, at least, measured against Az. And post-CAD they also had some improvement. It was a more modest improvement. They didn't have as much to improve.

Now, finally, there's this reader up here. This reader had a nearly perfect pre-CAD performance. This does just go to .96, not all the way to 1 so they weren't absolutely perfect. What you worry about with a reader such as this is you don't want CAD to cause them to change their impressions so they get worse and they did not.

So moving onto the primary analysis this is the average reader ROC curve. Again, here is the pre-CAD line, the post-CAD line, and the area in between is the Azdelta. I'm just going to focus in on this part right here because it is an important point about whether or not the curves cross.

The curves do not cross and so you can see that they are always apart. Especially in this area here I think is the area where people are most likely to have their individual operating points, although,

as you saw, they might go all the way out here.

These are the same 15 dots just plotted against a different axis so this is sort of how far away they were from that line. You can see individual reader improvements ranging from about .06 to zero to no improvement. And then the idea behind the Dorfman-Berbgaum-Metz ANOVA-after-jackknife analysis is to create a confidence interval and computed p-value that would allow us to figure out what might happen with a new reader with a new case.

I mean, that's really the idea of this confidence interval is what kind of performance would we expect from a new reader with a new case. You can see that both the individual readers as well as the average delta and the confidence intervals are well on the side of CAD better as opposed to the side of CAD worse.

Now, we went ahead and did a number of robustness analyses and these were basically about repeating the primary analysis varying different assumptions to demonstrate that the primary results are not sensitive to study design. I think these are

very, very important because there is a considerable literature that you can tweak different things and end up with different results. If we had found that, we would have been in a difficult position because we wouldn't have known whether or not we really did have a robust result.

I'm going to talk about this with reference to the statistical methodology, specifically the ANOVA approach versus the bootstrap approach. There are lots and lots of different iterations on this but I'm just going to focus on these two. I'm going to talk about the reference truth. I'll focus on the consensus standard versus the majority standard but there are a number of other reference truths that we examined and I'll just focus on those two.

And then panel variability. I've talked about the confidence interval being a way of getting at what would happen with a future reader with a future case. What you really want to know is what would happen with a future reader and a future case evaluated against a new truth, right?

That means that you don't just have to

have the random reader and the random case components of the ANOVA model. You also have to have some way of evaluating your truth against the random panel if you are going to fully capture the variability.

So the ANOVA-after-Jackknife compared to the bootstrap, I'll run through this quickly because you heard this earlier. The ANOVA-after-Jackknife is based on leave one out samples. Again, the leave one out here is cases. A case is being left out of each sample as opposed to a quadrant.

The Az end of the curve has been computed for each reader case combination and then analysis of variance random effects model is fit. This is the standard analysis of variance random effects model with full interactions described by Dorfman-Berbaum-Metz.

The bootstrap, I think nonstatisticians a lot of times find the bootstrap a little bit more intuitive. The experiment is replicated in 1,000 random samples so from our sample of readers in cases, we generated random samples of readers in a random sample of cases and for each sample we matched our

random readers with the random cases and repeated the entire analysis.

It is very computationally intensive but it gives you a way of coming up with confidence intervals that allow a nonparametric -- fully nonparametric approach to evaluating what would happen with a future reader in a future case. I do want to point out that the ANOVA-after-jackknife is semi-parametric. The ANOVA piece is parametric but the jackknife piece is nonparametric.

So these are the confidence intervals for the ANOVA versus the bootstrap. You can see that the confidence interval for the ANOVA is a little bit tighter. For the bootstrap it's a little bit broader.

One of the things that the bootstrap is known for is being able to come up with confidence intervals that are not actually symmetric about the mean because often there is not really any reason to believe that the competence intervals would be symmetric about the mean. In this case you can see it actually goes out further on the CAD better side. Even though the competence interval is wider, it does

not in anyway diminish the results.

So returning again to the primary analysis, the primary analysis, as I showed you earlier, is based on a delta Az of .024 and a p-value of .003. I just showed you a different methodology using the bootstrap and came up with .0246, very close, and a p-value of less than .001.

Then we went on to a different reference truth. The different reference truth that I'm talking about here, and I apologize that it's not on the slide. We didn't want to make it too dense, but this different reference truth is majority so this means that a quadrant would be considered nodule present if there was at least one majority or consensus nodule and it would be considered nodule absent if it did not have any majority nodules in it.

A really important thing to point out here is that the majority quadrants, the ones that two our of three radiologists in the panel consider to be actionable. They are included in every single analysis so that means that when we're talking about the unanimous truth, they go in to the false positive

side of things, as somebody calls it.

On the other hand, if we talk about this reference truth, they go into the true positive side. We felt like we don't know if those are nodules or not and so the most conservative approach to take is to always put them in every analysis.

The delta Az here is a little bit lower but the p-value is actually more significant, to use a loaded term. This has to do, I think, with this sample-sized paradox that Bob Wagner was describing earlier. The final step was to do the random reference truth.

We did the random reference -- actually, before I go to that, I want to mention on the different reference truths in addition to majority and consensus, we also looked at a minority reference truth which is sort of the loosest possible standard we could come up.

We also did a tighter truth based on having a second panel of five people look at the cases and define the truth more tightly. In all four of those cases we came up with a similar statistically

significant result. So the random reference truth is based on picking two panelists at random to review each case.

Pretend that the three-member panels didn't exist. Redo the truth assuming that third person just wasn't there in their room. When you bring together the first-pass findings, their data doesn't come in. When you go to the second-pass it's only the two out of two consensus. This allowed us to come up with competence bounds that captured that piece of the variance. It ended up being fairly similar, although the delta Az is somewhat diminished from that of the primary analysis.

So all variations gave statistically significant results. I'm a statistician so that's what I know best and that's why I'm best prepared to talk to you about. I take the point of some of the panelists that -- by panelists here I'm referring to you all as opposed to any of our other panelists.

You want some sense of what does it all mean. What does this Azdelta of .02 mean? For myself, I find it useful to think about individual

operating points. This is the pulled curve where we pull all of the readers together. You can't really translate this to a new reader and a new case.

These are analyses that you don't do to find statistical significance or to get a particular competence interval or particular estimate. There are analyses you do to try to understand the data. There were analyses that we put in our protocol that we would be doing but they were secondary analyses just to try to get some sense of what's going on here.

So this is the operating point of 20. Recall that we have this 0 to 100 scale so 20 reflects sort of the most aggressive end of the spectrum. We could go all the way out to 0 but 0 is just all the way at that end. Twenty was an area where you could imagine a fairly aggressive reader would say, "Even for a 20 I might want to do some kind of follow-up."

Fifty was indeterminant on our scale so that is one operating point that is interesting to look at. Eighty would reflect sort of the least aggressive reader. This is by no means all readers. If I put this plot out with all 15 of the readers, you

get sort of that weird scatter plot similar to what you saw earlier, but just to get a rough sense of what kinds of improvements are maybe plausible

So this dotted vertical line here is the line that corresponds to having the same false positive fraction. This is saying that if you started out at 50, your sensitivity could increase by this much without sacrificing your false positive fraction at all. Not one iota. If you think of the false positive fraction as your measure of safety and you think of the true positive fraction as your measure of efficacy, that is saying you can go up and get efficacy without any safety tradeoff.

Now, it's probably more likely that people are going to go a little bit up and over so maybe they are going to call more things. That's what we see with our individual rating. You can go up and over and still have the same positive predicted value. Even though you are giving up a little bit on the false positive fraction, you still have the same positive predicted value.

This 50 here is still a little bit over

from that so it's not exactly the same positive predicted value but the basic point is that you can go up and over without having a sacrifice or without having a substantial sacrifice.

So these are the analyses that I mentioned. They were in our protocol as analyses that we were going to do, but I really am very sympathetic to what Bob Wagner said about these numbers. It's so hard to say what they mean. What are these numbers. I don't want anybody to run too far with these numbers but I do feel like it's necessary, especially for people who aren't statisticians, to want to understand what's going on with some of the raw data.

If we take 20 as the threshold for where somebody -- pretend that all readers treat 20 as their criteria for actionability, then we would have had 16 percent of the total nodules so there were 1, 125 positive quadrants that the 15 readers looked at. Sixteen percent of those would correspond to misses. With this very aggressive cutoff I think odds are those are, in fact, observational oversights.

Post-CAD that goes down to 11 percent so

the 16 percent versus 11 percent, that's a 30 percent reduction in misses at that threshold. Now, that is a very aggressive threshold. Probably most readers aren't at that threshold. Fifty might be closer to where most people are at. It goes from 20 percent down to 16 percent. That's a 22 percent reduction in misses.

Then finally if we imagine that 80 is sort of a higher-end threshold of what might be called a miss, there is still a 15 percent reduction in misses. Now, these numbers are presented without confidence intervals, without p-values. Take them with a grain of salt. But in terms of understanding potentially the clinical importance, I think that maybe this may satisfy some of the desire to see a different number than just the delta Az.

I also wanted to show you what happens if we look at the true positive fraction and we look at the false positive fraction in a way that is probably more similar to the way that a lot of academic studies are done where you look at the cases where you are most likely to see an effect on the true positive side

and you look at the unambiguous nodule absent quadrants on the other side.

Here I really am throwing out quadrants. As a statistician I hate to throw out data but I'm throwing them out just to get a clearer idea of what's going on here. So if we are looking at the true positive fraction just for the smaller nodules, and I'm just using -- they are not really small.

I think a lot of people would define small as less than 4 or less than 3, but the intermediate-size nodules as a proxy for difficult to find nodules or easily overlooked nodules. Then you can see that you get more of a rise in the curve without quite as much of a tradeoff early on in terms of the false positive fraction. This is analysis that was not included in our protocol. It's just something that I added to try to get a little bit more understanding of what is taking place here.

So the study conclusions. Again, the study conclusions go back to the primary analyses that we did and the robustness analysis. The study conclusions are that the imaging checker CT improves

reader performance for the detection of actionable nodules. That was our objective and that's what we feel that we demonstrated. And specifically the results are robust to the analytical methodology, to the choice of the reference truth.

Again, it wasn't just looking at consensus and majority. We looked at minority, majority, consensus, and sort of a super consensus. Then it is also robust to the additional variation associated with selection of panelists. I described identifying two random panelists. We also did it with a single random panelist, with three random panelists and came up with very similar results.

With that, I'll turn it over to Dr. Delgado. Thank you.

DR. DELGADO: Thank you and good morning. I am Dr. Pablo Delgado. I'm clinical associate professor of radiology at the University of Missouri, Kansas City. I also practice at St. Luke's Hospital. I'm here to describe the beta experience that we're involved with.

First of all, I'll tell you a little bit

about where I practice in the setting, where the beta site was performed. I am a private institution affiliated with the university. We have a hospital setting as well as an affiliated imaging center adjacent to us. We practice with residents available and we have an on-site residency training program of which I am the program director.

Our patient base is quite varied and I think rather common place for the region. It's a typical mid-west community base of private as well as community patients. Our CT equipment for our radiology department, we currently have two four-channel multi-detector CT scanners which happen to be GE QXI light speed scanners, although I don't think that's of importance to this device as long as it's DICOM data and meets the collimation thickness.

We currently perform anywhere between 20 and 30 CT studies a day of the chest and these different diagnostic indications including CT pulmonary angiography, high resolution CT of the chest, detection of other lung diseases, as well as multi-organ disease workups.

The beta study that we performed was between the times of June and August of 2003 for a total of eight weeks. We processed numerous studies. However, the goal of the study that we agreed upon and embarked upon was to assess the functionality of this image checker, CAD software, and how we would work with it to answer the R2 developmental group questions about radiologist preferred reading practices as well as work flow issues of how this would be incorporated into our practice. And to determine future applications of training needs in training radiologists in how to use this device. It should be noted that we were not asked to assess the clinical effectiveness of the CAD system.

The design of the system involved retrospective review of CT chest cases from our institution from previous months that have already been acquired and already been interpreted outside of the study and that met the collimation thickness which, I think, was already mentioned, 3 mm. or less and were contiguous slices of the chest.

The cases were read by faculty

radiologists as well as residents so we got feedback from both experienced radiologist as well as radiologist in present training.

For the training of utilizing the device, we had an R2 application specialize on site for an entire day who got to work with most if the radiologists. A few that were not available for that time were given the training subsequently by those who experienced the training from the application specialist. That training process involved the description of the CAD algorithm, what indeed it does and what it doesn't with the review manual.

We also reviewed several institutional cases. First R2 had some cases of their own. Then we through the DICOM hookup were able to push some of our cases to the R2 device and process them so they were our cases. We also performed shadowing of retrospective reading sessions where the radiologists were able to work with the CAD device and subsequently ask questions if they felt that they were necessary or encountered any questions.

Our observations from using the beta

product demonstrated that most radiologists, in fact all, demonstrated a rather rapid learning curve for using the CAD device. In a rather short period of time most people felt very comfortable in utilizing the product as is intended.

We encountered no specific technical errors or malfunctions. We had no difficulties. We did, indeed, use it in the way it was intended and we asked radiologists to first look at the case in a soft copy reading mode and then subsequently push the CAD button and activate it and then review it immediately thereafter. We found that all radiologists missed nodules that were detected by the CAD.

There certainly are false CAD positive marks as Dr. Castellino pointed out. However, most of these are easily dismissed by radiologists and that includes both faculty and residents.

Of course, I would agree with the comments made by other panel -- excuse me, other presenters from R2 that we feel that radiologists definitely should review all images initially without CAD and then a subsequent read with CAD. The reason for this

is that CAD is not really made to detect every single nodule and, No. 2, the algorithm is such that it does not detect every single lung abnormality and radiologists are still responsible for detecting any lung abnormality.

In conclusion, I think that this product is very timely in what radiologists are facing on a daily basis. The development of multi-detector CT has led to an explosion, if you will, or significant increase in the number of images that are very detailed and radiologists are asked to interpret.

Numerous published studies have already documented there are limitations in radiologists' ability to detect lung nodules. I believe the detection really is the limiting factor of eventually determining actionability whether it is related to further diagnostic or therapeutic or interventional workups. We found CAD to me an effective tool in assisting the radiologist in the detection of lung nodules with multi-detector CT.

I will now reintroduce Dr. O'Shaughnessy of R2 Technology.

DR. O'SHAUGHNESSY: Thank you very much. I just have a couple of summary slides kind of to bring it all together at the end. I just wanted to reiterate the main conclusion from our clinical study for multi-detector CT exams of the chest, that the image checker CT CAD software system significantly at a p-value of .003 improves radiologist ROC performance for detecting solid pulmonary nodules between 4 and 30 millimeters in size.

And as both Mr. Miller and Dr. Castellino talked about and Dr. Wagner this morning, we feel that is a good measure for -- a reasonable measure for evaluating both a safety and efficacy aspect of the product. Also from the safety aspect, the product is intended to be used as an adjunctive device and with appropriate training we don't think there are any issues there.

Just to summarize, I'll put up again the same slides of the proposed indications for use. We thank you very much for your attention.

DR. IBBOTT: Thank you, Dr. O'Shaughnessy.

We are going to have time this afternoon

for detailed discussion of this presentation but let's take a few minutes now to see if there are any questions for the previous speakers or clarification that's needed.

DR. STARK: I have a few questions. Other panelist, please jump in. Dr. O'Shaughnessy, thank you. By the way, it was a fabulous presentation.

DR. O'SHAUGHNESSY: Thank you.

DR. STARK: Very interesting subject and I think everyone is interested in seeing this technology succeed. Certainly I am so forgive me. Some of my questions are, I guess, by nature going to be -- are intended to be challenging.

Mr. Miller talked about, as the panel did, what the word significant -- he used the term significance is a very loaded term. Later on when we discuss the marketing materials and things like that, I'm worried about the pressures on radiologists to buy and use a technology and want to shift the significance to what really is clinically significant.

In your presentation you pointed out -- I believe several of your experts pointed out that the

real clinical problem is that we're missing about 24 percent of nodules or we are missing nodules at a significant rate. I think it was something like 24 percent or something, perhaps you can refresh me, were seen in retrospect.

One significant figure of merit here would be what fraction of those nodules that are missed, that 24 percent that are detectable in retrospect, are now detected with this technology given that the technology by itself has a sensitivity of about 50 percent for detecting majority and unanimous nodules and a 50 percent detection rate? I'm just asking. It's very, very low.

That would suggest to me that at best the technology is going to reduce that 24 percent missed rate to about a 12 percent missed rate at the cost of generating 100 percent false positives and then having a radiologist groom through and sort all this out by basically being said, "Do it again."

I'm wonder if we had a placebo in this FDA trial of, "Radiologist, just do it again, " or, "Here is the sugar pill. Just read it again," would we

achieve the same presumptive 50 percent improvement in finding half of the lesions we know the current standard of care is to miss?

DR. O'SHAUGHNESSY: Right. I would like to answer that sort of in two parts. The first part I would like Dr. Miller to go over what we measured in our study and then have Dr. Castellino talk about translating that to the clinical environment if that's okay.

MR. MILLER: I guess there were a number of questions there. Is there one you would like for me to start out with?

DR. STARK: I think you will do a great job.

MR. MILLER: Okay. So the analyses that I showed at the end with the percent reduction in misses are sort of approximated percent reduction in misses where an attempt to get at that very issue. I suppose that it is to some degree your job and, to some degree, our job to determine what is clinically significant.

Now, the numbers that I showed you were

sort of in the range of a percent reduction in misses of somewhere close to 20 percent. Actually more like 20 percent on the low end. That is similar to what the experience has been with CAD for mammography.

For CAD in mammography the percent reduction in misses has been in that range. I think if you are a person that's affected -- I guess I'm drifting off from statistics here. I should have handed it over to a clinician but, I mean, my hunch is that is a number that would be meaningful.

As far as the stand-alone sensitivity, I do want to sort of bring us back to the fact that we evaluated two modalities here. The two modalities that we evaluated were the readers stand-alone performance and the reader plus CAD. The whole MRMC framework is developed around those particular modalities.

CAD as a stand-alone modality is not something that anybody is recommending that people use. Therefore, those stand-alone numbers, I think, are less valuable but are more valuable if they pick up some of the more important things.

Also I think some of those things in the 4 to 10 millimeter range that readers react to and say, "Oh, I missed that. I'm glad CAD pointed out." It's more about what did CAD find than it is about exactly what the percentage is.

DR. STARK: Did you answer the core question of if the radiologist right now standard of care I would suggest, and clinicians can debate this, is that we miss a quarter of the lesions that are actually there in retrospect. If we can accept that as a statement, then as you design the experiment, what data are there to suggest we would cut that miss rate and by how much?

MR. MILLER: Will you permit me to go back to the slide? Sorry. I'll get there soon. Okay. This, again, is presented as an analyses that was specified in the protocol that we would do, but you don't have competence intervals there so these are numbers that you would want to put competence intervals on if you were going to put a lot of weight behind them.

Also, they make the presumption that

readers all read with the same threshold cutoff and we know that's not the case. At a threshold cutoff of 50, let's focus on 50 for just a second, there were 228 missed quadrants. In other words, out of the total number of quadrants that the radiologist looked at, 75 positive quadrants times 15 so there are 1,125 times that one of the readers looked at a positive quadrant.

They gave a rating less than fifty 20 percent of the time. That is actually kind of a nice number because that number is not radically different from I think what we see in the literature. It may be a little bit lower. I think there's a little bit of a relaxed environment in the readings that they may be a little bit more likely to identify things. But 20 percent of the quadrant something is missed.

Post-CAD it goes to 16 percent so that's a 22 percent reduction in the misses. That is, I think, the number that is closest to answering the question that you raised. Is that correct?

DR. STARK: I think so. Let me see if I understand it and then I'll ask you about the affect

on this analysis of the quadrant versus the lesion methodology.

MR. MILLER: Okay.

DR. STARK: I think that prejudice thinks in favor of the technology. I'm not sure. So you're saying if the standard of care currently is to miss a quarter of lesions, then of that 25 percent we'll miss one-fifth less so now we'll miss 20 percent of the lesions.

MR. MILLER: Yes. Their miss is defined loosely as you are not actioning a nodule that a consensus panel believes should be actioned. I don't think that they are actually missing it in every case. Sometimes they are giving it a low rating.

DR. STARK: Correct. But as far as --

MR. MILLER: Yeah.

DR. STARK: You can debate the inference but the literature talks about a missed rated of 25 percent which we are going to equate with actionable nodules. As we talk about the parent efficacy of this, and I appreciate your honesty, is that we are taking a standard of care of a 25 percent missed rate

that juries and patients think is horrible in retrospect and we are going to cut that to a 20 percent missed rate. We can judge the -- that's the efficacy.

MR. MILLER: I should also add this is just based on jumping from one 50 to the other 50 on the curve. We did another set of analyses based on what happens if you jump from 50 to the other point on the curve where you -- I'm sorry.

I should say jump from 20 from one point on the curve to the other point with the same PBD and jump from 20 to the same point without sacrificing the false positive fraction. That also was a protocol specified analysis and the numbers go down a little bit. I don't remember how much but it may be five or 10 percentage points.

DR. CONANT: May I interrupt or just jump in for a second because you are the slide that I'm curious about. You mentioned it's similar to mammography. It is but it's so different. I'm very interested in the by-case analysis of this compared to by quadrant. The reason being I think you have a

little bias in your case selection and I'm not sure if that is okay or not.

You have the majority of your cases, 62 percent of the nodule present cases, as people with extra-thoracic disease. I'm not sure I really care about the absolute number of quadrants you've missed because once you've got three nodules in both lung fields, who really cares? It's metastatic disease so I would want to see these numbers by case.

I also think the comparison to mammography is very different because I think that, again, chest analysis is much more multi-focal and reflective of systemic disease than mammography in terms of a bilateral fairly somewhat independent process. I would just like your comments on that if you could take this another step and then do it by case.

MR. MILLER: We did not do these analyses by case. I suppose the data are there to do it. I think the challenge with doing it by case is that the way -- I should let a physician get up here in just a second but the way that one would action a case where you had one lung where you had a very high likelihood

of it being something bad, using my simple statistical language, and you had the contralateral lung where you had something that was probably bad. That one that's probably bad may actually be the one that drives the care of the patient.

Figuring out how you sort of wrap this all up and do something like this at the patient level with something that was sort of beyond the scope of what I was able to imagine. I absolutely do not disagree that it's something that would be useful to try to investigate in some way. Having said that, I think I really need a physician to answer the question.

DR. CONANT: I'm not sure what the answer is, though. However, in your cases it's very different if a person -- if you're looking for a primary lung carcinoma versus metastatic disease so they are very different clinical questions.

MR. MILLER: Yes. Let me let Dr. Castellino answer that.

DR. CASTELLINO: I'm not going to answer any statistical questions. I can guarantee you that.

It is hard to answer that question. I would like to put it more in a clinical context of how we read cases every day.

I agree that if you have a patient with a soft-tissue sarcoma and you find three, four, five nodules, unless you are in a setting where you have surgeons who aggressively pursue that, as I was at Sloan-Kettering, at times it is important to find a six or seventh nodule. There is a spectrum of surgical behavior.

Let's assume that you find six or seven you don't have to find the last three. We had very few cases like that. The second thing is that we are not positioning this product as a lung cancer detection product, although it does work that way. Patients with lung cancer who had a nodule, it was not necessarily the primary lung cancer. They may have had lung cancer before treated post-op, post-radiation.

We accepted those cases and had a lung nodule in the lung for whatever reason so it wasn't really as a primary detection issue. I'm not sure I

answered that completely and I do recognize that certain mammography is quite different, as I think we have discussed before, than chest CT.

I would like to go back to a couple of comments you made. If I understood you correctly, I think you said, Dr. Stark, that the issue was that we had a 50 percent sensitivity for consensus nodules. As I recall from looking at that, I think, with consensus we were closer to 80 or 83 with the classic nodule definition. I'm looking at the -- you'll see that later with Petrick.

If you stratify those nodules with what would be more definition that radiologists would call classic nodule. It ranges from 83 to 59 I think is the number. Is that correct?

DR. STARK: We can study it but I'm trying to draw data from table 10. When I suggested 50 percent, it was based on this so maybe over lunch you can --

DR. CASTELLINO: We can go through it. I thought it was about 59. But I think it's a good point. We would love to have developed an algorithm,

to be very honest, that was 100 percent sensitive but this is the best we've come up so far. I think the issue to me as a clinical radiologist is how would this affect me or my colleagues in practice to find more nodules that we look at a year later and say, "My goodness. How did I miss that? Why did I miss that?"

The ROC study, to some extent, I think, approaches that. I think this table here to some extent also would address that. These are nodules potentially that could be missed or are missed that the radiologist would say, "I would have liked to have seen that nodule to make a decision as to whether or not it's actionable or not." I don't know if I'm addressing the myriad of questions that you had but I would like to try to -- if you can rephrase some of them I would like to try to answer them.

DR. STARK: If the chair and the panel think we have time.

DR. IBBOTT: Let's wait until after lunch and we'll have that detailed discussion this afternoon.

DR. CASTELLINO: Can you write them out so

I can think about them?

DR. STARK: I'm not sure of the protocol. I'll ask for advice.

DR. IBBOTT: I don't think there is any reason why you shouldn't present those questions and let them think about them over lunch.

DR. CASTELLINO: That would be very helpful because they are a lot and I think they are important questions. Thank you.

DR. IBBOTT: Again, I'll take this opportunity to ask Dr. Mehta if he has any questions that require clarification at this point.

DR. MEHTA: No, I don't.

DR. IBBOTT: All right. Thank you.

DR. SOLOMON: Do we have time for anymore questions?

DR. IBBOTT: Well, certainly. Especially if it's appropriate now to get clarification on something before we break.

DR. SOLOMON: I guess I have a couple of questions for Dr. Delgado. I guess they start off by asking you a little bit more about what your

experience was with the system and then, more specifically, did you find that you as a radiologist or any of your colleagues were using the CAD system or becoming more dependent on the CAD system and not quite giving it the same kind of read that you would give ordinarily? Also, what was the impact on the time that you spent on a case? Did it make it longer or shorter? Why don't you answer those.

DR. DELGADO: Okay. Thank you. I think those are good questions. First of all, we did not do any time analysis with and without CAD or separate, just soft-copy interpretation and then soft-copy interpretation without CAD and then subsequently with CAD.

I think it goes to say that if you are doing the second review that there might be a time factor that would be slightly increased and that may be something to be quantified. However, in my experience I think, first of all, the first question is people were instructed through the training phase that this device was to be utilized through a primary read in which you make decisions on whether you see or

detect a lesion and then there is a way for you to mark it. Then you activate the CAD and then you go through, as Dr. Castellino said, really not the whole entire study again but only those images that identified a lung nodule. It might be on average three per case or so where you might click on a button and that would take you immediately to that axial's image and show you a lesion of which then the radiologist would make a decision, "Did I miss this? Is this a significant mark that I would consider actionable?"

Or, if not, then easily discharge and be done with it. If it was a mark that is consider a false positive, that would be discarded easily. I think we did have a few of our radiologist which initially asked the question, "Well, is this benign or malignant?"

Yet, we made sure and I as the principle doctor in charge of this made sure to remind them that this was not the purpose of this device. It's really only to present you with a nodule that you may have missed and give you the ability to either add that to

your findings or completely discard it. Does that answer your question perhaps?

DR. KRUPINSKI: This will probably be more for Dave. On point of clarification, you've got a quadrant and suppose the CAD during the initial view the reader says there's nothing there. There really is a nodule and then the CAD comes up and points out the nodule and a false positive.

Now the reader increases their confidence and now do you consider that in the analysis and how can you be sure? Do you consider that a true positive and an increase in behavior when, in fact, the radiologist was looking at the false positive? Is there anyway without localization to establish that?

If you were then to take your cases and throw away any instances where the CAD marked a true and a false positive and the reader went from "false negative to true positive" what then happens to the ROC curves? Admittedly, although you've got statistical significance, those curves are pretty darn close and you've got these ambiguous cases now. How do you deal with that?

MR. MILLER: Well, the short answer is that we don't know precisely what happens in those instances. It was not captured. Bob Wagner talked about this best of both worlds scenario. We really tried in the way that we did the study not to take the readers out of their normal reading environment.

We felt that was very important and so capturing additional data was something that we thought could take them outside of their reading environment and create some kind of placebo effect essentially. We don't have that data on which one of the nodules or which one of the findings, I should say, which one of the CAD marks they are reacting to.

Now, having said that, we did after we completed the ANOVA-after-jackknife analysis you can pull out from that analysis which cases are the ones that were most favorable in terms of producing a CAD effect and which cases are least favorable in terms of producing a CAD worse effect.

I sat down with a dozen or so of those cases with Ron Castellino, our chief medical officer, and went through them and said, "Is it obvious what

they're reacting to here?" In the overwhelming majority of the cases it was obvious what they were reacting to.

The number of marks per case is small enough that it is fairly unlikely -- I should say fairly. The case where you have multiple close to positive findings in a quadrant is not very common. It's common to have two in a quadrant but most of the false marks are very easily dismissable.

I mean, our engineers hate it when I say this but there are some vessels. I mean, not a statistician I look at it and I say, "That's a vessel." So the radiologist, it's really easy for them to dismiss those.

I guess the short answer is we did not do the analysis that you are suggesting but I completely take your point that it's important to figure out what was really going on in the ratings. I think I have a pretty good feel for it that they were reacting to true positives.

DR. KRUPINSKI: So you rate them all as true positives?

MR. MILLER: Yeah. I mean, the only thing that -- I mean, just from a programming perspective, the only thing that is fed into the analysis is the truth for the quadrants and the ratings. Whether there were or were not CAD marks there is not actually in the analysis.

You could do an analysis that was more of a parametric model and a fixed effect model where you tried to capture whether it was the quadrants with CAD marks that were causing the increase, but I think it's reasonably obvious that they are in trying to model that it gets pretty messy building that on top of the models that we already did.

Just while I'm up here, I did really quickly want to comment on the issue about the sensitivity, the back and forth about that table. I think you were doing a weighted average of some numbers in a table and we'll come back to that later, I think.

The sensitivity number -- I mean, it's just incredibly variable depending on sort of which reference truth you use and so if you hear different

numbers going back and forth, it's not necessarily inconsistent. Two people may actually be both reading sort of off the same page but in a slightly different spot on the page. Thanks.

DR. IBBOTT: Thank you. At this point Dr. Stark has a couple of questions he's going to raise now to be discussed later this afternoon.

DR. STARK: Actually, it's a response to Dr. Castellino's question which I respect and it's fair. I have been working very, very hard for this because, as we'll discuss later, I have spent 15 years wondering why my ROC based prediction that MRI for detection of liver cancer in 1985 was significantly better than CT. That was wrong. I think I know why and I think this group here, the industry group and the panel, I think, were at the nub of it.

Dr. Castellino, rather than have us giving the formality and the importance of this scratching on pieces of paper, I've asked the chair to allow me to read. I've formed a question and I'm going to read it into the record and I'll give you my handwritten copy of what I'm going to read just so that we're clear on

this. Forgive me. You've seen me scrambling over three minutes here. If any of this is unclear, I'll rephrase it. Thank you for offering to do this.

Would you please calculate from the data and/or literature discussed or presented here today, and in your submission, the net decrease in false negative rate which we have here today estimated to be 24 percent for practicing radiologists working by themselves when those radiologists in the future, we're projecting, are to add this technology and these results, these data to their practice, specifically accounting for what Dr. Conant was just asking about, accounting for and not crediting as a detection or improvement with the addition of CAD those quadrants or patients as you compile the data where CAD marked a false positive lesion in a quadrant where the radiologist alone had a false negative.

Where that radiologist, in other words, failed to recognize a true lesion false negative for the radiologist that was not subsequently marked by the CAD.

I have this written down. I think that

translates into English and I would be happy to clarify. Feel free to grab me during lunch if there is some nuisance of that that would make a better question.

DR. IBBOTT: All right. Thank you. At this point then, we'll call this session to a close and break for lunch and we will reconvene at 1:15, just a little less than an hour. Thank you.

(Whereupon, at 12:21 p.m. off the record until 1:18 p.m.)

A-F-T-E-R-N-O-O-N S-E-S-S-I-O-N

1:18 p.m.

DR. IBBOTT: Could I get you to take your seats, please, and we'll continue. Thank you. I would like now to call the meeting back to order and I would like to remind public observers of the meeting that while this portion of the meeting is open to public observation, public attendees may not participate unless specifically requested to do so by the chair. At this point Mr. Doyle has a statement to make.

DR. DOYLE: Yes. The R2 has approached me and indicated that they have developed answers to the questions that Dr. Stark proposed at the end of the morning session. In an effort to keep the meeting moving with the schedule we have, I have asked them to present those answers at the beginning of the discussion section this afternoon. They have the answers ready and I would just ask for the flow of the meeting to present those at that time. Thank you.

DR. IBBOTT: Thank you. We will now continue with the FDA's presentation on this PMA which will be introduced by Dr. Phillips.

Dr. Phillips.

DR. PHILLIPS: Well, in case you forgot what we're doing over lunch, we are discussing the image checker CT CAD by R2 Technology. It is a system that analyzes and displays to assist radiologists in review of multi-slice CT exams to the chest and in the detection of solid pulmonary tumors.

It is composed of several items. It's a combination of software and a computer. The system is a work station which is the image checker CT Model LN-500. This was approved for marketing under a 510(k) K023003, the software which is the operating system for the product that we are looking at today.

Again, the indications for use, and I don't need to read those. Then this was reviewed within FDA by a rather extensive team. Michael Kuchinski was the team leader; William Sacks was the clinical reviewer; Teng Weng was the statistics reviewer; Robert Wagner and Nicholas Petrick were

reviewed for analysis methodology; Joseph Jorgens reviewed the software; Larry Stevens did bioresearch monitoring; Fleadia Farrah did the manufacturing. That's the quality systems regulation; and Ronald Kaczmarek reviewed it from epidemiological basis.

Two people will present to you today, Bill Sacks and Nicholas Petrick, discussing the PMA. The other reviews were all found to be satisfactory and we are moving on from there.

With that, Bill Sacks.

DR. SACKS: I apologize for the jaundiced look of that. It wasn't so bad in the rooms we were testing this in. Okay. I'm going to just give some background. Then Nick Petrick will present the data from the clinical study and then I'll come back and draw some conclusions.

The outline of my introductory comments, I'll say something about the character of the device for those of you who did in fact, forget over lunch something about the clinical utility, a point about the instructions for use, and some issues that are new to this particular PMA.

First on the character of the device. Just to remind you, this is for chest CT scans and for CTs that are done for any indication the algorithm is trained to detect solid lung nodules, not, for example, ground glass opacities. It is trained to detect nodules between 4 and 30 mm.

Also there was a Hounsfield unit cutoff which is just CT numbers, the amount of radiographic attenuation that needs to be above -100. In particular, this is a computer-aided detector. Just to say a word about the difference between computer-aided detection and computer-aided diagnosis, a point I made earlier.

The difference between detection and discrimination lies not in the instrument but in the clinical use to which it's being put. The detector system, which is what we're talking about today, this left-hand column, scans entire images whereas a discriminator only scans portions that are selected by the user. The detector marks the images where a discriminator will give a level of suspicion that is just a number. As I say, the same device will do both

but it is thresholded to give you marks when it's acting as a detector.

On clinical utility, as we've heard, many nodules are missed in clinical practice for two major reasons. One, other pathology distracts and hundreds of images are present in one CT of the chest. Indeed, you may start out as a board certified radiologist and after reading 500 images you are certified board.

A CAD is intended to reduce the missed nodules, this CAD. That is, it is intended to increase the users sensitivity to detecting lung nodules. We will come back to this point.

Instructions for use. The important points are that the reader should review the films unaided first. Then the CAD marks the candidate nodules. Then the reader looks again in the vicinity of those marks.

If the CAD fails to mark a nodule that was judged actionable on the initial unaided review, the instruction in the labeling reads that the reader should retain that initial judgment, not back off just because the CAD failed to mark it. We will come back

to this in my closing comments.

Issues that are new to this PMA are should the particular choice of target for the CAD algorithm, the definition of truth, the unit of analysis and endpoints. I'll say something about each of those.

First, on the CAD target, the target is not malignant nodules but actionable nodules as we've heard which, among other things, means that the definition of truth is not based on biopsy or tissue histology which would be an external standard, but rather based on the judgment of an expert panel that is an internal standard based on the very images that are being evaluated here.

The unit of analysis, as we've seen, at one level of the statistical unit is the person but it's further broken down into long quadrants and Nick Petrick will say more about that.

Finally, the end points. One could do an entire ROC curves as was done and one could, as Bob Wagner explained this morning, in addition, or instead of, do the sensitivity and specificity of a particular action recommendation which was not, in fact, done in

this particular study.

In summary, again, just to remind you, the clinical study consisted of three expert radiologists drawn from a group of 11 but three at a time on a panel to determine what was called by the company reference truth for each nodule. Then there were 15 completely different radiologists with a range of experience, not necessarily experts, that were called the readers and they all 15 read all 90 cases and the 90 subjects were divided into 360 long quadrants. Those 15 readers used a 100 point scale for a confidence and actionability rating for each case.

Now I'll introduce Nick Petrick who will give you the clinical data.

DR. PETRICK: Okay. So my name is Nick Petrick and I will go through -- let me see which one of these work. I'll go through the clinical results that were done by the sponsor and some of our perspective. The outline of my talk will be first to talk about the applicability of Az in the analysis. Here I'm using the term Az which is somewhat more of a technical term but this is the same as the area under

the curve or AUC. Other people may call it area under the curve or AUC but I'm going to use that as meaning the same thing here.

I will also talk about and somewhat review what the sponsor presented on the pool of cases used for the clinical study. I'll talk about the definition of actionable nodules by the panel of experts. Then I'll go into the particulars of the clinical study.

In particular, I'll talk about the primary analysis which was analysis using a fixed panel of experts and then what is somewhat of importance here, the secondary analysis which was the analysis using random panels of experts.

Then I'll finish up my presentation by talking about the measurement of CAD stand-alone performance. When I'm talking about stand-alone performances this is the algorithm performance with no reader involvement.

Okay. So for the applicability of the agency here, I show one of the sponsor's curves for the average reader ROC from predisposed CAD and this

had a change in the area under the curve of .024 and a p-value as shown there .003.

What's important to note about the applicability of the Az is that degree in curve here is the pre-CAD and the reddish curve is the post-CAD. And what we're looking for is that the two curves don't cross. That is an important measure if we are going to use Az as an overall performance measure for ROC analysis. What we find from this average curve is that generally the post-CAD curve is higher or on the same order as the pre-CAD curve.

So just to summarize this, the pre-imposed CAD curves did not cross in the average performance I showed before. I think, more importantly, there was no substantial pre or post-CAD crossing in either the average or individual ROC curves. This is important.

That makes the Az statistically appropriate performance measure for this type of analysis. If they had a significant crossing, we would have had to look at some sort of partial area or some other measure of performance in that situation. Because of this conclusion the sponsor had used an Az

as a figure of merit in all their analysis that follows.

Okay. Now to talk about the pool of readers. Again, just sort of a summary of what the sponsor had talked about before. There is a pool of cases. There was a subset of that which was made of nodule cases. These were documented cancer cases so the primary neoplasm or extra-thoracic neoplasm with presumptive spread to the lungs. That is the set of nodule cases. The cases were allowed to contain non-nodule pathologic processes, things like pneumonia or emphysema and so forth were allowed to be part of that subgroup.

They took another set of cases. These were considered the non-nodule cases and what they term or what can be termed as normal cases where there was no nodule deemed present by the site PI and that site PI primarily relied upon original radiology reports in coming to that determination.

These cases could include a history of cancer, radiation therapy, or even previous thoracotomy were allowed to be in this data set. This

is a pool of cases that now the sponsor will pull out cases to run their ROC reading studies from.

At this point we're not going to talk about -- we are going to talk about actionable nodules or the object of interest in this application. In particular, there is a panel of expert radiologists that identified the actionable nodules. This was done in a two-stage process, again, just as a review as before.

In the first reading the cases were independent and blinded by three expert radiologists. The information provided to the radiologists were the subject's age, gender, and indication for the exam, obviously along with the exam as well.

Each individual radiologist marked all findings deemed to be lung nodules. Then the radiologist provided ratings for each of those nodules so there is a detection test and then there's a rating of the actionability of that nodule. It could have fallen into an interventional category. That is an actionable finding where further workup was advised.

A surveillance which is, again, considered

an actionable finding which was monitored with follow-up studies and this would probably be more typically additional CTs. Also, they could have rated as probably benign calcified. Again, no action required here, or probably benign noncalcified, no action required.

After the first pass was done, findings that lack 100 percent consensus after that first pass were reviewed unblinded by all three radiologists and basically they are going to reevaluate locations where either two out of three of the panel or one out of three of the panel call the location a nodule. then the radiologist would rate or rerate these on the actionability of the nodule candidates.

Along with this thresholding was applied to match what the general performance of the area where the algorithms should be performing and so thresholds of greater than 4 mm. in diameter for each nodule candidate and a peak density of greater than -100 Hounsfield units. This considers a CT number and is related to the attenuation coefficient in grayscales in the CT exam.

Then after each nodule was identified, each lung quadrant was categorized based on the highest actionable finding within that quadrant. Then subsequently the quadrants will be used in the observer studies.

Now, just to summarize what was found in that initial pass, again, this is three experts per panel. I'll show in this column the unanimous actionable. That's three out of three finding. Majority actionable two out of three. Minority actionable one out of three. You can see that for unanimous actionable there was 142 findings. For majority there were 168. For minority there were 149 findings.

This gives you somewhat of an indication that panel variability is an important component here. There's a lot of cases, almost a third -- only about a third of the cases were unanimously actionable and another third or so were two out of three, and another third were one out of three. This gave the FDA an indication that panel variability was an important component and probably should be taken into account in

the clinical study.

Now to go into the clinical study, there were multi-reader, multi-case ROC observer studies. Again, the test statistic was the Az or area under the curve. I'll present net results based on analysis of 90 case data set, 360 quadrants. The sponsor also performed a 32-case study and also presented pooled results of the 32 and 90 cases. I'll just limit myself to the 90-case study.

What's important the MRMC allows us to look at the variability, confidence intervals, and significance testing and we can take those into account. That is important obviously in this case to determine significance and then to try to get an idea of what the separation is between the reading without CAD and reading with the CAD device.

In order to analyze the variability confidence intervals and significance two approaches were used, ANOVA-after-jackknife and bootstrap analysis. So here is just the general flow chart to the clinical study and this will be followed for all the clinical studies. The study starts out with a

pool of readers. These are going to be the group of radiologists that are going to actually read the cases and give rankings for each quadrant.

There's a pool of cases and there's a pool of experts and the experts will be used to define truth. There will be a sample pulled out of cases. It will be used by the pool of experts to define nodules. There will be a set of readers picked out. Those cases will then be read using multi-reader multi-case ROC observer study and an estimate of the Az will be calculated. This could then be redone for different case sets, different reader sets, and potentially different experts on a panel.

So the important components here are how to measure the variability confidence intervals and do significance testing. Again, two approaches were taken, ANOVA-after-jackknife analysis. This is a parametric type of analysis and just jackknife if a leave one case out type of analysis.

Again, we're talking about leaving out a whole case so you're leaving out all four quadrants together and then performing a quadrant-based analysis

on that. So just as a quick example, if we had a case set of case one, two, and three, when jackknifing is performed or leave one case out, the first partition is going to be one and two. We've left out case three. The second partition may be set case one and three, case two has been left out.

Finally partition would be two and three leaving case one out. Then using those partitions and looking at the pseudo values that come out of that you can use ANOVA to estimate the variability confidence intervals and significance. The analysis assumes modality as a fixed effect and readers, cases, and all interactions as random effects in the ANOVA.

A second approach to doing this is bootstrap analysis and this becomes important to look at variability of the truth panel. This is, again, just to repeat, is a nonparametric analysis. What happens is randomly generated data sets are created based on the original data using replacement. Just as another quick example, with a case set of one, two, and three again when you run bootstrap you use replacements of the first partition, randomly pick

maybe case three, case two, and case three.

When you do the analysis you assume that case three and case three are really separate events and we bootstrap across those to get those potential partitions. The second partition you may pick case three, case one and case two. Here all the cases have shown up equally. Then a third partition may be case one, case one, and case two and so forth.

So the primary analysis, again, the same basic diagram as before but now there's a resampling scheme introduced into the analysis. The resampling is used for the pool of readers, again, the people that are going to -- the radiologists that are going to rank the quadrants and the pool of cases.

The truth is based on a fixed three-member nodule definition panel, again, based on unanimous consensus. The analysis will be based on ANOVA-after-jackknife. Also bootstrap analysis was also performed. What happens here is a pool of readers go in. It's resampled so it picks out a subset of readers. Likewise a subset of cases is selected using a resampling scheme. The cases go into the definition

panel where the panel is fixed and define the actual nodules of interest or the quadrants that are positive or those that are negative.

The set of readers are then randomly selected and go in and perform the ROC experiment. That gives one estimate of Az. This process is repeated either through jackknife or bootstrapping in order to get estimates for the variability and allow for confidence intervals and significance testing.

So just the result of the clinical study. Again, this is for a fixed three-member nodule definition panel. In the first column I show the pre-CAD Az for both jackknife and bootstrap. The second column is post-CAD, the change in the Az, the p-value for that particular test, and the lower and upper confidence intervals.

You can see that the results are fairly consistent between both jackknife and bootstrap with a pre-CAD Az of .881 or .879, post-CAD increasing to .905 or .903. With change on the order of .024 we see fairly small p-values for both the jackknife and bootstrapping. Then the confidence intervals also

fairly consistent.

We wouldn't necessarily expect the bootstrap and the ANOVA to give us the same values but it's nice actually to see that there is consistency here between the two analyses.

So just some conclusions on the primary analysis. The sponsor has shown a statistically significant improvement in Az from pre to post-CAD and that is on the order of .024 or change in area under the curve.

The ANOVA-after-jackknife and bootstrap analysis showed consistent performance in both significance and confidence intervals. The analysis, however, was limited because it did not take into account any variation in the expert panel. Variability of the panel would add uncertainty to the performance estimates, or we anticipate that variability in the panel would add uncertainty to the performance estimates.

This is, I think, an important factor because we don't have this cold standard of truth. We are dealing with a panel truth. We expect if we

sampled a new panel, they may come up with a different set of cases. They certainly would come up with some different nodules there.

One of the important questions is how would performance change with a different panel makeup. That is one of the questions that we had talked to the sponsor about addressing. In particular, looking at a different number of panel members so if you have a different panel makeup or a different definition of truth potentially and different sets. What happens if another set of experts was used.

So a secondary analysis was conducted here. I'll show there are many different types of analysis done by the sponsor. I'll concentrate on one set of random panel makeup. This will be based on a random three, two, or one-member panel, nodule definition panels and assuming the definition for truth is unanimous consensus.

Because of this type of analysis the ANOVA-after-jackknife isn't applicable at this point so only bootstrap analysis is possible. It follows a

similar scheme as before. We, again, start with a pool of readers, pool of cases, pool of experts. Here, however, bootstrapping is applied to the pool of experts as well so that we have a different panel makeup for defining truth. That adds variability into that definition of truth and we can use our MRMC ROC observer study to take into account that variability.

So we use bootstrapping to select a group of readers, a group of cases, and a group of experts. Again, with that particular combination we get an estimate for Az. That study is repeated a number of times to allow again to look at variability where we have included variability of the truth.

So, again, these are random three, two, and one member nodule definition panels. When I'm talking three-member panels I'm saying unanimous consensus. Three out of three have to agree. When I get results for two members that means two members.

They both have to agree. Obviously for one-member panel it is the opinion of one of the members. The sponsors randomly sampled that panel so that we get the added variability from having many

different experts involved.

Again, the same layout here. The pre-CAD Az, the post-CAD, the change, the p-value, and the lower and upper confidence intervals. We can see from pre-CAD this measurement of performance was .845 increasing to .868.

For the three-member random panel a change of .022. For a two-member panel it was .832 increasing to .854, again a change of about .022. One-member panel .817 increasing to .838. Again, a change of about .0. This is 21 but very similar 0.22 on average.

We also see fairly consistent upper and lower confidence intervals for all different definitions of the truth. Then we see the significance values which are fairly small as well. That's sort of interesting because what I talked about before was that we expected when we incorporate randomness of the panel in here, we would see an increase or a decrease in the statistical significance that this would be a harder -- that it would be harder to chose statistical significance.

Really we see similar p-values to what we saw when we had a fixed-member panel. One of the possibilities or one of the trade-offs that may have occurred was something that Dr. Wagner talked about this morning where when the definition of truth is varied, we have also varied the case mix or the differentiation between negative and positive findings so we have now moved ourselves potentially more off the curve where we have a more closer balance study which gives us effectively a larger number of cases or a larger number of effective cases.

That was traded off against the variation in the truth. Those seem to potentially have traded each other off where we don't see a big difference in the performance. This is one possibility. It's certainly not conclusive in any way but it is somewhat surprising that we didn't see a larger variation in the truth when we randomize it.

So just some conclusions on the secondary analysis. This analysis take into account the random nature of the expert panel for defining actual nodules. In particular, it took into account

different number of panel members and different panel makeup using a bootstrap selection of the panel.

All variations of the panel make up confirmed a statistically significant improvement in the Az from pre to post-CAD and this change was on the order of .02. And just a more general conclusion, this type of analysis where we actually tried to randomize the panel makeup is likely to be a more appropriate type of analysis for assessment of devices when panel truth -- when only panel truth is available. That's obviously the case here but we can anticipate other devices potentially coming in where this will again be an issue.

Finally, I would like to talk about CAD stand-alone performance. In particular, this is a performance of the CAD algorithm alone and it's the algorithm's sensitivity and specificity with no reader involvement so we are just going to measure the performance of the algorithm on some set of cases or defined nodules.

Why may this be important? Well, it's generally important because the radiologist can use

this information to appropriately weigh their confidence in the CAD marking so this is a measure. If you are a reader or a radiologist trying to purchase this device, you generally like to know how it would work. Or if you have the device to use, to get a feel for how it's performing and what it might be marking.

Likewise, it potentially can be used as a benchmark for future revisions of the algorithm so as an FDA perspective knowing some benchmark of performance may help us to determine how to evaluate new revisions of this particular algorithm when it comes in.

The question becomes what's an appropriate performance measure for this particular device and this isn't necessarily an easy question to answer. Anecdotally the sponsor went back and looked for the unanimous three out of three fixed-member panel and look at those on the appearance of the nodules that the radiologist marked.

What they found was that many of those 142 findings did not meet the criteria of solid discrete

spherical density. They subsequently went back and reconvened a second panel to reevaluate the nodule but only based on appearance. Not to find new nodules but just look at the appearance of those nodules defined.

They put together a set of five independent radiologists and they were asked to categorize the nodules into two categories, either what they define as classic nodule. These are discrete, solid, spherical ovoid nodules, or as nonclassic nodules. These would be nodules that may not be discrete. they may be hyperdense, irregular in shape. They may be potentially normal structures that for whatever reason may not be considered nodules at all. This new panel is only going to look at the appearance of the nodules and determine whether they are classic or nonclassic in appearance.

This is a performance. In the first column I'll show the number of panels defining the nodule as classic. Again, there was a total of five. I'll just group together zero, one, and two out of five. I'll give the number of findings. The true positive fraction, the sensitivity of the CAD

algorithm to those particular subset of cases.

In general I'll just summarize the CAD false marker rate. Then I'll give a final column to the median diameter of the true positives detected. This is just to give an idea if there is any bias on the size of the nodule based on how many panelist defined it as classic.

So in the first category less than three out of five there was a total of about 65 findings. The sensitivity was on the order of about 32 percent. For three out of five there was a total of 13 findings, sensitivity of approximately 70 percent. Four out of five of the panelists saying this is classic in appearance the performance jumps up to about 82 percent. All five the performance is about 83 percent.

If you just combined all these findings together a total, again, of 142 based on the definition of truth. The sensitivity is on the order of about 59 percent. The CAD false marker rate, it varied between two and three depending on whether the sponsor incorporated or didn't the equivocal nodule.