manual, uni- and bidirectional steering deflection, and computer-aided remote magnetic deflection when used with the Sterotaxis Niobe system.

It should be noted that NaviStar ThermoCool variant was the only one used in the study being reviewed today for atrial fibrillation. This was also the case in those studies supporting the approved atrial flutter and ventricular tachycardia indications.

The sponsor is seeking to add the indication for treatment of atrial fibrillation to all variants in the ThermoCool family.

The Panel will be asked to comment on the applicability of the study data to all devices in the ThermoCool family.

A preclinical review was performed for this device consisting of engineering, biocompatibility, sterilization, and other information that was previously accepted for prior PMA applications.

The proposed indication for the treatment of atrial fibrillation did not raise any new preclinical challenges. There were no changes to the design of the catheter specific to the proposed atrial fibrillation indication.

Therefore, there are no outstanding

preclinical issues for these devices.

The sponsor performed a single pivotal trial to support the adding of treatment of atrial fibrillation to the NaviStar ThermoCool catheter. The treatment group was patients undergoing ablation with NaviStar ThermoCool catheter. The control group was given antiarrhythmic drug that had not been previously prescribed.

The trial was prospective, unblinded, randomized control trial with two to one randomization. It was performed at 19 centers, 4 of which were outside the United States. There were 167 subjects treated, 103 of whom were in the ablation arm. Thirty-six patients crossed over from the control arm to the ablation arm per protocol after failing the primary effectiveness endpoint.

The primary effectiveness endpoint was chronic success through a nine-month evaluation period and assessed for superiority of treatment over control. The chronic success was defined as freedom from documented symptomatic paroxysmal AF episodes and from changes in drug therapy after a blanking period within each group.

The primary safety endpoint was incidence of primary adverse events within seven days compared

to a perspectively established performance goal.

I'd like to review what constituted an effectiveness failure for the treatment and control groups.

In the treatment group, an effectiveness failure consisted of documented symptomatic atrial fibrillation, a change in antiarrhythmic drug regimen after the blanking period, repeat ablation greater than 80 days after the index procedure, or acute failure which included use of a non-study catheter, more than two repeat ablations or failure to demonstrate entrance block to a pulmonary vein targeted for isolation.

For the control group, an effectiveness failure consisted of documented symptomatic atrial fibrillation, change in antiarrhythmic drug regimen after the dose-loading period or a discontinuation of the study antiarrhythmic drug.

Otherwise, if none of these occurred by nine months, the subject was a chronic success.

The study proceeded as follows: the ablation group of patients underwent ablation therapy followed by a 90 day blanking period during which time recurrent AF didn't count towards chronic effectiveness, and the antiarrhythmic drug, were up

to two repeat ablations could be used as needed.

After 90 days, new antiarrhythmic drug therapy used during the blanking period had to be discontinued and any additional symptomatic afib recurrence was considered an effectiveness failure.

Medical control patients had a 14-day period during which time antiarrhythmic drug therapy could be optimized. After that period, new or increased dosages of antiarrhythmic drugs or the current symptomatic afib was considered an effectiveness failure. Patients that failed chronic effectiveness were eligible for crossover ablation therapy.

Next, Laura Thompson will present her review of the clinical study from the statistical viewpoint, followed by Randy Brockman who will provide the clinical review. The FDA will conclude with Ellen Pinnow discussing the potential postmarket and the epidemiological review. Dr. Thompson.

DR. THOMPSON: Thank you, Dr. Eloff. I'll be presenting the statistical review of this PMA.

Here's an outline of what I'd like to talk about today. I'm going to give a second overview of Bayesian statistics to complement that of Dr. Berry, and I'll explain how it was used in the sponsor's

trial. I'll briefly overview the study design again and then turn to discuss the primary endpoint analyses. I'll discuss poolability across ties and then a summary of the statistical and design issues.

During the presentation, I will bring up several discussion items which we would like the Panel to think about when they review the primary endpoint results.

The first item relates to this trial being unblinded. To what extent did a placebo effect occur? Also patient symptoms were self-reported. This could lead to a reporting bias especially because the control group was eligible for ablation once they experienced a chronic failure. Also the time for randomization to initial treatment, either ablation or dosing varied among subjects and was longer for those who received ablation. Is it possible that the physician's decision of when to begin the treatment depended on the health of the patient?

Finally, the largest-enrolling site performed substantially better than the other sites. The question is, is there an issue of generalizability or should there be a recommendation in the application of the procedure?

I will revisit each of these items in more detail during the presentation.

So now I'd like to give an overview of Bayesian statistics because the primary endpoint is analyzed using Bayesian methods.

The Bayesian method is an approach for learning from evidence as it accumulates. The Bayesian approach uses Bayes' Theorem or Bayes' Rule, as referred to in Dr. Berry's presentation, to combine prior information with current information on a quantity of interest, for example, an adverse event rate.

Prior information on the quantity of interest might come from information from previous comparable studies, subjective ideas prior to running the study, or one can use Bayesian statistics with no prior information by using what's referred to as a non-informative prior distribution on the quantity of interest. This might represent lack of information about the quantity. Non-informative priors were used by the sponsor in their application of Bayesian methods.

So as a simple illustration, suppose we are making a decision about the adverse event rate in a population that uses a particular medical device, and

we'll run a small study. Before we run our study, suppose there are existing data from a previous generation of the device that would imply a distribution on the adverse event rate that looks like this figure, where the prior mean on the adverse event rate is .35. The probability that the adverse event rate or the prior probability on the adverse event rate, the probability that it takes on any particular set of values is determined by the relative area under the curve for those values.

Suppose a hypothetical performance goal for the adverse event rate is .4, and we want the study we're going to run to ultimately show a small probability of the adverse event rate being greater than a hypothetical target of .4. According to this prior distribution, the prior probability that the adverse event rate is greater than .4 is the shaded area, and it's about .38. So about .38 or 38 percent of the total area under the curve is greater than .4.

Now, suppose we run a small study with 10 patients, and we find that 1 patient has an adverse event by the end of the follow-up period. So that's an observed adverse event rate of 10 percent. Combining the prior distribution that I just discussed with the distribution for the study data,

for example, a binomial distribution, gives an updated distribution on the adverse event rate through the use of Bayes' Theorem called the posterior distribution of the adverse event rate.

Now, with the new information from the study that gave an observed adverse event rate of 10 percent, the posterior mean for the population adverse event rate is lowered to .21, and the posterior probability that the adverse event rate exceeds the hypothetical threshold of .4 is now lowered to .04.

Also from the posterior distribution on the adverse event rate, we can get a credible interval. This is the analog to a confidence interval and describes uncertainty about the knowledge of the adverse event rate. Here with this posterior distribution, a 95 percent equal tailed credible interval on the adverse event rate runs from .6 to .42. And the interpretation is that there is a 95 percent chance that the adverse event rate fails in the interval of 6 percent to 42 percent.

Now, I'd like to describe the predictive distribution. The Bayesian predicted distribution is a special name we give the posterior distribution of an unknown outcome but one which can potentially be

observed in the future. For the hypothetical study we just ran, there was 1 failure among the first 10 patients. What is the likely result for the next 10 patients?

The predicted distribution for these next 10 patients can help answer this question, and here is the predictive distribution which describes the relative likelihood of there being anywhere from 0 to 10 failures in the next 10 patients.

We see from this predicted distribution that there's a fairly high probability that there will be 0 failures in the next 10 patients, and then a somewhat smaller probability of there being 1, but it looks like it's pretty likely that there will be a small number of failures in the next 10 patients.

Now, I'd like to explain how the predictive distribution is used to get a predictive probability of study success similar to what was calculated by the sponsor.

The predicted distribution can be used to collectively impute unknown subject outcomes in a trial. For example, we could impute the number of failures for the next 10 patients in our hypothetical study. We would draw the value from its predicted distribution. So we would draw it from, for example,

this distribution. For each drawing, we compute the posterior probability that the adverse event rate exceeds .4, our hypothetical threshold, and then we compare the computed posterior probability to a prespecified criterion of study success. For example, .025. If the calculated posterior probability is less than .025, the study is determined to be successful.

Now, to account for variability in the imputation, we can performance many imputations, perhaps 1 million imputations to get 1 million comparisons to the criterion, and here the criterion is .025. The proportion of the 1 million comparisons that beat the criterion or are less than the criterion is the estimated predictive probability of study success after the 10 patients are collected. So note that we obtained this result, the predictive probability of a successful study, after getting all 20 patients without actually collecting the next 10 patients.

An assumption that's made in using the predictive distribution is that subjects already in the trial with known outcomes are not distinguishable overall from subjects with unknown outcomes with respect to the primary endpoint. However, this

assumption is reasonable for many medical device trials because when we conduct the primarily endpoint analysis for a trial, we often don't distinguish subjects as to when they were enrolled into the study. So the assumption given here in this slide is not much different than the assumption used to justify combining data across enrollment times to conduct a primarily endpoint analysis.

One can apply predictive probability to adaptive design. Adaptive designs are trial designs that use accumulating data to decide how to modify aspects of the design during the course of the trial. In particular, one can use a predicted probability at an interim point as the rule for stopping enrollment into the trial. If the predictive probability that the trial will eventually be successful, once all enrolled patients complete follow-up is sufficiently high, then enrollment may be stopped and follow-up can continue only on patients already enrolled into the trial.

One can also use predictive probability at an interim point as a rule for stopping for effectiveness. If the predictive probability that the trial will eventually be successful, based on results at an interim point, is sufficiently high,

follow-up may be stopped and the trial declared successful before its planned completion.

One can also combine these two applications as the sponsor did for their trial. Note that to calculate the predictive probabilities, the sponsor needed to impute outcomes for subjects who didn't already have outcomes. We will see that they employed an imputation model that used available time to failure information. I'll describe that model in a bit more detail in a few slides.

So the sponsor used Bayesian predictive probability to decide whether to stop the trial early for effectiveness. A time-to-event model, where event was defined as chronic failure, was used to model the data and impute unknown outcomes. No external prior information on model parameters was used to obtain the posterior distribution. So even though predictive probability was used to stop the trial, posterior results based only on observed data are also in favor of treatment over control.

So even though using predictive probability is in fact valid for the sponsor's trial, we will see that in their trial, its use did not have an appreciable influence on primary endpoint results.

I'd like to conclude this section on

Bayesian statistics with a comment on how CDRH regards Bayesian trials in general.

CDRH supports the use of Bayesian methods for medical device trials. However, Bayesian methods do require planning, especially if external prior information is used. Sponsors are encouraged to discuss potential Bayesian methods with FDA prior to planning their trial.

In May of 2006, CDRH held a public meeting to discuss their "Draft Guidance for the Use of Bayesian Statistics in Medical Device Clinical Trials," and the finalized guidance is due to be issued soon.

Now, I'd like to discuss the primary effectiveness analysis.

So I'd like to remind you that, this is actually Ben Eloff's slide, but I'd like to remind you that the primary effectiveness endpoint was chronic success through a nine-month evaluation period, and it was assessed for superiority of treatment over control. Chronic success was defined as freedom from documented symptomatic paroxysmal AF episodes and from changes in drug therapy after a blanking period within each group.

The primary effectiveness endpoint

evaluation computed the posterior probability of superiority, and compared it to a prespecified threshold. So if we let P_T here indicate the treatment chronic success rate and P_C indicate the chronic success rate for control, then if the posterior probability that P_T exceeds P_C, exceeds .98, then effectiveness would be claimed.

Note that this is actually a posterior criterion. This is not the predictive probability criterion which I'll get to later.

Prior distributions on the parameters P_T and P_C are non-informative roughly speaking. They were uniform from 0 to 1, equally likely across the range. The maximum sample size was 230 subjects, and the sponsor used two types of interim monitoring.

First they used monitoring for sample size. When accrual would reach sample size of 150, 175 and 200, an interim analysis would be performed. Now, at this point, the time they stopped for accrual, not all enrolled subjects will have known outcomes. So we need to use the predictive distribution for those subjects. If the predictive probability of trial success for all enrolled subjects is at least .9 at the 150-look, or .8 at the 175 or 200-look, accrual would stop at that sample size.

When accrual stops, an analysis for an early claim of success is done when either 4.5 months have passed, where 4.5 is half of a 9 month evaluation period, or at least 50 of enrolled subjects have complete effectiveness outcomes. Then at that point, if the predictive probability of trial success is at least .99, effectiveness would be claimed.

Now, we call that trial success, means that the posterior probability of superiority exceeds .98. The trial success is described here. The .99 here is the criterion for predictive probability of trial success. It's the one where unknown outcomes are multiply imputed. So we're actually getting the predictive probability of a posterior probability exceeding another threshold. That threshold was .98, but the predictive probability threshold is .99. So actually they are two different numbers.

Okay. I'd like to describe the sponsor's model for imputing unknown chronic failure outcomes.

The sponsor actually imputed the time to chronic failure with a model for time to chronic failure that used the failure times from subjects with known outcomes. Failure would then be determined based on when the imputed failure time was

given. So if a subject's imputed failure time occurred after nine months, then they would be deemed an imputed success, and if it occurred before their evaluation period of nine months, it would be deemed a failure.

Failure time was assumed exponential with rate varying piecewise across time and separate across treatment groups denoted by the letter G. So G equals either treatment or control.

Here time is in months. So the failure rate was assumed constant at θ_1 from 0 to 2 weeks, then changing to θ_2 from 2 weeks to 2 months and then changing to θ_3 from 2 months to 9 months. All three rates were given identical prior distributions with overall prior means of one failure a month.

With the sponsor's model, there are several assumptions which should be explored, two of which are that the failure rate is piecewise across the given time periods, and that the prior distribution on the rates has a mean of 1. There's also the assumption of the exponential form of the model.

The sponsor checked out assumption number 2 and tried different prior distributions on the failure rates. However, they did not find any influence of the prior on the ultimate results.

FDA checked assumption number 1, and instead assumed a constant probability of failure from 0 to 9 months, different for each treatment group, with a uniform prior on each rate.

We also modeled binary success failure outcomes instead of time to failure information thus making the imputation model more consistent with the primary analysis model.

Nonetheless, FDA reached an identical conclusion using this model which is perhaps somewhat simpler as did the sponsor using the time to failure model.

The adaptive design was introduced into the trial midway. As mentioned before, the sponsor was having significant enrollment problems in their U.S. sites. In addition to extending enrollment to OUS sites, they proposed to replace fixed sample size design with an adaptive sample size design plus interim monitoring for effectiveness. 106 patients had been enrolled, with the sponsor blind to results at the time. And the sponsor also changed the criterion for success to Bayesian posterior probability instead of frequentist P value.

The FDA review team believed it was potentially problematic to introduce the Bayesian

adaptive design after the trial had already begun. However, the sponsor emphasized that not all enrolled subject had even completed their nine-month evaluation period, and furthermore that they were blinded to any chronic results.

After several meetings with the sponsor, FDA agreed to change from fixed to adaptive design but recommended that the sponsor treat the first 106 patients as an interim look with appropriate statistical penalty. The penalty resulted in an increased posterior criterion for effectiveness in order to maintain the one-sided, type 1 error rate at .025.

Normally, FDA also does not recommend changing from frequentist to Bayesian or vice versa midway through a trial. However, in this trial, the switch does not greatly impact the results because there is no external prior information used, only non-informative priors.

Even so, I later show a tipping point analysis that conducts the original frequentist comparison of proportions. This tipping point analysis will show that the unknown outcomes from future subjects, the ones that were not ultimate collected because the design was switched to an

adaptive design, would have to be quite different from the current results where the original frequentist analysis did not show conventional statistical significance. So it is likely that the original frequentist analysis would have shown statistical significance.

At the first interim point, there were 160 subject enrolled with 148 of them eligible for analysis. Due to timing, the first interim point occurred at 160 subjects instead of 150 subjects as planned. I present the sponsor's Kaplan-Meier curves for time to chronic failure by randomization group where the red line is the ablation group and the blue line is the control group. The Kaplan-Meier estimates of probability of chronic success are .62 and the ablation group .8. In the control group, there were at the time of the first interim point, there were 55 subjects in the ablation group who had yet to either reach nine months follow-up time or have a failure and there were 8 such subjects in the control group.

The first interim analysis was actually to calculate the predictive probability of concluding superiority when all enrolled subjects reached an event or nine months follow-up. This was calculated

by the sponsor as exceeding .999, which exceeded the .90 threshold for stopping for enrollment. So the sponsor could stop enrollment at 160 subjects.

At that time, also 50 percent of enrollees had had an effectiveness endpoint outcome. So the sponsor made an early claim of success because the predictive probability also exceeded .99, which was the threshold for stopping for effectiveness.

At the time of PMA submission, the sponsor updated the Kaplan-Meier curves. Seven patients were subsequently enrolled during the period before the trial was officially discontinued due to a time lag in completing the interim analysis. So there are 167 enrolled with 159 of those subjects eligible for analysis.

I present the sponsor's updated Kaplan-Meier curves. Now, only 14 ablation patients have yet to reach an outcome of either success or failure. We're still in the evaluation period.

The sponsor also computed the posterior probability of superiority using the updated dataset at the time of PMA submission. They found that the posterior probability of superiority at that time was greater than .999, which exceeds the posterior criterion of .98. They also presented the 95 percent

posterior credible interval for a difference between the treatment and control probability of success at nine months. This ranged from 31 percent to 58 percent with a median of .46, implying that there is 95 percent chance that the actual difference in the chronic success rates falls within the interval .31 to .58.

To see this again in graphical form, below is plotted a box plot version of the posterior distribution of P_T minus P_C, where P_T is the probability of chronic success for the ablation group at nine months and P_C is the probability of chronic success for the control group at nine months.

Note that the entire distribution falls to the right of the superiority line at a difference of 0, indicating a high posterior probability of superiority based on the results collected. And then I've also included the points of the 95 percent credible interval.

As a check on the Bayesian results, FDA also performed what's called a tipping point analysis. Here we determined how poor the results could be for the unknown ThermoCool patients in order for a classical comparison of proportions to still yield a significant P-value. So this was the

originally planned frequentist analysis for the primary endpoint before the introduction of the Bayesian adaptive design.

So I have a couple of scenarios here. Suppose that for all of the 14 censored ThermoCool patients, and these were the ones who haven't yet reached an outcome, suppose that they're all failures, if we do that and we calculate a classical comparison of proportions, the P-value is less than .001.

Now, suppose accrual went to the originally planned 230 total subjects. Because randomization was 2 to 1, that would imply 25 control subjects and about 38 ThermoCool subjects. Suppose conservatively that of the 25 control subjects, 13 of them are chronic successes. That's about a 50 percent chronic success rate, which is much greater than what was actually observed in the control group. Then of those 38 ThermoCool subjects, only 4 of them would need to be successes in order to obtain a P-value at the conventional criterion of .025.

This tipping point analysis is consistent with the sponsor's result that yielded a high predictive probability of trial success if all 230 patients had been followed.

So I guess just as an aside, this in some sense answers the question directed to the sponsor previously about, you know, what did we do with those 14 patients who don't actually have an answer. Well, granted this doesn't do a Bayesian analysis, but I can assure you that the Bayesian analysis is still reaching "significance." For all you frequentists out there, this should be, you know, comforting to know that even if all of those failed, we'd still get the P-value that's less than the conventional significance level.

Okay. Given all of that, now we get to some discussion items.

So although the posterior probability of superiority of ThermoCool over control and chronic success was very high, even nominally close to one, where ThermoCool subjects achieve chronic success more often than control, there are some limitations in the design of the study such that caution must be used to interpret those results.

First the trial was unblinded, and as we all know, unblinded trials can be plagued by a placebo effect because the subjects can be led, intentionally or not, to believe that they are receiving the better or worse treatment even before

the study results are known. It is not known to what extent the effectiveness results are due to a placebo effective.

Second, the determination of symptomatic AF was not measured entirely objectively. Subjects had to first report their symptoms in order for symptomatic AF to be considered as having occurred. Otherwise, symptomatic AF was not investigated, at least not routinely.

It is unclear to what extent the effectiveness results are due to bias in reporting symptoms. Because control subjects were eligible for the newer treatment, once they experienced a chronic failure, they might be more inclined to indicate symptoms in their reports.

Now, I'd like to discuss the time from randomization to initial treatment.

The sponsor has already noted that the evaluation periods for the effectiveness endpoint began at different calendar times for the ThermoCool subjects and the control subjects. Specifically, time 0 for the nine-month evaluation period began after a three-month blanking period for ablation subjects and after a two-week dosing period for control subjects. After time 0, which should be here

or here, the two groups were compared against each other with respect to chronic success.

However, the beginning times of the blanking and dosing periods after randomization, which would be at this point, varied from patient to patient intended to be delayed longer for treatment patients. As already noted, the longest delay was 331 days from randomization until ablation, median 28 days, mean 43 days. For control, the longest delay was 76 days from randomization to dosing with median at 10 days.

As noted, we would expect some difference in the timing due to scheduling the ablation period as opposed to just beginning dosing, but that doesn't imply that there weren't be consequences such as imbalance between groups at the beginnings of their respective evaluation periods. So from randomization, where they're supposed to be equal by the way the randomization is done, until this point or from here to here.

It is unknown if subjects in each group became different from randomization to the beginning of the evaluation period. Within each group, the subject's initial treatment was allowed to start whenever the physician deemed appropriate. Allowing

an arbitrary starting point for the initial treatment could be a source of potential bias in an unblinded study. It is possible that ablation was delayed until subjects were healthier, thus making the treatment subjects start time 0 at an overall healthier position.

So although the timing of evaluation periods for this trial might have been consistent with that of similar trials, the effectiveness results should be interpreted within these limitations.

As already discussed quite a length, the largest enrolling site performed substantially better than the other sites.

There was site variation in both effectiveness and safety results, and I'll talk about both. OUS sites overall performed better than U.S. sites, and this appeared to be primarily due to the better ablation results at the highest enrolling site. However, treatment effects across site groupings are all consistent with ablation performing better than control.

You've already seen these graphs. This is the Kaplan-Meier curve for time to chronic failure for the largest enrolling site. You saw that there

were no treatment patients who failed. So it's 100 percent estimate of the chronic success rate and the control estimate was 11 percent. And the other sites treatment is 47 percent versus 18 percent in the chronic success rate.

And I believe there was a question before about whether these control groups differed with respect to chronic success rate. If you just did a test comparing the proportions, the answer is no, it's not significantly different. But nonetheless we see this. We've already talked about this. These are disparate results.

Using a logistic regression model for the probability of chronic success, along with non-informative priors on the regression coefficients, FDA found that the posterior probability of positive interaction between an indicator for the largest site versus other sites and randomization group on the probability of chronic success is effectively 1. What all those words imply is that there's a likely difference in the magnitude of treatment effect at the largest site versus the other sites, with the larger magnitude being in the largest site.

However, as was noted by the sponsor, excluding the highest enrolling site, the primary

effectiveness endpoint is still met. That is the posterior probability of superiority exceeds .99. It might have one extra 9 in there.

There was also an observed difference in primary safety results across site groupings. The highest enrolling site had 2 out of 46 ablation subjects with what were termed as primary adverse events. That's a 4.3 percent rate versus 12.9 percent in the other sites.

The FDA clinical reviewer, Dr. Randy Brockman, will discuss possible clinical difference between the largest site and the other sites.

Given the different magnitudes of observed treatment effects, it is unclear whether the overall results generalize to a solely U.S. population.

So to summarize, the primary effectiveness endpoint was met according to a prespecified statistical criterion, after a statistical penalty was paid for changing the design from a frequentist fixed sample design to a Bayesian adaptive design.

However, it is unknown how much of the observed treatment difference is due to placebo effect or bias in reporting symptoms.

Also variability in time from randomization to the initial treatment time could be a source of

bias.

And finally, the treatment effect in OUS sites might be different than in U.S. sites.

I'd like to turn the podium over to the FDA clinical reviewer, Dr. Randy Brockman.

DR. BROCKMAN: Good morning. I'm Randy Brockman. I'm an electrophysiologist with the Agency, and I'm going to provide our clinical review.

As you know, atrial fibrillation is the most common tachyarrhythmia we see in clinical practice. It's been estimated that it affects over 2 million Americans. The prevalence of AF has been established to be between .4 and 1 percent in the general population, and it increases with age as it can be seen in this graph. It really is a major public health issue.

AF affects a broad spectrum of patients, people both with and without other heart disease. AF is associated with an increased long-term risk of stroke, heart failure, and mortality. Some patients may have severe symptoms while others may be relatively asymptomatic.

According to published guidelines and the HRS consensus document, the principal reason to ablate for afib is to treat symptoms.

While catheter ablation for atrial fibrillation is gaining wider acceptance, differences in technique remain. According to the HRS consensus document, strategies which target the pulmonary veins remain the cornerstone of AF ablation procedures. Additional approaches include left atrial linear lesions, ablation of complex fractionated electrograms or ablation of ganglionated plexi. Right atrial cavotricuspid isthmus ablation is only recommended if atrial flutter is identified. Again, that's according to the HRS consensus document.

This slide presents two examples of lesion sets used for the treatment of atrial fibrillation. These images represent the electroanatomic maps.

In this picture, we're looking at the back of the left atrial. This is a PA view. This is the left superior pulmonary vein, left inferior pulmonary vein, right superior pulmonary vein, the right inferior pulmonary vein. The red dots represent a classic lesion set for encircling and isolating the pulmonary veins.

The orientation of this image is similar. In addition to encircling the pulmonary veins, this lesion set includes linear lesions at the roof of the left atrial and down to the mitral isthmus line.

These images are just to present an idea of some of the various targets involved in AF ablation in general. In a few minutes, I'll discuss the lesion set involved in the study we're discussing today.

The sponsor conducted a pivotal clinical trial that studied the use of the NaviStar ThermoCool ablation catheter for the treatment of medically refractory paroxysmal atrial fibrillation. This was a prospective, multi-center, unblinded, controlled trial. It was randomized two to one, ablation therapy to medical therapy. Primary effectiveness was compared between the two arms, and primary safety was compared to a performance goal.

Enrolled patients had to have symptomatic paroxysmal atrial fibrillation with at least three episodes within six months prior to enrollment, but as has already been discussed, only one of those episodes had to be documented electrocardiographically. They had to have failed at least one antiarrhythmic drug. That could be class I, II, III or IV.

Enrolled patients could not have AF episodes lasting more than 30 days. They could not have had a prior AF ablation. They could not have

advanced heart failure symptoms, could not have substantial left atrial enlargement or substantial left ventricular systolic dysfunction.

167 patients were consented and randomized. Initially it was 106 to the ThermoCool group and 61 to control. There were seven excluded patients. Excluded patients were enrolled but either didn't have the study catheter inserted or didn't receive the newly prescribed antiarrhythmic drug. Exclusions occurred in accordance with the protocol.

One additional patient was discontinued from the control group after consent was withdrawn. This left a primary effectiveness cohort of 159 patients, 103 from the ThermoCool group and 56 from the control group, and it left a primary safety cohort of 139 patients, 103 from the ThermoCool group, and it represented 36 patients that crossed over to ablation therapy from the control group.

Enrolled patients averaged 56 years of age. About a third were women. Most were reported to have New York Heart Association class I symptoms. They had preserved left ventricular systolic function without left atrial enlargement. Overall, baseline demographics were generally well matched between the two arms.

The protocol allowed enrollment of patient that had failed or been intolerant of rate control therapy, class II and IV drugs, as well as membrane active drugs, class I and III. I'll just point out that only 16 percent of patients had failed only rate control therapy.

According to the protocol, pulmonary vein isolation was required using electroanatomic mapping. The protocol required the use of the NaviStar ThermoCool ablation catheter with the embedded location center compatible with electroanatomic mapping, and again, I'll remind you that the sponsor is seeking an AF indication for all ThermoCool ablation catheters including those without an embedded location sensor.

So pulmonary vein isolation was required, but the protocol allowed left atrial roof and/or mitral isthmus lines, targeting of non-pulmonary vein foci that initiate atrial fibrillation, linear lesions in the right atrium, if atrial flutter was induced during the procedure and isolation of superior vena cava potentials identified during the procedure that triggered atrial fibrillation.

Patients randomized to the control arm received a class I or class III antiarrhythmic drug

that had not been previously administered and that was approved for the treatment of atrial fibrillation. I list the drugs and the protocol recommended minimum daily doses in this table. Protocol approved antiarrhythmic drugs did not include amiodarone which is not approved for the treatment of atrial fibrillation.

Additionally, FDA's Guidance Document on Clinical Trial Design for Percutaneous Catheter Ablation of Atrial Fibrillation recommended excluding patients who had taken amiodarone within six months prior to enrollment.

The Panel will be asked to comment on the generalizability of the control arm therapy to the general practice in the U.S.

I'd like to move onto a discussion of the results and we'll start with safety.

The primary safety cohort is comprised of patients that underwent an ablation procedure with the study catheter. This group includes patients that were randomized to the ablation arm as well as patients that were randomized to the control arm but, in the course of the study, became eligible for and underwent ablation therapy with the study catheter. So the primarily safety cohort included 139 patients.

As I mentioned, it was 103 from the ThermoCool group and 36 from control.

The protocol included a performance goal less than or equal to 16 percent. That's the 95 percent upper confidence boundary, and that represents the proportion of patients that could experience a primary safety event.

The primary safety endpoint is the incidence of early onset, within seven days of the ablation procedure.

Primary adverse events, you may have seen this referred to as catheter-related adverse events or CRAEs, as that was the terminology used in the clinical protocol.

This list includes the following adverse events. I won't read all of them, but it's death, myocardial infarction, pulmonary vein stenosis, diaphragmatic paralysis, atrioesophageal fistula, neurologic events, and then certainly there were others.

In the first seven days following the ablation procedures, several serious adverse events occurred that were not included in this list. So in addition to the primary safety analysis, I'll also present an analysis of serious adverse events in the

first seven days that weren't listed in this table. I'll also present several other safety analyses.

For the primary safety endpoint, there were 16 primarily adverse events reported for 15 patients. The proportion of patients who experienced a primarily adverse event was 10.8 percent and the 95 percent upper confidence boundary was 16.1 percent. The safety endpoint specified in the protocol had an upper confidence boundary of 16.0 percent. Therefore, the result didn't meet the protocol's established performance goal for the primary safety endpoint.

This slide shows the primary adverse events that were reported. Seven patients experienced a hospitalization. These consisted of several episodes of AF recurrence, anemia, pulmonary edema, hematuria, and pneumonia. There were five vascular access complications, several AE fistulas, a pseudoaneurysm, hematoma, and one simply identified as lower extremity pain. As has been pointed out, there were no deaths, stroke, esophageal fistula or myocardial infarction within seven days of the procedure.

So I'll briefly discuss several secondary safety analyses. The serious adverse events within seven days, again these are the events that were not

included in the primary safety endpoint because they weren't included in the protocol specified list. There were serious adverse events reported in five patients. In one patient, intro-procedural evidence of either a left atrial thrombus or an atrial septal tear, it wasn't clear on the intra-procedural electrocardiogram which it was, it resulted in termination of the procedure but no other clinical sequelae were reported for that patient. There was one episode of hemoptysis 48 hours after the procedure. That was felt to be possibly procedure related, but again it was conservative management, and no other clinical sequelae were reported.

The other serious adverse events I list here, but they were likely unrelated to the device.

The serious adverse events that occurred within 90 days that were tabulated, this includes all serious adverse events within 90 days including those that were captured in the primary safety endpoint. The proportion of patients that experienced a serious adverse event within 90 days was 20 percent in the ThermoCool group and 38 percent in the control group. Many of the serious adverse events for the ThermoCool group were actually AF recurrences. I'll point out that there were five serious adverse events

identified as life-threatening arrhythmias in the control group.

Serious adverse events after 90 days, the percent of patients that experienced a serious adverse event after 90 days was similar between the groups. I list them here. I won't go through all of them. I will point out that there was one death, and it was in the ThermoCool group, and I'll go into more detail on that.

So one patient died during the course of the study. This patient was randomized to the ThermoCool group. It was a 71-year-old man with a complicated medical history, included symptomatic atrial fibrillation, coronary artery disease, prior myocardial infarction, and bypass surgery. He had an ischemia cardiomyopathy with left ventricular hypertrophy, hypertension, and diabetes. About nine months after the investigational procedure, went to bed, despite experiencing chest pain. The following morning, his wife was unable to wake him. The EMS was summoned, but attempts to resuscitate him were unsuccessful. This patient's death was considered unrelated to the investigational device and procedure. The FDA had no reason to disagree with the investigator or the sponsor's assessment.

Pulmonary vein stenosis was defined in the protocol as a greater than or equal to 70 percent reduction in the diameter of the pulmonary vein from baseline. No pulmonary vein stenosis as defined in the protocol was reported.

While no patients developed severe pulmonary vein stenosis, according to the protocol definition, a number of patients did have some degree of narrowing based on baseline and follow-up imaging of the pulmonary veins. At three months post-ablation, 82 of the 139 ablated patients had follow-up imaging, 5 of which showed no substantial PV narrowing and 77 showed less than 50 percent narrowing. At 12 months, 29 of the 139 ablated patients had follow-up imaging, 27 of the 29 patients showed mild PV narrowing, one patient each showed no PV narrowing and moderate narrowing. I will note that no symptoms were reported in association with the observed degree of PV narrowing.

Acute success was defined as the confirmation of entrance block in all targeted pulmonary veins. In addition to failure to achieve entrance block, other reasons for classifying a patient as an acute effectiveness failure included undergoing a repeat ablation more than 80 days after

the index ablation procedure, use of a non-ThermoCool ablation catheter, or undergoing more than two repeat ablations.

Of the 103 NaviStar ThermoCool patients that underwent an ablation procedure, 2 were classified as an acute effectiveness failure because they had an ablation procedure more than 80 days later, and I apologize, your printed slides show 3 here. The projected slide is the correct number. So this left 101 patients as acute effectiveness successes with a simple proportion of about 98 percent.

The primary effectiveness endpoint was chronic success, which was defined in the protocol as freedom from symptomatic AF based on electrocardiographic data and no changes in the antiarrhythmic drug regimen.

For purposes of determining chronic effectiveness of the ablation or antiarrhythmic drug treatment, beta blockers, calcium channel blockers, digitalis, angiotensin receptor blockers and angiotensin converting enzyme inhibitors were considered antiarrhythmic drugs according to the protocol.

Chronic effectiveness monitoring was based

largely on transtelephonic monitors or TTMs. TTM transmissions were to occur on a prespecified schedule and for all symptomatic episodes. Other electrocardiographic assessments included periodic Holter recordings and periodic 12-lead ECGs. A Core lab was used to classify data from the TTMs and the Holter recordings.

According to the analyses presented by the sponsor, the ThermoCool group demonstrated a posterior mean success rate of about 63 percent. The control group demonstrated a posterior mean success rate of about 17 percent. The primarily effectiveness endpoint comparing superiority of NaviStar ThermoCool over control was met with a posterior probability of greater than 0.999.

This figure shows the Kaplan-Meier curve for chronic effectiveness comparing ThermoCool, the solid red line to control, the dashed blue line. The curves clearly separate and remain separated. The vertical green line is placed at nine months, the point at which the analysis was performed.

Data was stratified by the largest enrolling site versus the remaining sites. The largest enrolling site, which was outside of the United States and was conducted by a highly

experienced investigator, was selected for further analysis due to the high enrollment and successful outcomes. The largest enrolling site enrolled about 30 percent of the patients in this study.

The largest enrolling site reported 100 percent chronic success. You can see the red line here, while the remaining sites had an average chronic success rate of just under 50 percent. So as you can see, and has previously discussed by Dr. Thompson, there appears to be a substantial difference in chronic effectiveness when the largest enrolling site is compared to the remaining sites.

The sponsor offered several reasons for the observed site difference in outcomes. Rigorous conformance to the protocol requirements by the largest enrolling site resulted in no protocol adjudicated chronic failures, meaning failures due to something other than AF recurrence. So there were no protocol adjudicated chronic failures in the ThermoCool group in the largest enrolling site versus 12 failures, that's out of 72 patients or 17 percent, in the ThermoCool group from the remainder of the sites.

In terms of protocol approved medical management, ThermoCool patients at the largest

enrolling site were typically prescribed previously failed class I or class III antiarrhythmic drugs post-ablation which was allowed according to the protocol. A typical practice at the remaining sites was reported to be beta blockers and calcium channel blockers prescribed post-ablation which resulted in most patients protocol adjudicated antiarrhythmic failures. And also the lead investigator at the largest enrolling site had substantial experience in using catheter ablation for the treatment of atrial fibrillation prior to this study.

FDA identified another possibility. The largest enrolling site performed prophylactic right atrial ablations on most of their ablation patients. At the largest enrolling site, that was 23 out of the 31 ablated patients or 74 percent, whereas the procedure meaning prophylactic right atrial cavotricuspid isthmus ablation was performed on a much lower proportion of patients in the remaining sites, one out of 72 or just over 1 percent. It isn't clear to what extent this particular procedure deviation influenced the outcomes of the trial.

The FDA had some concerns about the adequacy of antiarrhythmic drug therapy in the control group. Four control patients received less

than the protocol recommended minimum antiarrhythmic drug dose. Eleven control patients received a previously ineffective antiarrhythmic drug. One patient was common to both. So the total was 14.

A sensitivity analysis was performed to assess the impact of these protocol deviations on chronic effectiveness. That analysis indicated that the insufficient antiarrhythmic drug therapy provided to the 14 control group patients did not materially impact the chronic effectiveness result of the study.

This table presents the chronic effectiveness data according to the antiarrhythmic failed for purposes of enrollment, in terms of whether the failed drug was a membrane active drug or rate control therapy. Note the relatively low numbers of patients that failed only a class II/IV antiarrhythmic drug. This is a total of 20 patients for whom a chronic effectiveness endpoint was known at the conclusion of the study.

The Panel will be asked to comment on whether you believe the study provides sufficient experience to support failure of only rate control therapy in the indication statement.

According to the study protocol, patients were required to transmit a minimum of 15 TTM

recordings during the nine-month chronic effectiveness evaluation period. They were required to transmit once a week for the first eight weeks, and following eight weeks, to transmit one a month.

The compliance index for each patient was calculated as a percentage based on the number of TTM transmissions within an expected timeframe divided by the total number of expected TTM transmissions per patient. As you can see, overall compliance with TTM transmissions was 88 percent, and it was similar between the two groups.

I'll just point out that TTM compliance was similar in the U.S. and outside the U.S. as well as the largest enrolling site, and it was relatively stable over time.

I'll just briefly mention protocol deviations. Fourteen control group patients received antiarrhythmic drug therapy that didn't adhere to the protocol. I already discussed this issue. Four patients received amiodarone during follow-up; three were in the ThermoCool group and one in the control group. Three of the four that received amiodarone were ultimately classified as chronic treatment failures. The one patient who was declared a chronic treatment success received amiodarone for only two

days. That patient was in the ThermoCool group. We didn't feel these deviations substantially impacted the study results.

I'll just remind you that according to the protocol, right atrial cavotricuspid isthmus ablation was only to occur if atrial flutter was identified during the procedure. However, prophylactic right atrial linear lesions were performed in 24 ThermoCool patients, 23 of which occurred at the largest enrolling site that also had the highest reported success rate.

The Panel will be asked to comment as to the importance of this modification to the ablation strategy.

I'll just summarize by saying NaviStar ThermoCool was shown superior to medical therapy in terms of reducing recurrent symptomatic atrial fibrillation at nine months. The largest enrolling site did have greater effectiveness than the other sites. While the primary safety endpoint was not met, review of individual safety events did not raise substantial concerns for FDA.

Now, Ellen Pinnow will discuss the proposed post-approval study.

MS. PINNOW: Okay. Thank you,

Dr. Brockman.

As an epidemiologist on the PMA review team, I'm responsible for working with the sponsor on the development of the post-approval study protocol. The sponsor has submitted a post-approval study outline. In the event that the device is approved, we will continue to work with the sponsor to develop a protocol that both the Agency and the sponsor can agree upon.

Here is an outline of my presentation today. First, I will discuss general principles that we utilize when thinking about the need for in designing post-approval studies. Then I will comment on the rationale for postmarket questions that the premarket study was not designed to answer that maybe answered in a postmarket study. Then I will summarize the latest version of the sponsor's post-approval study outline for the NaviStar ThermoCool catheter and the assessment of this post-approval study outline.

Before we talk about the post-approval study, we need to go over a few things. The discussion of a post-approval study prior to formal recommendation on the approvability of this PMA should not be interpreted to mean the FDA is

suggesting the Panel find the device approvable.

The plan to conduct a post-approval study does not decrease the threshold of evidence required to find the device approvable.

The premarket data submitted to the Agency and discussed today must stand on its own in demonstrating a reasonable assurance of safety and effectiveness in order for the device to be found approvable.

There are two general principles for post-approval study. The main objective of conducting a post-approval study is to evaluate device performance and potential device-related problems in a broader population over an extended period of time after premarket establishment of reasonable evidence of device safety and effectiveness.

Post-approval studies should not be used to evaluate unresolved issues from the premarket phase that are important to establish device safety and effectiveness.

The reasons for conducting post-approval studies include to gather postmarket information on longer-term performance of the device, data on how the device performs in a real world, broader patient population that is treated by community-based

physicians, as opposed to highly selected patients treated by investigator in clinical trials, evaluation of the effectiveness of training programs for the use of devices, evaluation of device performance in sub-groups of patients, since clinical trials tend to have a limited number of patients or no patients at all in certain -- sub-groups of the general patient population. In addition, post-approval studies are needed to monitor adverse events, especially rare adverse events that were not observed in smaller premarket trials.

And finally, we conduct post-approval studies to address issues and concerns that Panel members may raise based on their experiences and observations.

Post-approval studies should contain a fundamental study question or hypothesis, safety endpoints and methods of assessment, acute and chronic effectiveness endpoints, and methods of assessments. The post-approval study should specify the duration of follow-up.

There are three questions the FDA review team considered important in the long-term safety and effectiveness of the device that may be addressed in a post-approval study.

The first question is what will the real world performance of the device be in a more general population of patients and providers?

The second question is what is the long-term durability of effectiveness and the safety profile in patients treated with the device postmarket?

It is important to evaluate the long-term durability of effectiveness because a procedure acts to damage the heart tissue creating a permanent scar, we view this as having the same effect as a permanent implant even though there is no device remaining in the body.

The third question is, is there a difference in the effectiveness outcome in subjects in whom cavotricuspid isthmus ablation lines are placed in addition to pulmonary vein isolation?

This table presents an overview of the sponsor's latest post-approval study outline. The objective in the proposed post-approval study is to provide additional corroborative long-term safety and effectiveness data for the ThermoCool catheter in the treatment of symptomatic paroxysmal atrial fibrillation.

The sponsor proposed a perspective multi-

center cohort study with a non-inferiority design and historic controls.

The study population consists of 145 ablation post-approval study patients and 139 control subjects. The ablation post-approval study group are subjects who will be treated with the ThermoCool catheter in the post-approval study while the controls are subjects who were treated with the NaviStar ThermoCool catheter in the pivotal trial.

The sponsor proposes to follow the subjects for five years after the procedure.

The proposed safety endpoint is the occurrence of adverse events at seven days.

The hypothesis for the study is that the proportion of post-approval study patients with an adverse event at seven days is no worse than the ablation treated patients in the pivotal trial with a non-inferiority delta of nine percent.

The secondary safety analysis proposed is a descriptive analysis of the occurrence of adverse events at five years.

The proposed effectiveness analysis includes a descriptive analysis of the occurrence of symptomatic afib at five years and an evaluation of the effectiveness outcome in subjects in whom a CTI

lines are placed in addition to the pulmonary vein isolation.

For the short-term safety objective, the sponsor proposed to assess the adverse events at seven days. The sample size of 145 patients in the post-approval study is sufficient for the non-inferiority hypothesis. The anticipated dropout rate of 10 percent is a conservative estimate and is acceptable. However, the non-inferiority delta of 9 percent proposed by the sponsor may be too large.

During the afternoon deliberations, we will be asking the Panel to discuss what is the appropriate trial design to assess the short-term safety of the device?

For the long-term safety objective, the sponsor proposed a descriptive analysis of long-term safety of the device. The assessment would include a description of the occurrence of adverse events such as death, stroke, MI, et cetera, up to five years following the ablation.

In the current post-approval study outline, there was no stated hypothesis for long-term safety. The sample size of 145 patients will not be sufficient to characterize less common events such as stroke. The anticipated dropout rate of 10 percent

at five years is very optimistic, and no control group was posed to evaluate long-term safety.

The questions that have not been addressed in the outline include what is the appropriate long-term safety endpoint? What is an appropriate length of follow-up? And what is an appropriate control group?

For the long-term effectiveness outcome, the sponsor proposed a descriptive analysis. The assessment would include a description of recurrent symptoms of afib at five years.

In the current post-approval study outline, there is no stated hypothesis for long-term effectiveness, and no control group was proposed to evaluate long-term effectiveness of the device.

The questions that have not been addressed in the outline include what is the appropriate follow-up, and what is the appropriate control group needed to evaluate the long-term effectiveness of the device?

To evaluate the impact prophylactic right atrial ablation had on chronic effectiveness, the sponsor proposed a descriptive analysis of CTI patients. There is no stated hypothesis for the evaluation, and it is not clear how many ablations

would include CTI. There is no comparator population described in the outline.

In addition, the cohort design may not be the ideal study design for this evaluation.

The questions that have not been addressed in the outline include is there a need to address the differences in the effectiveness in the postmarket period? And is it appropriate to randomize patients to prophylactic right atrial ablation?

This concludes my presentation as well as the FDA's presentation for this morning. We welcome any questions you may have.

DR. BORER: Thank you very much. That was a very clear and helpful presentation.

I'm going to ask the Panel for any questions of the FDA. Earlier I suggested we could hold the questions about statistical approach until this point, but I want to change that because Dr. Naftel, who is our committee statistician, will be making a formal statement a little bit later, after which the sponsor and the FDA also will receive questions. So I'd like us to hold the questions about statistics until we've heard all three presentations, and then if there's anything that's unclear to us frequentists in the audience, we can

ask the questions then.

So are there any other questions, non-statistical methodology questions to the FDA? Dr. Kelley.

DR. KELLEY: I have a question for Dr. Brockman. As I understand it, there were 14 of 52 control patients with protocol violations as far as the antiarrhythmic drugs. So I wondered if there are any data as to whether they affected the results? Did you analyze the other 42 separately or did those 14 have a higher incidence of failure or do we know?

DR. BROCKMAN: There was a high incidence of failure obviously across the board on the medical arm.

DR. KELLEY: Did it change though or did we look at the 42 didn't have protocol violations?

DR. BROCKMAN: Well, we looked at it more from the perspective of the 14 and looked to see if we classified all of those as failures, we classified all of them as successes, and each stage in between. Any point we chose along that line of successes in the control group or failures didn't impact the final analysis.

DR. KELLEY: Okay. Thank you.

DR. BORER: Dr. Somberg.

DR. SOMBERG: You mentioned that you picked up, that there was a CTI difference between the OUS 1 and everybody else. I guess you picked that up from a review of the case report forms. Did you communicate with the sponsor? My concern is sometimes people don't always list things in case report forms or case report forms aren't properly -- is this a real phenomena the sponsor says or is this some sort of artifact of reporting?

DR. BROCKMAN: It was picked up from the line listings which are extracted from the case report forms. I guess you would have to ask the sponsor if they disagree with my analysis. As far as I know, this is a real difference.

DR. SOMBERG: Would it be okay for the Chairman to address the sponsor?

DR. BORER: Well, the sponsor will -- perhaps we'll wait and talk to the sponsor after lunch, but I think the key point is, and I had exactly the same question, are we absolutely certain these were prophylactic because the protocol didn't allow for that, and until I saw the FDA presentation in our book, I didn't realize they were prophylactic. So I think we need some statement from the sponsor, not now but after lunch, after you've had a chance to

think it through. Dr. Bilazarian.

DR. BILAZARIAN: Dr. Brockman, I have a question on slide 37, the serious adverse events at 90 days. Since both groups include multiple AF recurrences as an adverse event, of course, they could be adverse events, but they also could be an efficacy failure. If those are excluded, can you give me insight about what the rate of serious adverse events would be excluding multiple AF recurrences?

DR. BROCKMAN: I'd have to actually crunch the numbers. I would suspect it would be somewhat similar just based on the numbers of the AF recurrences and the randomization scheme. I think it would be similar.

DR. SOMBERG: So similar amount of --

DR. BROCKMAN: I think a similar amount of adverse events would remain if we pulled the AF recurrences out.

DR. SOMBERG: The other question I have is on slide 81 regarding the exclusion of therapies which are also obviously anti-hypertensive therapies, slide 81. Is there any data in the case report forms about blood pressure control regarding these patients who are not allowed to have use of beta blockers, and

obviously non-dihydropyrimidine calcium channel blockers were excluded, I assume also may have limited the ability to control blood pressure. Do we have any reports on that?

DR. BROCKMAN: I don't have data on outcomes according to blood pressure.

DR. SOMBERG: And the last question I have is on slide 85. Do you have any sort of insights about, much has been made about OUS 1 and the differences in outcome in regards to treatment arm, but in terms of the control arm, they have half of the success rate with the control as well, and I'm not sure why that could be explained by a difference in patients so-called healthier patients or patients with smaller atria. Do you have any thoughts about that?

DR. BROCKMAN: I don't have an explanation as to why that occurred. The numbers are relatively small especially when you're looking at a single site. Even though it was largest enrolling site, their single control numbers are relatively small. So the confidence intervals around those are relatively large

DR. BORER: Dr. Jeevanandam.

DR. JEEVANANDAM: I want to discuss slide

78 with the pulmonary vein narrowing. I notice here at 3 months and 12 months, you know, greater than 90 percent of patients have less than 50 percent narrowing. Is that 45 percent? Is that 1 percent?

DR. BROCKMAN: Actually I believe it was Dr. Waldo presented a very nice slide. I liked the way he did it. This is the way it was presented to me. So this is the reason I presented it as such. Just looking at the line listings again, many of those varied 10 to 20 percent either way actually. I don't show where there was reported enlargement. Many of them were 10 to 20 percent.

DR. JEEVANANDAM: So that is something that could potentially be progressive, and notice that at 12 months, that much fewer people actually were studied. So it would be something to look at long term because that could be a potential complication.

I guess my other question is, you know, when you look at the study design, you look at the primary safety analysis, it was done with patients who got therapy with the ThermoCool, and then even in the control arm, it was compared to people who got therapy with ThermoCool. Because there were 56 patients who got controls, right, and then out of that 36 were crossed over and got ablations. So the

complications were patients who got crossed over and ablated or patients who just got ablated right away.

DR. BROCKMAN: So the primary safety endpoint was for anyone who got treated with the study catheter, and that was in the first seven day. And that's, from a practical perspective, that probably applies to the serious adverse event that I presented in the first seven days as well. In subsequent analyses, I believe that patients who crossed over were censored in terms of the safety analyses at that point. So once they crossed over to ablation, they no longer contributed to the control group. Does that answer your question?

DR. JEEVANANDAM: Thank you. Yes.

DR. BORER: Okay. I have a couple of questions for you, for anybody on the FDA team who wants to answer. First of all, I thought you made a very important point with your first slide. You said the principal reason to ablate is to treat for symptoms. And that's what was done here in a sense; people had symptomatic AF and they underwent ablation.

The question is whether this criterion as it stands would be sufficient for the FDA for approval of a device? The way I understand current

thinking, you would really need to have symptoms that bother you a lot because it's not clear that anything besides being bothered by these symptoms is a sequelae of symptomatic AF other than the things that couldn't be measured because the numbers and the time and whatever strokes and what have you. So I'm wondering, and we may get back to this later, if this is an approvable device and if the FDA chooses to approve it, a label is going to have to be written. The label as it is in our book says this is for symptomatic AF, and I'm asking you whether that is an acceptable standard to the FDA?

DR. BROCKMAN: I think there are a couple of reasons, and this is my own opinion. I think there are a couple of reasons to treat atrial fibrillation. There are many reasons obviously. We would all love to see a reduction in clinical outcomes in terms of stroke and mortality. That's a pretty big study with a long-term follow-up. So what else is reasonable? Well, I think a lot of experts, people smarter than I am, said that the reason to ablate is for symptoms, and I think to show a reduction in symptomatic AF is a reasonable endpoint.

DR. BORER: Okay.

DR. ZUCKERMAN: Dr. Borer, I'd just like to

clarify for the Panel for a moment because you've hit upon a key point which comes later in the day, but I would underline Dr. Brockman's point about effectiveness being the symptomatic relief being appropriate with the following caveat. We always look upon device technology within a risk-benefit framework, and certainly if the Advisory Panel believes that the risks are minimal for an endpoint that is perhaps not as hard as others that could have been chosen, then your advice is very helpful to tell us whether that risk-benefit profile is an appropriate one.

DR. BORER: That was exactly the point that I was trying to drive at obliquely. So you've said it out loud, and I guess we will get to that. You know, I mean I have no problems with relief of symptoms as an endpoint. The question is are these the kind of symptoms that the devices is intended to be used for? That's all.

There was an analysis in the book, and I missed it if it was discussed here earlier, and I think it was done by the FDA. There were the crossovers from AAD to device, and it was allowed by the protocol, and there was a follow-up of those people. And the results I thought were interesting.

So can you talk about that a little bit?

DR. BROCKMAN: Sure. There were, as you say, 36 patients who crossed over at the final dataset. We had endpoint data on 33 of them. Twenty-one of those thirty-three were classified as successes. So the simple proportion, which is all I can give you, was I believe it was 64 percent. It was very much in line with the chronic success rate in the ThermoCool randomized group.

DR. BORER: That was my point. I just wanted that confirmed because it seems like something important for us to consider.

Are there any other -- yes, Dr. Bilazarian.

DR. BILAZARIAN: I have a question for the post-approval study that Ms. Pinnow presented, and just a general question. I would think as a clinician it would be very helpful to have data from a postmarketing study about the different in outcome in safety and efficacy based on the experience of the site and the operator. And the general question is, is there a precedent for that with FDA? And is that something that could be incorporated in the postmarketing study?

MS. PINNOW: That's a very interesting suggestion. I don't know of any precedents, but that

is something that we could consider in the post-approval study.

DR. ZUCKERMAN: Yes, there are multiple precedents recently for looking at that question, Dr. Bilazarian, because I think what you're getting to is again how generalizable are the procedure results in less experienced hands or not the traditional centers of excellence, and the study can be designed that way. A good example is the way the carotid stent studies were designed, to look at generalizability in the U.S. population.

DR. BILAZARIAN: So would that be a bifurcated part of a single study or a second study, or how was that done previously in carotid stent trials?

DR. ZUCKERMAN: Again, at the appropriate point in time, we want this Advisory Panel to indicate to us what are the key questions. If one of the key questions is how well do the procedure results translate to centers that are not considered the so-called star centers, then the Agency and sponsor can design that study with your help, of course.

DR. WEINBERGER: Just one comment about this postmarketing approval concept. This is turned

into a postmarket approval trial rather than a postmarket approval registry, which was a model that we had in angioplasty and then in stenting, where we garnered a lot of information about real world outcomes. And the design that has been enunciated by both the FDA and sponsor revolve around replicability in the outside world of the very narrow set of questions that were answered here rather than a much broader set of questions which you would like to know about the device. So the question is, does a postmarket approval study preclude or include the possible of a registry?

MS. PINNOW: It actually does include the possibility of a registry. We're looking for input from the Panel on what they would think would be the most appropriate study design for the post-approval study.

DR. ZUCKERMAN: Okay. I just would like to clarify one thing for Dr. Weinberger and the Advisory Panel. Again, Ms. Pinnow in her introductory remarks was very clear that prior to any approval decision, we must have a reasonable assurance of safety and effectiveness, and reasonable would be defined in the context that there's a reasonable likelihood that when introduced into the general U.S. population, the

risk-benefit profile is appropriate. We don't use post-approval studies to figure out what the pre-approval data should have been.

Now, I know it's a little bit confusing because we do have this segment at this point in time, but it's important to first of all think about what the data show for the pre-approval study in appropriate detail.

DR. BORER: Okay. I think we've pretty much exhausted our -- I'm sorry.

DR. JEEVANANDAM: I have a question, and I don't know if it's the appropriate session to ask this in, but the difference between OUS 1 and the rest of the data is very disparate. So let's say this device does get approved and then future devices will have to, let's say a 510(k) be compared to this device, would that OUS 1 data be included in that comparison of another device? You would probably start setting a benchmark of success for other devices, and we just set that benchmark with OUS 1 data included or not because it's so disparate?

DR. ELOFF: At this point in time, the ablation catheters are PMA devices. If in the future a petition was submitted to FDA for a down classification, then we would evaluate whether or not

the devices could be evaluated as 510(k) devices where the standard for market clearance is substantial equivalence to a predicate device.

Right now, these are PMA devices, class III PMA devices which the ablation catheters are, the standard for approval of the device is a reasonable assurance of safety and effectiveness. In this trial, that was done through a randomized control trial versus available therapy. In future trials, as our guidance document on atrial fibrillation trial design and last year's September 20th Panel meeting on atrial fibrillation trial design suggested, once there are one or more approved catheters, a future trial design could potentially randomize against any approved treatment including a catheter or a medical therapy.

DR. ZUCKERMAN: Okay. Dr. Jeevanandam, let me try to take a crack at answering your pivotal question in terms that I'd like the Advisory Panel to think about this afternoon and to discuss because it's a very important question.

Point number one is that we're here to discuss this particular device and this PMA today. Don't worry about device X, Y or Z. We'll contact you at a time in the future.

With respect to this device, we have a very interesting problem as you framed it and others. There appear to be disparate results. As you know, at an Advisory Panel meeting, we never show you the easy stuff, and so Dr. Borer and others will be asking this Panel this afternoon to comment on what this means from a clinical trialist perspective, to comment on generalizability of the results consequently, and potentially how these data could be accurately portrayed in a label if the Panel thinks it's appropriate for approval. But, you know, the key thing that you want to concentrate on has been what's been alluded to recently by yourself and other Advisory Panel members, what do you make of these data given that this is what we have? How generalizable are they? How can you best portray what we have?

DR. BORER: Dr. Kelley.

DR. KELLEY: One more comment. What I find a little troubling is if we look at the disparity between OUS 1 and the other centers. The other centers are centers of excellence, big busy labs with very experienced electrophysiologists. So what's a little worrisome is if we go from OUS 1 to the other centers and our effectiveness drops by half, what are

we going to see when this is generalized? Is it going to fall by half again, which gets us very close to the antiarrhythmic drug people?

DR. BORER: Okay. Well, that's going to be a subject of discussion for later I'm sure.

If there are no other questions, then we'll break for lunch. The only anchor we have in this schedule is the 3:30 open public hearing, which we must hold at 3:30, but we can stop for lunch now. It is 9 minutes and 10 seconds after 12:00. So 9 minutes and 10 seconds after 1:00, we'll reconvene and being the Panel deliberation.

(Whereupon, at 12:09 p.m., a luncheon recess was taken.)

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

(1:12 p.m.)

DR. BORER: Please sit down. We'll start again in 11.5 seconds. We'll have the two panel presentations first, the Panel reviewer David Slotwiner, and the statistical reviewer David Naftel, we'll begin with their formal presentations and then we'll ask questions of the FDA, the sponsor. The sponsor can answer the questions that we asked before lunch.

So why don't we begin with Dr. Slotwiner.

DR. SLOTWINER: Okay. Thank you, Dr. Borer, and I'd like to thank the FDA and the sponsor for such excellent and clear presentations earlier today.

As an electrophysiologist who regularly treats these patients with paroxysmal atrial fibrillation with both antiarrhythmic drugs and off-label use of ablation catheters, particularly the ThermoCool catheter, it's very clear to me how important the data contained in this trial is.

The data presented by the sponsor and the FDA touches upon many of the key questions that have challenged all physicians and healthcare professionals working in this are, and I'd like to

bring up five points that I thought would be helpful to discuss amongst the Panel members. What I'd like to do is bring them up now, and then when, Dr. Borer, you think it's appropriate, we can open them up to discussion.

I have just a few slides, and the first question I have is how important is the CARTO 3-D mapping system? Is it critical to the safety and success that was demonstrated in the trial? I'll go through those pictures up there in a moment, but first, let me just point out that in this study, every patient received either a CT scan or a MRI prior to their first ablation. So that gave the electrophysiologist an enormous amount of information about their anatomy, which is particularly variable when we're talking about the pulmonary veins.

Presumably that 3-dimensional CAT scan was incorporated, merged with the electroanatomical map that was created in each patient, and that was used at the time of ablation. So the 3-dimensional reconstruction of their left atrium was used and then encircling of the pulmonary veins, within the left atria, was performed first. Then the pulmonary veins were isolated.

So the sponsor now is requesting that the

catheter is approvable for use without the 3-dimensional mapping system, and I think that this is question that we as the Panel need to consider. It may sound a little bit technical, but I think it gets at a very key problem in atrial fibrillation ablations, which is partly are we modifying a large amount of substrate or are we just isolating the triggers that initiate atrial fibrillation?

I think based upon how the sponsor designed the trial, it's clear that they believe that left atrial modification with the encircling is important, but to perform that without 3-D mapping system, especially for less experienced operators, I think may be a challenge, and that's why I wanted to show the Panel these slides up here. I don't have printouts unfortunately.

This is a picture I took from Biosense Webster's website. I could have taken from our CARTO system. This is three images of the left atrium, 3-D reconstructions from a CAT scan of the left atrium, and here you see the left pulmonary veins coming off and here are the right pulmonary veins, and here you see the ablation lesions encircling the ostia of both left veins and here you see the right. You also can see, we're able to visualize within the left atrium,

and what you can't see on this picture is that we can actually see our catheter tip in real-time within the 3-D reconstruction of the CAT scan. And the third image is just a slightly different angle of the same picture.

So this is what the investigators in this trial were looking at when they were performing their ablation.

This is a picture, a fluoroscopic picture of what one would see without the 3-D mapping system using just what's called a Lasso catheter, which is a 10 or 20 pole catheter that can be place in the os of the pulmonary vein to map the electrical connections into the pulmonary vein, and here you see next to it our ablation catheter. And you'll have to take my word for it, this is in the left superior pulmonary vein. It just takes practice to know.

Now, if you don't have the 3-dimensional mapping system and the 3-dimensional reconstruction of what the left atrium really looks like, this is what you are left with, and this was done by the investigators in additional to the 3-dimensional mapping system, but what I'm showing here is that same chest x-ray or fluoroscopic imagine with the Lasso catheter, and I want to show you what we have

to look at, what we record from that Lasso catheter.

This is our surface EKG here. There's a little P wave and a QRS, and here you see the recording from the left atrium, and this is a far field ventricular activation, and within this recording of the left atrial activity, what we're seeing is left atrial activation and then conduction into the pulmonary vein, the pulmonary vein potential. So let me just see if my animation will work here.

So the green arrow is indicating where we're recording in the left atrium and the red arrow in the pulmonary vein. And if you bring that over to the recordings, intracardiac recordings, that first bump represents the left atrium, that second bump is the pulmonary vein potential.

So this is before ablation. As we apply radiofrequency energy, you see a delay between that first and second bump as we start to damage conduction or slow conduction into the pulmonary vein. So there's a slightly increased time delay, and then post-ablation, you see we've lost conduction completely into the pulmonary vein. It's very obvious, I'm sure. So this takes a little bit of practice, and for less experienced investigators, I'm

not sure that the same results as demonstrated in this trial could be obtained.

So this again is the 3-dimensional image that we use today, and some people use just the Lasso catheter, very experienced operators, but using the 3-dimensional electroanatomical maps allows us to ablate within the left atrium much more clearly to make sure we're not in a pulmonary vein. So it has the potential to improve safety by not ablating in the pulmonary veins. It can be hard to tell where the os of the pulmonary vein is on that imagine as you can imagine.

Additionally, it helps us avoid certain parts of the left atrium, that is the 3-dimensional reconstruction, helps us avoid certain parts of the left atrium, particularly the left atrial appendage which is easily torn by a catheter, by a less experienced electrophysiologist or even an experienced one. Sometimes we just don't know where our catheters are, and so I think it's important for the Panel to realize that this study was performed using 3-dimensional electroanatomical navigation with CAT scan or MRI images, but the request for approval is to use the catheter without, and does this study support that? I think that's something we need to

discuss.

Another issue is 3-dimensional mapping is required. Does it have to be this system, the CARTO system? That was the only system used in the study. There is one other competing system at the moment which provides essentially identical information which I think could be interchangeable with this CARTO system, but it wasn't used in this particular study.

A second question that I had was the difference between the five variations of this catheter that are manufactured. Are they all equivalent? The study was performed using one version of the catheter. The five versions, just to make sure we're all on the same page, two of the catheters can only move in one direction. So they're unidirectional. And two of the catheters are bidirectional. Two catheters have the ability to be visualized within the 3-D map. Two of them do not. So they can't be used with this 3-D mapping system. And then the fifth catheter is one that is not navigated by hand but is navigated completely remotely by a magnetic navigation system. There's data not included in this study to suggest that that is very safe, and I think there's limited data to

demonstrate how effective it is, but should we give approval to all five versions of this catheter when only one was really studied in this data, I think is an important question.

Let's see. I think, you know, we've all touched upon the European site with discrepant results, but when I was looking at the data, I was struck by the fact that all of the European sites taken together had a better success rate than the U.S. sites, and they've had this catheter longer. So just another question about how important experience with the catheter is for success rates, and how do we measure that going forward to make sure that investigators have enough experience to produce acceptable safety and efficacy results.

The issue of the cavotricuspid isthmus line was discussed this morning. That is an ablation within the right atrium to prevent atrial flutter. It wasn't required as part of the protocol. It was recommended that if atrial flutter could be electrically induced after the atrial fibrillation, that cavotricuspid ablation be performed, but that one European site with rates that were far higher than everybody else's in success did it in most patients.

I personally don't think that that's necessary, but it's a question that's brought up by the data.

And I think the last question I had was the sponsor points out, and the FDA, that afib affects about 2.3 million Americans, yet it was very difficult to enroll 167 patients, and I understand why. One of the reasons is that this catheter was available clinically off-label, and so it was hard to enroll patients because they had the ability to get the ablation otherwise. I think patients also, when they were referred to these centers of excellence for an ablation, were expecting an ablation, but my concern is that the FDA approval of the ThermoCool, if the FDA approves the ThermoCool catheter, it will rapidly be used in a much broader population of patients with atrial fibrillation, patients with structural heart disease, patients with persistent or chronic AF, patients with heart failure, et cetera.

So with a conservative estimate of 2.3 million people in the United States suffering from atrial fibrillation, the sponsor's proposal to study 145 patients who meet the same profile as the patients enrolled in the pivotal trial seems a little bit inadequate, and I'm looking forward to discussing

the postmarket trial options.

Overall, I really want to commend the sponsor for what is clearly one of the most important atrial fibrillation trials performed to date. I personally believe that the primary questions that we need to address as a Panel are not whether the catheter is safe or effective. Of course, we need to address that, but I think we need to discuss whether the 3-D mapping system has to be part of an approved system, whether the data supports approving all five models of the catheter, and how do we ensure post-approval when the catheter becomes used by less experienced operators, in more complicated afib patients, that we obtain the necessary data to monitor the efficacy and safety. Thank you.

DR. BORER: Great. Thank you very much, David. Those were really cogent comments.

I'll ask Dr. David Naftel now to make his comments about the statistical aspects of the study, and then we'll open up the discussion to the Panel.

DR. NAFTEL: So I thought this was really, really interesting as I read this. I wasn't sure if it was a device trial or if it was a statistics test. It's just absolutely incredible. When I read through it the first time, I thought, man, I'm going to have

to have a series of lectures to explain all this once I understand it myself. But I want to just go ahead and tip my hand up front and say that I think the sponsor did an excellent job at conducting the trial statistically. I thought Dr. Berry's tutorial in Bayesian statistics and then the application was extremely good, and then Dr. Thompson, Laura, I thought your explanation was so good. So I just want to say right up front, I thought that the methods were good and the results were also.

So I only have a few comments to help put some of this in context. All of my main concerns have already not only been asked but been addressed. So these are mainly small, but I do want to put it a little bit in context.

The purpose of statistics, this is going to be philosophical for a second, the purpose is to take a whole bunch of numbers and make some sense, to condense them, to look for relations, to test relationships, and the statistician has a difficult job. He or she not only has to know the technical aspects. That has to be done but then they have to be able to convey the results to an audience of educated people, but not statisticians, and that's a very difficult thing to do, and a lot of

statisticians stop short. Again, I'd like to commend both parties, FDA and the sponsor, in that I think you did a nice job at presenting it, and I plan to then copyright rules. I plan to use a lot of those slides.

Now, let's talk about basic stuff. Kaplan-Meier estimation. We all know Kaplan-Meier estimation, and we're all comfortable with it. The standard process is, let's say we're comparing two heart valves, prosthetic valves. You have a moment that the valve goes in. That's the time 0, and then if you're looking at say death, it's easy to define, maybe calculate some kind of P-value and you go home.

This is very different. The guidance document or the draft version makes a really good and I think useful attempt to take this whole clinical issue and put it into something that a statistician can work with. For instance, just the mere definition of time 0 is problematic. We have the 90-day blanking period, and I think that's appropriate, but it's not something that you're taught in a statistical book. You know, what's this blanking period? Suddenly at 90 days I'm at time 0, and I can look at events. That's strange, but I think it's necessary.

Then let me go a little further on this particular issue, and this will be one of my questions later one. The key Kaplan-Meier curve that we all look at, in one version it says events that occurred prior to the 90 day in the treatment group, were counted at time 0. So I should have been able to understand exactly what that meant, and you can tell from the curves, the quick drop off in the ablation group, you can tell that something happened. So I just need to understand that a little better.

Now, the history of randomized clinical trials is you work very hard to set up the sample size, hypotheses. You do everything, and once you've done it, then you monitor closely but you stand back and you do nothing. You don't look. You don't do interim analyses. Just one day the trial is over and then you look.

Now, in the history of clinical trials, people started to realize that there was a waste there that once you finally looked, you'd say, oh, my goodness, there's a huge difference. You know, we could have stopped this sooner, or maybe there's no difference. It was a futile trial. We still should have stopped it sooner. So that's when people started moving to adaptive designs, to Bayesian

analyses, and the world is moving there very quickly, and I think it's appropriate, absolutely appropriate but it must be done carefully. Stopping early is, I'll use the word, a dangerous thing. You have to be careful.

If I may use one quick example, I'm an Auburn football fan, which is not a good thing these days. After the first seven games, we won three games. The other four games we were ahead by more than 10 points at half-time. I lobbied for early termination. And I didn't get it, and we got massacred in the second half in all games. And it actually does apply to this situation. We've said several times as we impute expected results in the patients who haven't been through the whole 90 days, it was said that there's really no difference between the first patients and these that we're imputing.

Well, we actually know from the design of the trial that huge differences occurred, and it's what we're talking about the whole time, and that is the addition of the non-U.S. sites. Now, most of the sites have been entered and the effect was done, but you have to be a little careful to say that the future is going to be predicted by the past, if you want to look at my stock market portfolio.

But on the other hand, the sensitivity analyses that Laura did are perfect, totally handled that issue. So it's really good. So once again I had a question and had a good answer.

But not a few particular details. In the trial design, it was randomized two to one. I couldn't find how the randomization was done, but I'm sure that it was stratified by site, and I would imagine most of us that do this will do some kind of block randomization where every three patients are split two to one. That's a little too obvious. Maybe every six patients, you have four to two. The randomization does not look correct at several of the sites. Several are five and one. There's something I don't understand about the randomization, and usually when you do stratify, you end up with darn close to what you set up with, a two to one. And this is off more than I think it should be. So that will be one question that I ask.

So the big issue, of course, was recruitment. Things were not going well, and I think you encountered unprecedented challenges to timely enrollment. I think you're being liberal in your use of the English language. There are plenty of trials that would say, I've been unfortunate, too,

unprecedented changes, problems with enrollment. Just to be absolutely honest, and that's what we need to be, if something like this happened, it meant some planning up front was not correct, you know, that your estimates were not correct for whatever reason. So your feasibility study or whatever you did, you know, you were wrong.

So that's fine. So you make a midcourse correction. You go outside the U.S., and I have no problem with that. You change to the adaptive design, and I actually have no problem with that. It was done quite well. So not a huge point for me to make, but just something to think about.

Now, the results. Pooling. It used to be that pooling, you only talked about when you were combining sites. You only talked about two things. Did they have the same protocol? And were the patients managed the same?

But now there's the additional thing. Are the results the same?

And there was one quote, I had to write it down, and I forget who said it, but it said the statistical results were insensitive to the exclusion of OUS 1. Well, your final conclusion may have been insensitive, but the results are totally sensitive.

I mean that one site really makes that difference large. But again you've shown me pretty conclusively that without that large site, you still have the same conclusion.

I wanted to comment on the gender analysis. Again, you know, you're in a catch 22. You just have to do a gender analysis. There's no choice, but you need to always remember the study is powered for one overall comparison. So if you don't find differences in gender, that's good. And it was relatively convincing but I just want to remind you, you can't be too quick to claim a great success because the study just simply is not powered for that.

Then as far as the results, I think everybody here keeps looking at that Kaplan-Meier. How can you help it? The Bayesian analysis is at nine months, and if you try to match the Bayesian analysis with the endpoints from the Kaplan-Meier, they're incredibly close. For those of you that haven't grown up as statisticians, you may not know this, but there's almost a religious difference among statisticians between frequentists and Bayesians. It's really amazing how it splits. The Bayesians seem to be winning, which may tell you where I stand, but I will say this. A Bayesian with a non-

informative prior is so similar to a frequentist, the results are so similar and Laura showed that, and I appreciated it. But again, you know, whether you're a statistician or not, the Kaplan-Meier curves are just so convincing.

And then finally, I want to comment, I know we will directly on the postmarket study, if we get to that point, and first I want to compliment FDA. You're so good, and with Bram helping, you're so good at reminding us of the purpose of a postmarket study, and I think we've got that clear, and I think it's really good. And I know the FDA has worked so hard to have better and better post-approval studies. My question will be, and it's always my question in these Panel meetings, is there any result that would happen in the planned postmarket study that would make you take the device off the shelf? Are there any ramifications to the company of the postmarket study, and that will always be my question, and if not, you kind of wonder when we're going to look at the postmarket results and what good they are.

So again, if I may reiterate, I thought it was really a nice analysis, the plan, the implementation. Laura, you really helped shore up a lot of the questions, appreciate that, Laura, and the

rest of the statisticians at FDA. Thank you.

DR. BORER: Great. Thank you very much, Dr. Naftel. Let me ask the Panel, let's switch the order. First, let's ask questions of our two reviewers, but we'll start with the statistical review because we've been putting off the statistical questions, and maybe it's time, and then we'll go back to David's points.

Are there any questions or comments, issues from the members of the Panel, now that we've heard all three statistical presentations about how this analysis was planned and performed? Any concerns?

DR. KARASIK: Yes. I am not a statistician, but I have lots of questions that I'm hoping perhaps you can clear up for me. My first is, I still, despite everybody's reassurance, have concerns about a trial where the statistical method is changed two-thirds of the way through or half the way through the initial, you know, recruitment process, and that makes me very uncomfortable, and perhaps someone could reassure me that this is a legitimate thing to do in a clinical trial.

DR. NAFTEL: I might give a small answer but then leave it up to other people. Clearly FDA was very concerned. In part, if I understood it

right, part of it was they penalized the analysis a bit in that they made the P-value had to be a little bit smaller. So they did that, but I think they were equally concerned and I don't know. I'd love to hear answers.

DR. BORER: Dr. Berry, Dr. Thompson, you want to add to that? Both of you. Either or all.

DR. THOMPSON: Either one. Well, to address your question, certainly we're concerned at FDA when the analysis is changed midway through a trial, and I tried to convey our concern. The process of changing the trial took place over several meetings we had with the sponsor, and some of the things that changed our mind were, one, the sponsor was supposed to be blinded to results, and we had no reason to suspect that they weren't.

Two, we did impose a statistical penalty, if you will, which just basically means the threshold has to be higher. Their evidence has to be greater to kind of account for the fact that suppose they did look at this time, and then decided to change the trial, you know, and do things to make it more likely to reach a successful endpoint.

And third, well, changing methodologies from frequentist to Bayesian, Dr. Naftel indicated

that, and I also mentioned in my presentation, that if you've got what's referred to as a non-informative prior in a Bayesian analysis, it might be a different statistic or a different way of calculating things but you're not getting any new information. So it's almost like you've got a non-informative prior, which means you're not really giving any weight to any other information outside of what you're looking at in terms of the trial. You're sort of repackaging it in a different way, but in some sense, it's sort of like calculating a different statistic. Actually, in this sense, the statistic itself is almost identical because we're still talking about a comparison of proportions.

Another point is if you really still feel uncomfortable with the change, this isn't exactly a panacea, but had nothing been changed at all, I did present an analysis that showed, you know, had things been worse, in the remaining set of patient than they were already, and they had to be kind of a lot worse, I gave pretty conservative scenarios, they still would have met what was the original frequentist endpoint.

So although I do have to say that we really don't recommend that a sponsor change a trial midway,

I have to reiterate that we had several discussions with the sponsor, and we tried other sorts of avenues, you know, extending recruitment in other sorts of ways, and they did do that, but they still didn't quite get the recruitment that they wanted.

Interestingly, I did before the meeting, I asked the sponsor for some information on study dates, when a subject was actually enrolled in a trial, because I was curious to see at the point at which they approached us to change the design, what could they have looked at if they were unblinded to the trial, and actually the results, I do have some slides, but I can probably just say this.

It really wasn't that compelling. If I would have seen these results, I don't think I would necessarily think that I could stop, and the sponsor can probably correct me if I'm wrong. I'm just going by what they gave me. So they came to us around October 20 of 2006, and so I was looking at some of the date information. There were actually around 53 or 51 patients enrolled. There could be one or more, but this is what I got. Thirty-three in the treatment group. So in the treatment group of the 33, there had been 15 chronic failures, and there were 6 successes. So there were 6 patients who had

completed 9 months without a failure, and then there were still 12 in the evaluation period. With the control group, there were, of the 18 enrolled, there had been 15 failures, 2 successes, and 1 was still in the evaluation period.

So I calculated what was the predictive probability of trial success or treatment superiority, only using those 53 enrolled. So the imputation was only to account for the 12 plus 1 still in the evaluation period. And when I did that, I got a predictive probability of .88, which, you know, it's large, but I don't think I would necessarily come to any kind of strong conclusion because it was supposed to be .99. So it could have gone either way.

And again, I'll just have to say we don't like when sponsors come in to change. That's why I don't want to say that it's something we recommend, but I think we can probably feel comfortable about this.

DR. BORER: Great. Thank you. Dr. Berry.

DR. BERRY: So I agree with everything Laura said. The two additional points, one is at the time that we first met, this was very early on. It was about a year before we eventually got an

agreement as to what the protocol would be. There were a small number of patients who had achieved the endpoint as Dr. Thompson indicated. I, of course, knew nothing of the endpoint, and I don't know that the company knew anything of the endpoint.

If we had been approached to design a prospective trial, it would have been exactly the trial that we, in fact, designed, and with respect to company input, we, Berry Consultants, did the entire design. We evaluated various things for the company, but the company never said, oh, well, let's do this instead of that. This was a design that we built, and we certainly built it prospectively.

DR. BORER: Okay. Let me start out by saying you convinced me that the statistics are fine, but let me tell you what my concern was coming in, and maybe you can respond to it just in case there's a little kernel of this in the minds of some of the other panelists.

I read the guidance. I asked Bram for it. I got the guidance. I read a report of the trial, a device trial that had been submitted previously to the FDA and had used the Bayesian approach. I still didn't quite understand.

What I was left with was sort of the

feeling that everyone involved in cardiovascular disease and clinical trials constantly requested the FDA to be more creative and innovative in the use of statistics so that the size of trials can be reduced without altering in a detrimental way the quality of the information that we get, and here the FDA did that. They went and borrowed something that had been done in the cancer field and ran with it.

I mean that's something we should be thankful for, but when I looked at this, it sort of sounded as if, and it brought back something that Dr. Naftel said about the power issue, what happens when people like me see the results of a trial and nominally X is different from Y and a P-value is recorded using conventional statistics that I now understand as frequentist statistics, the P-value comes out .13, and somebody says, well, obviously, you know, that intervention worked. If there would have been more patients, it would have been significant and, you know, that's ludicrous. That subverts the purpose of statistics.

Either you believe that 1 chance in 8, that the result is due to chance alone is compelling, or you don't, and conventionally we say it has to be less than 1 chance in 20.

And clearly the problem is that the power of the trial was not sufficient to be able to see a result of the magnitude that we were looking for, the way it was done.

Well, okay. That's all a preamble to what at first seemed to me was that with using the Bayesian statistic, you took a look and you said, well, we're close but we're not there. But we know if we had a few more patients, we'd be there. So we're going to get a few more patients. We're going to look again, and then we're going to see and maybe we'll get what we want.

That's the way it seemed to me, and with frequentist statistics, that would be a sin, but you convinced me that that's not what's happening.

However, the assumptions that I saw when I looked at the Bayesian analysis was, number one, the issue of exchangeability of the patients. You know, one patient whenever, is he same as another patient whenever, and I guess that's sort of an assumption you have to have whenever you do a trial or otherwise you can't do a trial and draw conclusions, and that the boundary conditions are always the same. And maybe they are and maybe they aren't, over time and as you look and things change.

What jumps out at you here is that the magnitude of the effect at the end of the day was rather large, no matter how you looked at it. So I thought that was pretty good, but I'm not sure that I've been clear, but you see the quandary that I'm in. And, Dr. Berry, you're standing up. So why don't you just relieve me here.

DR. BERRY: Yes, I feel compelled to address the issue, and I think it's absolutely right on. I mean your feeling is what everybody feels, and we understand it.

So you're at .13 and you attempted to go on. If you do, and it's not in the protocol, you don't know what you've got. What we do and, you know, you said a frequentist sin, what we do is we build a better mousetrap, but then we validate it with the frequentist paradigm, false positive rate being paramount. And the FDA is very concerned about false positive rates.

So what we do is this. We say if we come to .13 and the sample size is still rather small, we're not there yet, we're going to go on. If it's .001 or something, we're going to stop. If it's .99, we're going to stop for futility. We write down very specifically exactly what we're going to do at each

of these points, and then we simulate.

We say, let's suppose there's nothing going on, there's no benefit of the treatment, and we generate patients accordingly and we do this, you know, tens of thousands of times, and we come to a point that says .13, according to the protocol, you go on. Okay. We go on. We take the hit so that when we evaluate the proportion of those trials that showed a positive benefit, and that proportion is above .05, we say it's not good enough. We've got to go back and change the design in some sort of fashion like, for example, the hit that you take in the penalty, in the usual frequentist approach, is incorporated within.

So you saw in our study the .98. The .98 was elevated specifically to address this false positive rate, and then we do other simulations to show that indeed we've got a sample size big enough in the range of sample sizes such that we, in fact, do get power.

And the benefit of the Bayesian approach, you know, this kind of thinking is that if you do hit a favorable result with sufficient confidence, then you have a smaller sample size, but you get the benefit of being able to go to the full sample size

in case you didn't.

But all of these are validated in the sense that we simulate and we show that we've got the right false positive rate and power.

DR. BORER: I'm happy.

DR. ZUCKERMAN: Dr. Borer, but your point is a critical one, and it really is incumbent on the Panel to understand that a lot of the necessary simulations are done up front. Perhaps Dr. Thompson wants to comment also.

In fact, the main problem that we get in with Bayesian is that many sponsors don't, so up front the simulations that Dr. Berry is talking about or within FDA, we have problems with reproducing the code and simulations, et cetera.

I want to convey to this Panel that that was not a problem with this particular trial, but Dr. Thompson can comment.

DR. THOMPSON: Well, I don't know if I have anything else to say based on what was already said, but there's one thing that wasn't said. You know, stopping a trial midway for effectiveness is not inherently a Bayesian thing. You know, we have lots of frequentist designs that do that. So I don't want to mix up the two. It is problematic to introduce

that kind of stopping midway but changing from frequentist to Bayesian, you know, it would only really be a big problem if you want to incorporate external prior information. We didn't even talk about that here today. I don't want to open up that can of worms, but I kind of want to separate the two, you know, stopping a trial based on a good result before you wanted to is something separate from Bayesian, and kind of keeping those separate is kind of a good thing.

DR. BERRY: Ten seconds. So I agree. This is not specifically Bayesian. One thing that was, however, is the predictive probability calculations that would allow for stopping accrual and continuing follow-up. That's something that is not standard in the frequentist world and is something we can do with the Bayesian approach because we do the prediction of the future.

DR. BORER: Sounds good to me. Do any of the other Panelists have any questions or concerns about the statistics that you want to raise at this point?

(No response.)

DR. BORER: Okay. If not, then let's go back to David Slotwiner's presentation. David raised

five separate points that we need to consider. We can consider them in the context of the questions that the FDA posed to us, but if anybody has any questions for David about his presentation, now would be the time to ask. Seeing none, let me -- oh, I'm sorry. Go ahead, Dr. Jeevanandam.

DR. JEEVANANDAM: Well, David posed some specific questions about 3-D mapping and whether we should approve all five models or just the one that was tested. And I think in terms of the 3-D mapping, which probably, I mean in my opinion, we should be evaluating the model that was tested, and not necessarily predicting that the other models would work, and I think that if this was done with 3-D modeling and mapping, I think that is what you need to discuss and I think if the particular model that was used should be discussed unless there are very minor modifications to the other models, but I thought the 3-D mapping was pretty critical in terms of knowing where you're ablating. And I mean I don't know whether going in one direction or being able to go in two directions is better than being able to go in one direction. I don't know if that particularly makes a difference, but then, you know, we're bringing in a magnetically guided catheter which we