one year, two years, four years, six years, and eight years. And we are worried that going longer is going to make recruitment of patients and consented patients and the need and the likelihood of patients that have consented to actually come back less and less likely.

For the learning curve, physicians who were initially treating patients with the STAR, six weeks, six months, and one year. Patient follow up and compliance, again we are proposing the eight years based mostly on the difficulty of getting consent and follow up. As everybody knows, historical long-term follow-up rates decrease the longer the studies go out.

We do plan to take measure to improve follow-up rates including visit window reminders and patient cards with visit windows.

At each study, we'll do an assessment of operative site adverse events,

clinical examination, radiographs, and patient evaluations, including the BP score, VAS, SF-36 as well as AOFAS score.

Thank you very much.

CHAIR KIRKPATRICK: Thank you.

I don't know if it is intrinsic to your New York background but that was a good summation at the end.

I would like to now ask the Panel if there are any specific factual points of clarification that you would like to ask the sponsors. Shall we just go around the table then?

Dr. Mayor?

DR. MAYOR: Yes, my largest area of confusion and lacking in insight is the question of the polyethylene's history and its handling through its fabrication, sterilization, and storage. Is there any one of the group presenting that can clarify my understanding of exactly how the polyethylene in these bearings is sterilized and how it is

handled subsequently?

DR. COUGHLIN: In my -- and I'll see if anybody else needs to chime in -- Hoist is the company that makes it in Europe and I understand plates of polyethylene are shipped to Link where they are cut, and then they are CNC cut, and then laser cut to make the implant.

Then the implant is put inside a package that is a plastic-wrapped package in a nitrogen environment exposed to, my understanding is, somewhere between 25 and 29 kiloGrays to sterilize it.

And there has been extensive testing at different periods of time of contamination of the implant has fallen way, way below whatever the standards are -- 100 or 1,000 CSUs, depending on the length of time and how it was exposed. So that's what I understand about the implant.

Any other questions regarding that?

DR. MAYOR: Yes, there seems in the

presentation to be a dearth of focus on the consequences to the mechanical properties of the polyethylene with time subsequent to the sterilization.

Is the package in which the implant contained, granted that it is nitrogen infused, but is it oxygen impermeable on the shelf? And have there been any efforts made to assess the consequences to the gaseous sterilization in regard to polyethylene long-term accountable properties?

CHAIR KIRKPATRICK: If I might, Dr. Mayor, would you mind if they prepare that answer for you?

DR. MAYOR: That's quite all right.

CHAIR KIRKPATRICK: Okay. Because that might require some additional discussion that they may need to consult with other people --

DR. MAYOR: I understand.

CHAIR KIRKPATRICK: -- on their team.

DR. COUGHLIN: We'll get back to you.

DR. MAYOR: All right.

CHAIR KIRKPATRICK: I think the likely time would be during a discussion, we will open with a time where we can answer some of the detailed questions that we have brought up here is we can't get the quick answer.

All right, next? Specific factual questions of their presentation.

DR. PFEFFER: Good morning. How is range of motion determined in all of these patient preoperatively? And was both ankle and subtalar motion measured? And does the combined motion noted in the Buechel-Pappas score refer to the combined motion of the ankle? Or is that referable to hindfoot inversion and eversion?

DR. COUGHLIN: That's a decent question. And when we started this study -- I can you when we started this study, we had the goal of using radiographs as a means to

compute range of motion. That was our goal.

But, in fact, the inordinate amount of radiation that would be required for these multiple visits, we felt that that was not the right thing to do for these patients. And so a goniometer was used as sort of second best.

Now I think this issue washes both ways. We did not compute subtalar range of motion as part of the study. Obviously it was examined in these patients but we examined ankle motion, which we all know there is a component of hindfoot motion with it.

Now it is true. And people who have arthrodesis sometimes still note that they have some range of motion as well. So the question washes both ways.

DR. PFEFFER: May I ask another factual question?

CHAIR KIRKPATRICK: Yes.

DR. PFEFFER: How was osteoporosis determined in your patients? That was one of the exclusion criteria. And one patient was

eliminated from the study because of "severe" osteoporosis. How was that information -- what was the objective data for that measurement?

DR. COUGHLIN: Preoperative osteopenia was our means of inspecting radiographs. And when we had someone that was on the line, a DEXA test was performed. And that patient that was excluded was my patient who failed a DEXA test.

CHAIR KIRKPATRICK: Thank you.

Dr. Propert?

DR. PROPERT: I just have one question. And I'm sorry to have to ask about the imputation. And I suspect the answer to this won't be now but I was really confused about exactly what sort of imputation was done and would like some clarification on that specifically how the confidence intervals were adjusted for the imputation that was done.

And, again, I don't expect an answer at this moment. We could do this after

the break.

DR. COUGHLIN: I'd appreciate doing it after the break. As I said, as an orthopedist with imputations, we need some statistical help here. But we will report back. Thank you.

CHAIR KIRKPATRICK: Thank you.

Dr. Skinner?

DR. SKINNER: I have no specific factual questions at this time.

CHAIR KIRKPATRICK: Thanks.

Dr. Goodman?

MEMBER GOODMAN: Well, I have many questions that I'll ask later on but in particular I was wondering whether the sponsor is going to show some of the radiographic examples of loosening, subsidence, et cetera, at some point in the future?

CHAIR KIRKPATRICK: Thank you. We will ask them to try and prepare that for later discussion as they didn't have that readily available in their presentation.

Thanks.

DR. WRIGHT: I guess I have a lot of questions that I'll ask later also but I have two specific technical questions. One was the approach for the control group for an arthrodesis in that most of the complications had to do with the procedure were this anterior surgical approach. And was there any control for the arthrodesis group?

The second question was does this procedure include a fusion of the tibial/fibula joint like the prosthesis that is on the market now? I didn't see anything in the literature -- in the report on that also.

DR. MANN: The lateral approach was used for the ankle fusion. None of them were done through an anterior approach because the fibula was removed in doing the fusion and that obviously can't be done through an anterior approach.

As far as fusion of the distal

tibial/fibula joint, this is not required in this prosthesis. That is one of the big advantages of the prosthesis is that fusion does not have to occur in that region.

CHAIR KIRKPATRICK: Does that answer your factual questions, Dr. Wright? Thank you.

Ms. Whittington?

MS. WHITTINGTON: The results were significantly different with the new surgeons in the later part of your presentation. Is your training program that you have designed for additional surgeons to add to this, does that mirror the experience of what you gave those new surgeons that you indicated as a group?

DR. MANN: Well, the procedure is based on experience. There is no doubt about it. And we learned that the hard way to a certain extent.

As time goes on, we've learned many tricks and many pitfalls of the procedure that

we can point out to the surgeon who is learning how to do this. I think between the knowledge we have gained in this study plus the training programs that we plan on putting on, that the new surgeon starting out is going to be much farther ahead than we were.

They still run into problems. There's no doubt about it. This is technically a difficult procedure. And orthopedic surgery is a hands-on type of event.

And when you are learning something new, you make mistakes. The question is how can you minimize them. And I think that is by pointing out errors that we have made along the way and also be giving them good training.

MS. WHITTINGTON: Maybe I didn't clearly ask my question. The data on the slide showed that the outcomes of new surgeons that did some of the continued access procedures, that their complication rate, their rates, their outcomes were better than

the initial cases that were done by those investigators. What kind of training was done for that new surgeon group? Or, more specifically, is that training comparable to what you have proposed for training new surgeons to do this procedure?

DR. MANN: Those surgeons in that group were part -- were usually associated with a surgeon already doing the procedure. So that they had scrubbed in on probably 30, 40 cases by that time before they did their own case.

In a perfect world, that is what you would do. But it is not a perfect world.

MS. WHITTINGTON: Okay, thank you.

DR. MANN: Sure.

CHAIR KIRKPATRICK: Ms. Adams?

MS. ADAMS: No questions.

CHAIR KIRKPATRICK: Thank you.

I have just a couple. You mentioned the advantage of the bone resection being 10 to 12 millimeters. I'm assuming that

was total resection based upon your slide as opposed to 10 to 12 off each bone? Is that correct? I see you nodding. That's adequate, I think. Thank you.

So for the record, they indicated yes, it was total resection.

The second question which, if it requires your preparation, that's fine, I didn't catch the causes of the four deaths. If you could look that up and let me know, I'd appreciate it.

And then the final question on facts is the HO incidents overall, you mentioned that you had a couple of reoperations for heterotopic ossification. Can I have the total incidents? And then the number of those that went on to surgery?

And also if we could have a radiographic or a clinical description of where that HO occurred, that would be beneficial for us. Thank you. Is that something you have available right away? Or

do you need time? You need a little bit of time? Okay. Thank you.

With that, I'll just review once again, are there factual questions from the Panel for the sponsor?

(No response.)

CHAIR KIRKPATRICK: Seeing none, we will now take a brief break. Oh, I'm sorry, one more.

DR. PFEFFER: Could you clarify for us the use of body mass index versus absolute weight in the initial study? You don't -- you refer to body mass index in the study but you say that the exclusion criteria for the ankle is 250 pounds, which is independent of body mass index. What was the reason for that?

DR. SALTZMAN: I think, you know, the concern when the initial inclusion/exclusion criteria were developed was that a patient would be under a certain weight, not a certain size but a certain weight because the weight is going through

that limb when they are in single stance. So if you weigh 500 pounds, you could -- if you were 14 feet tall, have a pretty good body mass index.

So I think that is the main issue is the weight through the limb rather than sort of the size of the person -- rotundness of the person which is what the body mass index shows you.

DR. COUGHLIN: The body mass index was probably my interest. And I agree with what Dr. Saltzman said. But finally it became too difficult to get our whole group together on this. So we drew a line in the sand on a certain weight. And just said you can be three feet tall or seven feet tall. But this is how much you can weigh and beyond this, it is an exclusion.

DR. PFEFFER: So you would agree then that body mass index might be more accurate because as BMI goes up, the size of the prosthesis probably goes up, whereas

weight goes up, the size of the prosthesis doesn't necessarily go up. No? We have disagreement.

Well, you guys think about it because you disagree. And then we can get back to that question.

CHAIR KIRKPATRICK: Thank you.

Last opportunity for factual questions. And please raise your hands so I don't miss out on you.

(No response.)

CHAIR KIRKPATRICK: Thanks. We will now take a brief break. In the interest of making it easy, my watch has three minutes until. Why don't we say we will have a 13-minute break and resume at ten minutes after. And if you want to synchronize your watch for clarity, that would be good.

Panel members, please remember that there should be no discussion of the PMA during the break among ourselves or with any member of the audience.

So we will, again, resume at ten minutes after. Thank you.

(Whereupon, the foregoing matter went off the record at 10:00 a.m. and went back on the record at 10:13 a.m.)

CHAIR KIRKPATRICK: Thank you. It is now ten minutes after ten and we appreciate everybody getting back. I would like to call the meeting back to order.

The FDA will now give their presentation on the issue. And I believe it is going to start with Mr. Pinder. You have an hour. Thank you.

Oh, I'm sorry. The Panel members, in our blue folders, you will find copies of the slides if you wish to look at a paper copy as well as the screen. Thank you.

MR. PINDER: Good morning. My name is Bryan Pinder and I'm the lead review for the STAR ankle PMA.

My section of the presentation will

be brief and will cover reasons for going to Panel, device description, preclinical testing, and STAR study design.

Dr. Popovic will present the clinical results. And Mr. Zhou will summarize the statistical information in the PMA.

This morning's presentation will conclude with Dr. Wang and an assessment of the applicant's post-approval study. And after lunch we will present eight questions for Panel discussion.

And I'd just like to mention that there will be some repetition in my discussion but we feel that a little bit of repetition will be necessary to frame the questions that we will be presenting later on.

So the main reasons for today's Panel meeting are as follows:

First, the STAR ankle is a Class III device and is the first of a kind, non-constrained, mobile bearing, total ankle system seeking a PMA.

Next, there is a preclinical issue regarding the adequacy of wear testing.

Finally, there are several clinical issues regarding a revised analysis of radiographic success, continued access follow up, surgical modifications, and learning curve determination.

So the applicant has already adequate described the device so I will be skipping this slide.

So the applicant performed several preclinical tests and of these, I would like to discuss wear testing. As the applicant has already described, wear testing was performed on five samples in a joint simulator. Compression force was held relatively constant at 3,000 Newtons while the joint simulator rotated and translated the device throughout normal ranges of motion. All samples survived ten million cycles without failure.

The Agency has questions about the loading regime that was used. Articles were

cited by the applicant showing that ankle joint forces range from two to 5.5 times body weight during normal gait. According to the exclusion criteria for the STAR ankle, the maximum weight for a STAR patient is 250 pounds. This results in worst case compression loading condition of 6,116 Newtons, which was not utilized in wear testing.

In addition, clinical-observed fractures of the mobile bearing were not mimicked with this testing. So we will be asking the Panel a question about the adequacy of wear testing -- of the preclinical testing.

So the applicant has already adequately described the indications for use. And they have also gone over the major contraindications. So I'll just move it along.

I will now discuss the study design and history, which will be the subject of several of the Panel questions discussed later

this afternoon.

The applicant's IDE protocol was conditionally approved in June of 2000 and fully approved in November of 2000. The study was designed to be a prospective, multi-center, nonrandomized, concurrently controlled clinical study to evaluate the STAR ankle. The study was also designed to be a non-inferiority study to test the hypothesis that the STAR ankle is as safe and effect as arthrodesis.

Efficacy and safety were the two primary study endpoints. The primary efficacy endpoint was the mean total Buechel-Pappas scale score measured at 12 months with further confirmation at 24 months. The BP score is based on a hundred-point scale consisting of subscales for pain, function, range of motion, and deformity.

Success was originally defined as a minimum 40-point increase in BP score from baseline.

The STAR ankle patients have a natural advantage over arthrodesis control patients in the section assessing range of motion. Consequently, the applicant was asked by the Agency to conduct a post hoc analysis using a modified BP score which excluded the 15 points contributed by range of motion.

Also, the continued access cohort was intended to have an identical assessment with the addition of the American Orthopaedic Foot and Ankle Society scale for STAR patients.

Just one point here, although it wasn't listed as a primary safety endpoint in the IDE or PMA explicitly, the Agency always considers safety as a primary endpoint. And as was previously stated, the primary safety endpoint per the original IDE protocol was a composite endpoint derived from three criterion: no device failures, revisions, or removals, radiographic success, and no major complications.

And it should be noted that there were proposed modifications to the original radiographic success analysis and these modifications will be discussed shortly.

Overall patient success was another measure calculated by the applicant. This was not considered a primary endpoint. Patient success was defined as both success in safety and efficacy.

As previously stated, the applicant performed post hoc analyses on the radiographic assessment. In the original analysis provided in the PMA, radiographic failures at six and 12 months were carried forward as failures, irrespective of possible success at 24 months.

In a revised post hoc analysis, the applicant identified seven patients who were radiographic successes at 24 months that had earlier failures carried forward. If these seven patients are included as safety successes, the success rate obviously

increases but the 15 percent non-inferiority margin delta was not met.

It should also be noted that a similar post hoc analysis was not performed on the control patients. An additional post assessment was made. In the original PMA submission, radiographic failure was defined as radiolucency, tilting or migration greater than four millimeters.

Under the revised assessment, radiographic failures at 24 months could be considered as successes if they are clinically successful at 48 months with no apparent progression of radiographic failure. This re-analysis effected five patients.

If these five patients are included as successes along with the previous seven patients, the success rate increases to the point where the 15 percent non-inferiority margin delta was met. So although these two subsets of patients are relatively small in size, you can see the effect that they have on

the safety delta.

And the Panel will be asked a question about the appropriateness of this revised analysis.

I'll conclude my portion of the presentation with patient follow up. In total, there were 627 investigational subjects and 66 control subjects. The pivotal study had two groups: unilateral and bilateral. At 24 months the pivotal study unilateral group had a 96.7 percent follow-up rate while the control had a 77.4 percent follow-up rate.

As was previously stated, results for the bilateral group were used for safety successes only so no new additional information will be presented concerning the bilateral group.

The continued access group had approximately 66 percent follow-up rate for those subjects that have reached the two-year endpoint. Radiographic analysis was performed on only 80 patients from the first arm of the

continued access study.

And now Dr. Popovic will present the clinical results.

DR. POPOVIC: Thank you, Bryan.

Good morning. My name is Neven A. Popovic. I'm an orthopedic surgeon.

My presentation will deal with clinical aspects of the study, primarily the operating or surgical data and results of the primary efficacy endpoint, the composite safety endpoint, and measurements of the overall patient success and secondary endpoints.

Anesthesia time, surgery time, and length of hospital stay were similar for the control and the STAR patients. Local anesthesia use was greater in controls than the STAR patients. Estimated blood loss was less in the STAR patients than the controls.

The continued access study patients had similar amount of surgery-related blood loss as the pivotal study STAR patients.

Fewer continued access patients were operated under general anesthesia. There was a slight decrease in the length of hospital stay for the continued access patients.

I will go into greater detail as to the difference in surgical results between the STAR pivotal study patients and the continued access patients when I discuss the results of the continued access study.

The primary efficacy endpoint is based on the Buechel-Pappas scale with the efficacy success defined as greater than a 40-point increase. It received an individual patient efficacy success rates at 12 and 24 months. For example, at 24 months, approximately 15 percent of individuals in the control group and 58 percent of the STAR patients had increases in the BP score of greater than 40 points.

Looking at the mean BP score, including the range of motion segment and the BP score without the range of motion segment,

we see that the STAR cohort has increased 40 points from the baseline value when the range of motion segment was included and approximately 37 points when the range of motion segment was excluded.

As previously noted, the composite safety endpoint is derived from various data segments which have been addressed previously.

Major complications are also noted by the applicant.

This slide lists the number of adverse events for the total patient population. For example, two bone fractures have been noted in the control group of 66 patients. Also note that an individual patient can have more than one adverse event. Comparing the more frequent operative site events in the pivotal study group, the STAR patients had statistically significant increases in frequency of bone fractures, bony changes, adjacent nerve injury, and general bone problems such as bone dehiscense, delayed

bone healing, or skin necrosis.

Twenty-one and a half percent of STAR patients and 16.7 percent of the control patients required additional surgical intervention. Major complications were noted in 8.9 percent of the STAR patients and one and a half percent of the controls.

Please note that the numerical data I have presented and will present in some of the future slides is based on the PMA submission. Also note that some of the numbers and percentages presented earlier by the applicant may be, in some cases, different. The difference may be due to variations or truncation of time used for data collection.

For example, in the pivotal study, group adverse events to 24 months do not include or capture all the adverse events that occurred during the study duration, thus, the difference in numbers of adverse events. This observation raises a question regarding

adequacy of a 24-month length of follow up as the numbers of adverse events and revisions have been noted after the 24-month observation period.

In this slide, we are comparing the numbers of STAR and control patients with various types of surgical interventions. Note that an individual patient can have more than one surgical procedure. Certain surgical interventions were more common in the STAR patients than controls. These additional surgical procedures included reoperations and device revisions. Revisions were done in approximately 11 percent of the STAR patients and approximately six percent of controls.

The control patients had a greater percent of minor procedures such as surgical hardware removal which was statistically significant.

In general, the STAR patients had statistically significant higher rate of major operative site procedures than the control

patients, 14.6 percent for the STAR and four and a half percent for controls. The most common major surgical procedure in the STAR patients was the device component removal noted in approximately 11 percent of those patients.

The most common major surgical procedure in the arthrodesis patients was hardware removal and/or fusion of the adjacent joint, which was noted in six percent of the patients. Once again, please note that a patient can have more than one surgical intervention.

The mobile bearing removal and replacement were the more common surgical interventions in the STAR patients and the numbers are listed.

Surgical technique was already addressed. And I should that the technique changes were made gradually over the course of the study.

This slide may be a good time to

compare the surgical results and adverse events between the pivotal study patients and the continued access patients. The continued access cohort, when compared to the pivotal STAR cohort, had a statistically significant decrease in the rate of bone fractures, post-surgical pain, and additional surgical interventions, including revisions or other types of surgical procedures.

The rate of major complications, although decreased in the STAR continued access cohort, was not statistically significant when compared with the STAR pivotal study cohort.

No appreciable reduction in local injury is noted between the STAR pivotal and the continued access patients.

The pivotal study control patients had a lower frequency of bone fractures, nerve injury, bone problems, and rates of major complications than either the pivotal or the continued access STAR patients.

The radiographic success or failure play an important part in several of the composite study outcomes such as the composite safety endpoint and the overall patient success. Therefore, a counting of radiographic data warrants some attention. Out of 158 STAR patients in the pivotal group, 151 patients had one or more radiographic evaluations at predetermined time periods.

For example, at six months, approximately 94 percent of patients had a six-month evaluation while approximately 85 percent of the patients had a 12-month radiographic evaluation. Not all patients with six-month evaluations had a 12- or 24-month radiographic evaluation.

The original radiographic criteria were noted by Mr. Pinder and the sponsor. The sponsor has requested changes in analysis of radiographic data as noted previously by Mr. Pinder. I will reemphasize those proposed changes.

The proposed changes include consideration of patients previously diagnosed as radiographic failure at six or 12 months, as radiographic success if they met the radiographic success criteria at 24-month follow up. Using the proposed analysis, seven additional patients would be added to the radiographic success column at the 24-month period.

In addition, the sponsor has proposed declaring five patients with radiographic failure at 24 months as radiographic success based on their clinical outcomes at 48 months and an apparent lack of progression of radiographic findings at 48 months. Using these criteria, five additional patients would be added to the radiographic success column at the 24-month period.

I should note that 48 month clinical and radiographic evaluations were not available for all surgical patients and the data on these five patients may not be

applicable to the radiographic or clinical results of the entire patient population.

Using the original radiographic analysis at various time points, six, 12, and 24 months, the STAR patients demonstrated an increase in radiographic failure over time while the arthrodesis group showed decreased signs of radiographic failure.

At six months, the STAR non-accumulated radiographic failure was noted in four patients for the rate of 2.7 percent while at 24 months, there were a total of 13 patients, or approximately nine percent of evaluated patients, meeting radiographic failure criteria.

Here we see the times of initial radiographic failure. Patients with radiographic failure at an earlier evaluation period were not carried as failures into the next radiographic evaluation period. With increasing time, there was a greater number and greater percent of patients with newly-

detected radiographic failures.

At 12 months, radiographic failure was initially detected in six STAR patients or approximately four percent of all evaluated patients at that time period while at 24 months, newly-detected radiographic failure was noted in ten out of 141 patients or approximately seven percent of the total evaluated patient population.

In the second column, we note the total number of patients with radiographic failure at each specified time point.

Looking at the radiographic success rates using the initial and the new proposed radiographic success or failure data analysis, we can see a significant change in percent of radiographic success in the STAR patients from approximately 85 to 94 percent. Once again, I should note that the radiographic success/failure rate is an important part of several composite study outcomes.

As previously noted, the overall

patient success contains various factors including the radiographic success analysis. The overall patient success rates are presented using the original radiographic analysis criteria and the proposed new radiographic analysis criteria. As noted, changes in radiographic analysis have a profound effect on the STAR overall success rates, going from approximately 71 percent using the initial method of radiographic data analysis to 79.6 percent using the new proposed method of radiographic analysis.

In general, the radiographic findings are useful indicators of total joint arthroplasty clinical success or failure. The Panel will be asked to address the radiographic evaluation criteria, the radiographic findings, and the proposed analysis of the available radiographic data requested by the applicant.

The secondary endpoints were noted and defined so I will go into the summary of

the secondary efficacy endpoints. Basically, the STAR cohort had higher function in range of motion BP subscales. There was similarity between the STAR and the controls in total BP scale. The STAR patients had a slightly higher VAS value for post-surgical pain.

The patient satisfaction survey, the quality of life surveys, as well as the medication usage were essentially similar between the STAR and the control patient populations.

In general, it is accepted that surgical outcomes improve with the surgeon's experience. The learning curve for any surgical procedure is dependent on a multitude of factors including, among others, the complexity of the surgical procedure, training of the individual surgeon, as well as the individual surgeon's motivation and general skill level.

In the current study, there were significant variables such as development and

modification of the surgical procedures as the study progressed. The surgeons participating in the pivotal study also participated in the continued access study.

The three new surgeons participating in the continued access study were trained in the implantation of the STAR ankle and had benefitted from the experience of the initial surgeon. They have also used the current surgical procedural modifications.

Thus, using the data provided by the sponsor, one is hard pressed to estimate the number of patients constituting the learning curve for the STAR ankle arthroplasty.

I thank you. This concludes my presentation. The next speaker will be Mr. Jack Zhou from the Division of Biostatistics. He will present the statistical overview. Thank you.

MR. ZHOU: Thanks, Neven.

Good morning. I'm Jack Zhou, the

statistical reviewer for this PMA. I will discuss the STAR clinical studies from a statistical perspective.

This is the outline of my presentation. First, I will talk about the pivotal study design and conduct. Then I will compare the demographics and baseline characteristics of the STAR and control patients. I will discuss the results on the primary efficacy endpoint, on the primary safety endpoint, followed by a brief discussion of the sponsor's meta-analysis. Finally, I will conclude with a summary.

The pivotal study was concurrently controlled but not randomized. Ten sites enrolled exclusively STAR patients. And five sites, they enrolled exclusively arthrodesis control patients.

Such design introduced confounding effects that were difficult to control for as the observed treatment difference between the STAR and the control could be solely

attributed to site difference.

To illustrate what I meant by that, on the slide I'm jumping ahead to show you the 24-months Buechel-Pappas score, which is the primary efficacy endpoint of the study by site. Note these are real data from the pivotal study.

On the left side, we have five control sites with various 24-months' BP scores that are statistically different from each other. In other words, the variability among different control sites cannot be explained by chance alone.

On the right side, we have a similar situation. The variability among the ten STAR sites cannot be explained by chance alone either.

Therefore, overall we have heterogenous STAR sites compared against heterogenous control sites. When the sites in the same group are pooled together, we do see a difference between the STAR and the control

group. But it is difficult to attribute this difference to the device or to the sites based on the BP scores alone.

Getting back to the pivotal study design on the next slide, the original sample size estimation planned to enroll 158 STAR patients and 79 arthrodesis controls based on individual patient safety endpoint. You might have heard slightly different sample sizes from the sponsor earlier this morning.

The difference is probably due to the estimated loss of follow ups. As in most clinical studies, this pivotal study budgeted 15 percent patient loss of follow ups at the planning stage. And the final FDA approved sample size was 158 STAR patients and 79 controls as shown on this slide. However, difficult was encountered in enrolling control patients and only 66 arthrodesis patients were enrolled by PMA submission, including three patients not due for their 24-month visit.

Please note the pivotal study is

technically not completed. All analysis could be considered unscheduled interim analyses which are subject to potential biases and Type I error inflation.

Patient follow-up at 24-months in the pivotal study are shown on this slide. Please note only 71 percent of the control patients had any 24-month follow-up data as proportionately more control patients were lost to follow up. Such a large percentage of missing control patients may reduce the power of the study and make it difficult to avoid biases in statistical analysis if the patients were not completely missing at random.

For example, if the missing patients were more likely to have experience favorable outcomes, ignoring these missing patients will result in a bias estimate against the control.

Since the pivotal study was not randomized, we do not expect balanced patient characteristics between the STAR and the

control group. Compared to the STAR patients, the control patients were younger, had more post-traumatic arthrosis, higher baseline Buechel-Pappas scores, and lower baseline VAS scores. The unadjusted p-values for these comparisons were statistically significantly or close to being significant.

Another way to compare baseline patient characteristics is to use propensity scores. A propensity score analysis evaluates each patient's probability of being assigned to either one of the groups if this had been a randomized study. It is a more comprehensive way to assess baseline comparability than to test each covariant individually.

The sponsor's propensity score analysis included several variables believed to have potential impact on patient outcomes and the propensity score quintiles are shown on this slide. Please note one of the propensity score quintiles has only STAR patients but no controls which means some STAR

patients with certain characteristics had no corresponding control patients to compare with.

This lack of complete overlap of propensity score quintiles raises doubts on the comparability of baseline patient characteristics of the STAR and the control group. And suggests imbalance that existed between the STAR and the control group may not be easily adjusted statistically.

As mentioned earlier, the primary efficacy endpoint of the pivotal study is the Buechel-Pappas score. The pre-specified non-inferiority margin delta is ten points in the original Buechel-Pappas scale. Therefore, the primary objective of the study is to demonstrate the average Buechel-Pappas score for the STAR patients is no more than ten points less than the control patients.

It is important to understand that it is not the observed difference between the two groups that must be less than ten points

but the lower bound of the confidence interval.

This is further illustrated on the next slide. In the pivotal study, we were trying to determine whether the average BP score for the STAR group is worse than the control by more than the non-inferiority margin delta. This can be accomplished by putting a one-sided 95 percent confidence interval around the observed average BP score difference between the STAR and the control group. And comparing the lower bound of the confidence interval with minus delta.

If the entire confidence interval lies above minus delta, as in Case A here, non-inferiority is achieved. If the confidence interval process minus delta, as in Case B, non-inferiority cannot be established.

As you can see on the next slide, STAR patients achieved higher Buechel-Pappas scores at 24 months than the control patients in both unadjusted and covariate adjusted

analyses. If you look at the last column, you will see non-inferiority was established in both intent to treat and per protocol population under several different scenarios.

However, by design, STAR patients had a natural advantage over control arthrodesis patients in range of motion, which is a component of the Buechel-Pappas score. It appears that removing the range of motion component from the BP score will make a better comparison for this non-inferiority study.

Therefore, similar analyses were conducted on the modified Buechel-Pappas score by excluding range of motion. And the results are shown on the next slide.

As you can see, the STAR patients showed similar modified Buechel-Pappas scores as the control patients at 24 months. And if you look at the last column again, non-inferiority was achieved in both the intent to treat and per protocol population under several different scenarios.

Please note, as previously discussed, the extent of the imbalance that existed between the STAR and the control population may not be easily adjusted by statistical modeling.

Moving on to the primary safety endpoint on the next slide, the primary safety endpoint is individual patient success whose definition you are already familiar with. Fifteen percent non-inferiority margin delta was pre-specified for patient safety success. And the sample size was calculated based on this endpoint.

Again, it is important to understand that it is not the observed difference between the two groups that must be less than the non-inferiority margin delta but the lower bound of the confidence interval.

As you can see on the next slide, STAR patients showed lower safety success rate at 24 months compared to the arthrodesis control patients. The non-inferiority margin

was not met in the per protocol or intent to treat population. Covariate-adjusted analyses gave similar results.

You've heard this multiple times. After initial PMA submission, the sponsor informed the FDA that seven STAR patients had early radiographic failures and were incorrectly carried forward as failures in the original safety analyses even though their 24-month radiographs showed success.

The next slide shows that if these seven patients were classified as successes, shown as modified Interpretation No. 1 here, the 24-months overall safety success rates for the STAR group would improve to 76 percent in the completers population, which includes all patients with 24-months safety data. However, when compared with the 83 percent success rate of the control group, the STAR group still did not meet the non-inferiority margin.

The sponsor also conducted analysis in which five additional STAR patients with

certain radiographic findings were counted as successes, shown as Modified Interpretation No. 2 here. In this scenario, the 24-month overall safety success rates for the STAR group further improved to 80 percent, which would meet the 15 percent non-inferiority margin in the unadjusted analysis.

The results on the intent to treat population and covariate-adjusted results are currently no available for the modified radiographic interpretations. Please note, the sponsor's modified radiographic analyses were conducted after seeing the results of the original analysis. You will be asked to comment on the sponsor's modified radiographic interpretations.

The sponsor claimed that surgical techniques and instrumentation improved during the pivotal study. And the safety success rate increased in the continued access study as shown on this slide. However, only 80 continued access patients received independent

radiographic reviews, which made direct comparison to the pivotal study difficult as a radiographic failure was a component of the composite safety endpoint.

To supplement the safety data of arthrodesis control patients, the sponsor conducted literature-based meta-analysis. Forty-two articles with more than 1,200 patients were reviewed. And 12 articles with 413 patients were included in the meta-analysis.

According to the sponsor analysis you saw earlier this morning, historical complication rates of arthrodesis were comparable to the rates observed in the control arm in the pivotal study. However, the post hoc nature of this analysis and the large number of excluded articles and patients made it difficult to assess the extent of selection bias.

To summarize, the pivotal study's nonrandomized design created confounding

effects that are difficult to control for. The incomplete control enrollments and poor follow up further weakened the value of the concurrent control group.

The comparability of STAR and control population is questionable. And the extent of the imbalance that existed between the STAR and the control patient may not be easily adjusted by statistical modeling.

Ignoring the comparability issue, STAR patients may have shown non-inferiority to arthrodesis control in the primary efficacy endpoint Buechel-Pappas score. However, depending on different radiographic interpretations, it is not clear whether non-inferiority was established in the primary safety endpoint.

Evaluating the safety profile of continued access STAR patients is challenging due to the incomplete follow up in this cohort. The sponsor's post hoc meta-analysis excluded a large number of articles and

patients, making it difficult to assess the extent of selection bias.

This concludes my presentation. I will now turn the podium to Dr. Cunlin Wang who will present the postmarket study.

DR. WANG: Good morning, distinguished Panel members and the welcomed guests. My name is Cunlin Wang. I'm an epidemiologist in the Office of Surveillance and Biometrics, CDRH. And Dr. Hefflin and I are the epidemiological reviewers for the STAR ankle post-approval study. And I will now present our summary and assessment of this study.

And first please note the discussion of post-approval study prior to a formal recommendation on the approvability of this PMA should not be interpreted to mean that FDA is suggesting the Panel find the device approval. The plan to conduct the 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 approval.

The main objective of conducting post-approval studies is to evaluate the device performance and potential device-related problems in the broader population over an extended period of time after premarket establishment of device safety and effectiveness.

Post-approval studies should not be used to evaluate unresolved issues from the premarket phase that are important to the initial establishment of device safety and effectiveness. And generally the reasons for conducting post-approval studies are to gather postmarketing information including longer-term performance of the device, community performance, which is the device performance

in the broader patient population treated by average physicians as opposed to highly-selective patients treated by leading physicians in the clinical trials.

Post-approval studies are also used to evaluate the effectiveness of device utilization training programs and the evaluation of the device performance in a subgroup of patients since clinical trials tend to have a limited number of patients and will not include all subgroups of the general patient population.

In addition, post-approval studies are also used to gather data on device real world experience and to monitor device-associated adverse events, especially rare adverse events that were not observed in the clinical trials.

And finally, post-approval studies also include issues and concerns raised by Panel members to be addressed.

And this slide provides an overview

of the post-approval study protocol submitted by the sponsor. The objective of the sponsor's post-approval studies are to evaluate the long-term revision or removal rate for the STAR ankle and assess the learning curve of physicians who are initially treating patients with the STAR ankle after device approval.

To achieve these objectives, the sponsor proposed a two-component prospective cohort study without a control group, a long-term follow-up component, and a short-term physician learning curve component.

In the long-term follow-up study, the sponsor proposed to recruit all surviving STAR ankle patients from the continued access study who have not had a revision or removal and follow them after ten years post operation with clinical evaluation at 48, 72 and 96 months post-operation in addition to the PMA data. The primary outcome is device revision and removal.

In the physician learning curve component study, the sponsor proposed to enroll five new investigators and 125 net STAR ankle patients to examine the learning curve of surgeons who began their clinical experience with the STAR ankle after device approval. These patients will be evaluated at baseline, six weeks, six and 12 months post operation. The primary outcome is major complications.

In addition, radiographic images will be obtained and interpreted by the treating surgeons during each clinical visit.

There are a few issues to consider regarding sponsor's post-approval study plan. First, the sponsor's proposed study is descriptive and not hypothesis-driven even for the subgroup analysis among STAR ankle patients. In general, we recommend the post-approval are hypothesis driven since it will provide greater scientific rigor and a firmer basis for the postmarket regulatory action if

indicated.

And the second, the sponsor's protocol lacks a control group. The absence of a control group significantly diminishes the scientific rigor of the study and also limits the meaningful interpretation and utility of the study results.

Third, the long-term follow-up component study only consists of patients from the continued access study and there is insufficient data on the representativeness of the patients and physicians in the continued access study. This may limit the generalizability of the study results, limit the ability to examine device performance under actual conditions of use, and prevent the fulfillment of some of the requirements of the study.

Fourth, the sponsor's study consists of eight years follow up. This will provide important longer-term data regarding device performance. But the challenge of

posed by loss of follow up over this extended period should not be underestimated.

In the PMA study, the loss to follow-up rate for arthrodesis control group was 23 percent. And only 66 percent of the STAR ankle patients who reached the 24-month post-operation time point in the continued access study had data collected. Significant loss to follow up requires comprehensive study, including measures that prevent a loss to follow up and the compensatory measures when loss to follow up occurs. A comprehensive plan to minimize loss to follow up is absent from the sponsor's current protocol.

If the device is recommended for approval with the condition of a post-approval study, there are a few issues related to the sponsor's post-approval study plan that we would like the Panel members to discuss. First, the appropriate control. Although STAR ankle is proposed as alternative to

arthrodesis, the sponsor's post-approval study plan doesn't have a control group.

Also published data that compares the long-term outcomes of the two treatments is lacking based on our literature review.

So you will be asked to comment on the necessity of using arthrodesis or other alternative treatment as an appropriate control in the post-approval study.

Second, the post-approval study protocol, including radiographic assessment at 48, 72, and 96 months post-operation by the treating surgeons, there is, however, no involvement of independent radiologists, no formal radiographic measurements will be obtained.

You will be asked to comment on the adequacy of this radiographic assessment plan and the potential need for radiographic measurements by an independent radiologist.

Third, published data from European study indicated a revision with post-STAR

ankle up to 30 percent during the postoperative period with the median 52 months. Data on the long-term outcome of STAR ankle patients who experienced STAR ankle revision or convert to arthrodesis after device failure are sparse.

You will also be asked to comment on whether the long-term outcome of patients who experience STAR ankle revision or convert to arthrodesis after failure should be addressed in the post-approval study.

Fourth, the sponsor's post-approval study plan proposed to follow STAR ankle patients up to eight years post-operation, however, total ankle arthroplasty is more difficult than total hip and knee arthroplasty due to limitations of the bone strength and the conical sides of the talus as well as the magnified compressive forces distributed across the ankle. All these pose a challenge to achieving long-term success for total ankle arthroplasty.

You will be asked to comment on an appropriate length of follow up to determine the long-term safety and effectiveness of STAR ankle arthroplasty in the post-approval study.

In addition, as described earlier, the follow-up rates for STAR ankle continued access study and arthrodesis control in the pivotal study were low. Given the longer-term follow up and importance of maintaining a high follow-up rate in the post-approval you will be asked to comment on potentially effective measures to minimize loss to follow up and compensatory measures when loss to follow up occurs.

Finally, the sponsor proposed to investigate the physician learning curve by enrolling five new surgeons and 125 new patients. And to follow them up to 12 months post-operation.

You will be asked to comment on adequacy of this study and the best methods to select the new investigators.

And this concludes my presentation as well as FDA's presentation this morning. We would welcome any question that you may have. Thank you.

CHAIR KIRKPATRICK: I'd like to thank all the FDA speaker for their presentations.

Does anyone on the Panel have a clarifying factual question for the FDA with regard to their presentation?

(No response.)

CHAIR KIRKPATRICK: Thank you very much. We will have another opportunity to ask the FDA question during the Panel deliberations later.

We will now begin the Panel discussion portion of the meeting. Although this is open to public observers, public attendees may not participate except at the specific request of the Panel.

This morning, Drs. Pfeffer, Skinner, and Propert will help focus our

deliberations by briefly commenting on the clinical, preclinical, and statistical aspects of this device. Following their comments, the Panel can ask questions of both the sponsor and the FDA that require preparation during the lunch break.

The Panel will resume deliberations following lunch. But first we will have the opportunity to hear from Drs. Pfeffer, Skinner, and Propert. First, Dr. Pfeffer.

DR. PFEFFER: Good morning. I would like to commend the investigators on an excellent job in their study. It was difficult, multi-center, involving hundreds of patients over a long term. You are certainly to be congratulated.

As Dr. Gill noted, these are all leaders in the field of foot and ankle. And we are honored to have them with us.

The FDA reviews are very thorough. Anything that I could say would simply be redundant in that regard. So what I would

like to do is add two views from 30,000 feet about ankle arthritis and this study specifically.

In this study, all patients were evaluated by multiple parameters including the Buechel-Pappas scale which quantifies pain, function, range of motion of the ankle, and deformity. As Dr. Coughlin and I discussed briefly early, subtalar motion was not measured but can be inferred by the total motion that was present in the hindfoot.

Subtalar motion, so we are on a level playing field here, occurs in the joint below the ankle. The normalcy of this motion has a significant effect on the outcome of a total ankle arthroplasty but especially on an arthrodesis. Patients with excellent subtalar motion will have an excellent result, at least short term, from an arthrodesis in regards to pain and increasing function.

This is because some of the motion in the subtalar joint contributes to

dorsiflexion and plantar flexion. This study compares the results of a STAR total ankle arthroplasty versus and ankle fusion. Now in order to make this comparison valid and helpful to this Panel, these two groups should be as similar as possible preoperatively especially in regards to motion in the subtalar joint or lack of motion in the subtalar joint, which I have commented so significantly effects the results and grading of an arthrodesis.

The preoperative groups are very similar, as Dr. Coughlin mentioned, in regards to gender, height, and body mass index. They are, however, very dissimilar in other key ways that may impact the study as a whole. If we look at the pivotal group numbers, the STAR group had 48.1 percent of patients with post-traumatic arthritis while the arthrodesis group had much more, 65 percent.

In other words, this is arthritis that develops as a result of trauma as opposed

to the hip and the knee where most arthritis develops primarily as osteoarthritis without trauma. A vast number of patients in the ankle develop their arthritis because of trauma from a previous ankle fracture.

If we look at the number of surgeries these patients have had, in the arthrodesis group, two or more patients -- 33 percent of the patients had two or more surgeries compared to the STAR group which had 29 percent.

Now most importantly, if you look at the combined motion in these two groups, if we look at those that had a severe loss of hindfoot motion, or what's called combined motion which is probably, as Dr. Coughlin noted, a combination of both ankle and subtalar motion.

We can only infer that from this data. But if you look at these two groups, 53 percent of the arthrodesis group had less than 14 degrees of motion while only 27.5 percent

of the STAR group did. This data warrants further examination.

Dr. Coughlin noted that there was no learning curve with the fusion group, which is true. Any one of us have done many fusions before but interestingly if you look at the data, the surgical time for the fusion group was identical to the total ankle group which was surprising to me.

And I can only infer from that with a 2.2 hour time for both approximately that these were very difficult ankle fusions because almost any orthopedic surgeon who does both total ankles and fusions would have a much less time in the fusion group. So again, this data goes to show the inferiority of the arthrodesis group.

These findings probably account for the lower than expected outcomes in the arthrodesis group. The lower than expected improvement and the pain VAS and the Buechel-Pappas scores, the arthrodesis group, for

example, may be because of the subtalar arthritis and dysfunction which persists after the ankle fusion is performed.

Dr. Coughlin said he was surprised that these patients did not do better. And that's because the subtalar dysfunction was not well understood. And the arthritis that was invariably present was causing these patients to not do well at least according to the data that we have available to us.

Further clinical issues which may effect interpretation of this data and a fair comparison of the groups is the disparity between the pivot STAR patients and the continued access STAR patients. So we're just looking now at patients who have received a total ankle. We are looking at the pivotal and those that are placed into the continued access group.

In regards to preoperative deformity, the pivotal group had 41.8 percent incidence of moderate or significant

deformity, 41.8 percent, while the continued access group had only 12.2 percent of significant or moderate deformity preoperatively.

Many of the improved successes in the continued access group attributed in our information to refined technique, improved equipment, and improved surgical ability may, in fact, actually be related predominantly to the selection of patients over time with less severe and more easily corrected arthritis and deformity.

On page 77 of our book that was handed to us, it is noted that, I think these are the company's words, "As baseline deformity may impact the surgeon's ability to correctly align the ankle at the time of surgery and ultimate patient outcome, the lower degree of deformity seen with continued access patients may explain the improved outcome of these patients as compared to STAR patients in the pivotal study. It should be

noted that prior to the initiation of the continued access study, investigators were cautioned that patients with a coronal deformity could be expected to do less well." This is the one time that this is mentioned.

These findings warrant further examination as they may effect the final FDA recommendations regarding appropriate patient selection.

Thank you.

CHAIR KIRKPATRICK: Thank you, Dr. Pfeffer, for your overview of clinical issues.

We will now here from Dr. Harry Skinner with regard to some of the preclinical data.

DR. SKINNER: Well, first of all, the Panel wasn't given the original preclinical data for evaluation. There were basically summaries of the preclinical data. And I'd like to point out that the purpose of the preclinical studies would be expected to point to potential problems that should be

addressed prior to clinical studies.

First of all, regarding the wear data, I have to agree with the FDA that the ankle was not wear tested for worst case scenarios. And this is particularly worrisome since early wear testing on hip simulators do not provide results that are comparable to wear seen in the clinical setting. And going by that, five tests of wear on ankle that may not be as rigorous as might be seen in the actual clinical situation would be somewhat concerning.

Worst case scenarios would include the maximum allowable weight for the patients found in the contraindications. Other worst case issues that might be addressed would be the placement of the components inside varus or valgus.

This is particularly the case when polyethylene fracture was quoted in the Panel information as possibly due to failure to properly align and stabilize the ankle during

initial STAR ankle placement surgery. That's concerning because the wear would be potentially higher if there was malalignment.

The second concern with wear would be the extremely thin poly in the keel trough and that would be a concern from a failure viewpoint also. I didn't see any information on where the fractures of the polyethylene component had occurred although there were reported cases of failure. High contact stresses would be expected to be present in the area of the keel trough.

That sort of leads into the pressure sensitive film data, the data borne out by the pressure sensitive film and the FEA were pretty similar and it is pretty much as you would expect if you take the stress that was used, 3,650 Newtons and distributed it over 344 square millimeters, you get about 10 or 11 megapascals.

But you would expect significantly higher stresses at the keel trough and

possibly at the edges. So the data is not surprising. What is surprising is that it would be useful and fairly easy to consider other scenarios in FE data since the FE data have been already calibrated with the contact stress data to consider things like the varus and valgus alignment.

The values of stress that were obtained in the FE analysis are getting close to levels, especially for the thinnest poly, and the situation with the heaviest potential patient of being in the area of plastic deformation, that would lead me towards thinking that the company should be thinking toward eliminating smaller polies and considering other options such as perhaps eliminating the wires from the polyethylene component which failed in the wear test.

And it might be stress concentration areas for failure for the polyethylene component. The internal stress levels of the poly are likely to be high

enough to lead to fatigue fracture of the poly, especially in the heavier patients.

Similarly, malalignment, severe enough to require osteotomy, occurred in two percent of the pivotal study, suggesting a higher rate of less severe malalignment, which could potentially contribute to wear and fracture. And that points to the need for particularly meticulous surgical technique.

I feel that further investigation could elucidate problems and lead to improved design of the component. Such things as further wear studies of the thinnest, smallest poly, the heaviest patient loading conditions, and slight varus and valgus malalignment, although some of these could be evaluated with FE studies to evaluate both the stress areas and the areas of observed with clinical fracture and areas of observed clinical maximum wear.

CHAIR KIRKPATRICK: Thank you, Dr. Skinner.

Dr. Propert?

DR. PROPERT: I am becoming accustomed to nonrandomized studies in this setting but when we do a nonrandomized study, we still want to mimic what is happening in a randomized study, namely our control group should differ from our study group first all in terms of treatment but secondly that all the other differences are based on things we can measures and, therefore, control.

And so my first comment about this is that the baseline differences between the two groups that we saw made me worry that there were unmeasured covariates there that were causing differences between the groups that we can't sort from the differences from the treatments.

And fancy statistical analysis such as the propensity scores, which is laudable, will never fix things that we don't measure. So that is my first statistical comment.

My second one is one that has been

mentioned a number of times but it is so important I want to say it again. It was outlined well in the FDA presentation which is this issue that the clinical sites and the treatments are completely alias, which is the statistical term meaning we can't sort out the treatment effects from the site effects.

And one question I have about this which I should have asked earlier this morning, and someone can answer it after the break, is I was unclear on exactly how which sites did which treatments was chosen. And if someone could clarify that for me after lunch, that would helpful.

There is conflicting data on whether there is site to site variability that actually matters here. We should note that there is not a lot of statistical power to assess that so statistical tests for that are not going to give us a lot of information here because you are essentially comparing ten things to five things.

My next set of comments has to do with the efficacy endpoint which I was fascinated by. I appreciate the review of the subscores this morning that was done by the sponsor that there are a lot of numbers here and that really helped me see what was going on.

I was a little confused by their presentation. Not the scores themselves but that there seems to be a lot of jumping back and forth between whether this is a superiority study or a non-inferiority study. And some consistency on that would have been helpful.

Furthermore, the study for the efficacy endpoints, the endpoints based on the BP score, is overpowered, as was mentioned this morning, because the sample size was based on safety. So I would expect to see a lot of significant p-values. And I hope my clinical colleagues on the Panel can tell me if some of these differences are, in fact,

clinically significant.

I do see the value in evaluating the BP without the range of movement score. I have no sense for whether patients really about range of movement. And I would suspect that a lot of what they really care about is contained in the functional aspect of that score which did, indeed, seem to favor the STAR group.

I was really struck by the tertiary efficacy endpoint of patient satisfaction, I think is what it was called, at 24 months, which did appear equivocal. And to me, it gives me information on the overall risk benefit of this.

One sort of subtle statistical issue is it was mentioned in a couple of presentations and also in my documents here that the fact that the STAR group had more severe baseline symptoms or severity of disease would tend to favor the control group. And I actually don't agree with this.

I think the potential for regression to the mean means that I might expect actually larger changes in the more severe group so I think you might be able to explain some of the differences, although they would be small, by the differences in baseline severity.

Finally, I want to talk about the safety endpoint and the statistical aspects there. That is one of the major concerns that has been raised this morning. I mean endpoints change. And it is unfortunate that they change but it's reality.

But when they change, we have to make sure that we change them uniformly across the board. And I just want to echo what has been said a couple of times this morning about being unclear that all subjects were reassessed for all endpoints after changes were made.

I do support using complete data if you have it. If you have 24-month data on how

a patient is doing, there is no reason to use their 12-month data. You may get some anecdotal information on what has happened, why subjects were slow to respond, but I do support using that 24-month data.

I was more concerned with the revised radiographic endpoint because it did appear to me that it was really clinical. And if I understood correctly, that it was based somewhat on 48-month data, then there is some sort of a mathematical no-no going on. You cannot take 48-month data and use to predict how someone was at month 24.

So I am a little concerned that there was some back calculation being done in response, if I correctly understood that 48-month data was used to predict 24-month response. And if I misunderstood that, I apologize.

My final comment is another statistical one. It does appear that the primary safety endpoint after all of the

changes is just reaching its non-inferiority margin. But there are two things that haven't been addressed. One is the multiplicity of endpoints and a lot of the slides this morning said not adjusted for multiplicity.

And the other is something I really didn't completely appreciate until the FDA presentation which is the issue of an interim analysis. That this is not complete data on everybody not because of final loss to follow up but because everybody hasn't reached their endpoint.

And either an adjustment for multiplicity or an adjustment for interim analysis through groups sequential methods or whatever would tend to widen confidence intervals and I think might very much change the conclusions on the primary safety endpoint. And so I'd like to hear discussion about that after the break as well.

Thank you.

CHAIR KIRKPATRICK: Thank you, Dr.

Propert.

I'd first like to see if any of our Panel members have questions for our lead Panel reviewers which would be for Drs. Pfeffer, Skinner, or Propert. Any Panel members have questions for any of those three?

(No response.)

CHAIR KIRKPATRICK: Thank you.

I'd like go around the table now then and see if any of our Panel members have specific questions for either the FDA or the sponsor. And these certainly can be things that they can prepare and answer during our break.

Why don't we start with Ms. Adams.

MS. ADAMS: My question is actually for you because I'm just -- point of clarification for the sponsor -- we have already asked them a lot of questions. When are those going to be answered.

CHAIR KIRKPATRICK: We are going to give them an ample opportunity after lunch.

And that's why, you know, we've already pointed out a few questions. We want to make sure whether there are any other questions or if you have comments about what you've heard or seen this morning.

MS. ADAMS: Okay. I have one that I'd like to add. We heard some talk earlier, I think it was the FDA, about the meta-analysis and the potential for bias in the selection of the articles.

I'd just like to add that to the list of questions for the sponsor because I don't think there is a complete discussion of that in the Panel pack. And maybe they can help us understand how they chose what to include and what to exclude.

CHAIR KIRKPATRICK: Is that a clear question for the sponsor? Do they understand the question? Thank you.

I will ask each Panel member to keep track of their questions so we can make sure that they are answered on an individual

basis rather than one of us taking a list right now, okay?

Ms. Whittington, do you have a question?

MS. WHITTINGTON: No additional questions.

CHAIR KIRKPATRICK: Thank you.

Dr. Wright, do you have a question or comment?

DR. WRIGHT: No.

CHAIR KIRKPATRICK: Thank you.

Dr. Goodman?

MEMBER GOODMAN: I have a series of questions. And please bear with me as I go through these because some are going to be repetitive from what we've just heard.

CHAIR KIRKPATRICK: Dr. Goodman, if I may --

MEMBER GOODMAN: Yes.

CHAIR KIRKPATRICK: -- if you have brief questions that they have answers to, can we go through them one at a time? And they

can acknowledge whether they have the answer available?

And so if we have the answer, we can get that off the list for the lunchtime preparation. And certainly if they are repetitive questions, just reemphasize your agreement with the previous questions. Thanks.

MEMBER GOODMAN: Understood. Thank you.

The first is one of the points that was brought up previously and that is the judgment by the surgeon with regards to osteopenia or osteoporosis, I would appreciate a little more detail into this topic because it is a contraindication that the authors have listed on page 7.

And going along with that is the obesity question. In the future study that is planned, what exact variable -- is it going to be weight? Is it going to be body mass index? In the future study, what are the parameters

that the investigators are going to ask their co-investigators to look at?

I know that Dr. Skinner and others have brought this point up but the factor about the wear analysis, the studies were done on a 163-pound individual or ten million cycles. And as we have just heard, the upper limit is 250 pounds. And people will cycle for more than ten years. So I was wondering if the investigators are planning any future studies with this regard.

CHAIR KIRKPATRICK: We will recognize an answer to that specific question if you have a specific answer to the question do you have worst case scenario plans at this time.

DR. SALTZMAN: No. And I'd like to answer the previous question just to take it on now.

CHAIR KIRKPATRICK: Yes, if we can answer in total the osteoporosis question at this time, is that what you are suggesting?

DR. SALTZMAN: I'll try to answer it. I don't know if it is in total.

The original guidelines written for the study were written by the medical monitors who are sitting behind me in 1999. At that time, there is no data about DEXA scans in total ankles. And there still remains no data.

The concept was if you saw a patient who had osteopenia on an x-ray, that you would consider whether this patient could be osteoporotic, get a DEXA of the hip, and if it was lower than 2.5 SDs, then you would probably not allow them to have this surgery. That was a soft call. We haven't developed hard guidelines regarding that.

PMI versus weight is also up in the air. If you look at fractures, which we will get to later, weight doesn't correlate very well to the fractures of the four patients who had poly fractures in our 600-and some patients. So that is still up in the air.

CHAIR KIRKPATRICK: Thank you.

Does that adequately address your question, Dr. Goodman?

MEMBER GOODMAN: Yes.

CHAIR KIRKPATRICK: Thanks.

MEMBER GOODMAN: Thank you.

CHAIR KIRKPATRICK: Okay. Now would you like to proceed with additional questions?

MEMBER GOODMAN: Yes, thank you.

I wanted to make a comment and perhaps change that into a question. We've all talked about the radiographic failures at six and 12 months that were later classified as successes.

I think one has to be very careful and separate out these two factors, that is looking at x-rays and how patients are doing, it is well known that patients who have a cemented total hip replacement can have a loose acetabular component radiographically.

You look at the x-rays. The

radiographic lucent lines are very easy to see. The cup has migrated. And yet a lot of these patients clinically do well.

And I'm a little reluctant, I think, and I think that the investigators should explain and justify the mixing of the two terms radiographic success and clinical success.

It seemed to me that they were using clinical success as a marker irrespective of radiographic failure at six or 12 months as a surrogate for success at a later time because the patient had not undergone revision. Is that clear?

That can be explained later or you can explain that now.

CHAIR KIRKPATRICK: I would suggest if it is more than just a few sentences, we would let you prepare a more thorough answer after lunch. But if you can handle it in a few comments, then please go ahead.

DR. SALTZMAN: It is more than a

few sentences.

CHAIR KIRKPATRICK: Okay, we'll look forward to hearing that after lunch. Thank you.

MEMBER GOODMAN: I thought so.

The other point is I didn't see it in the manual that we were given but was there resident and fellow participation in the surgeries? And, for example, for the surgeries which were fusions, if there were residents or fellows that participated in those operations but not in the STAR operations or vice versa, this might have an effect on outcome. And I was hoping that the investigators could clarify that point.

DR. MANN: The surgery was all done by the principle investigators. Most of us do have fellows. They would participate in it but at no time was the case handed over as a fellow case or a resident case.

MEMBER GOODMAN: For both types of surgeries?

DR. MANN: Yes.

MEMBER GOODMAN: Thank you.

The other two points that I wanted to raise our statistical significance versus clinical significance. Now I have taken three courses in statistics in the past and I think I got honors in all three.

But when we read the orthopedic literature, we commonly see p-values of .01 or .05 or something in that realm. And then when we look at the raw values, the difference in blood loss might be 10 ccs or 15 ccs. And I think when we read this manual in the presentations here, I think we have to keep in mind the difference between something that may be statistically significant and something that is clinically significant.

In this type of operation or in total hip or total knee replacement, blood losses which are minor might be statistically significant but they are of no clinical significance. And perhaps the investigators

when they report their data may want to emphasize that point. Probably later on I would think. Is that all right? Okay.

And finally, there has been a lot of discussion about range of motion. And sort of a re-manipulation of the data to exclude range of motion because obviously when you do a fusion, you want to get rid of range of motion. Range of motion is a failure basically after an arthrodesis whereas as far as I understand this operation, and part of the improvement in function and in mobility, et cetera, is some sort of preserved range of motion.

And perhaps it is a rhetorical question but I personally feel that it is not right to do that analysis without including range of motion because that is part of the procedure. And it sort of penalizes the investigators when we recalculate everything without including a variable that is basically germane to the whole argument of this

prosthesis. And that is preservation of range of motion as well as improving function.

And perhaps the investigators may want to emphasize this. We'll leave it up to them.

CHAIR KIRKPATRICK: Thank, Dr. Goodman.

MEMBER GOODMAN: That's it.

CHAIR KIRKPATRICK: We'll now go to Dr. Mayor and make sure he doesn't -- or see if he has any question or comment.

DR. MAYOR: Just a point of clarification. We can address the FDA presenters as well?

CHAIR KIRKPATRICK: Yes. Absolutely.

DR. MAYOR: Because I do have a couple of questions that I would like to address to two of the FDA presenters. The first one would be Dr. Wang who called for a hypothesis-driven study as an ideal or more appropriate goal to shoot for.

And I wonder if he would be willing or able to elaborate a little bit on what a hypothesis-driven study presents or provides over and above the strictly observational study might have been.

This is a point of educating me on the benefits of creating hypotheses to drive a study.

DR. WANG: And first of all, if we have the hypothesis-driven study, we can calculate a sample size and during the study design we can understand whether we have enough power to detect any significance that we target for.

Without hypothesis driven, then your sample size calculation is really subjective if I can use that word. There is no way you can make a rationale for the sample size.

And the second, I would say, with a hypothesis driven, you have a clear goal before you conduct that study, before you

collect data, before you do the analysis rather than after you look at the data, after you already collect data, after that you can do the post hoc analysis. Then you can make the statement, right? Is that -- you agree with that?

So I can bring up two points here regarding hypothesis driven. Maybe a sponsor can help with another perspective. But I just wonder whether you agree with those two points.

DR. MAYOR: Yes. And I wonder if just to further clarify my own concepts of the issue, whether you might be willing to propose at least one hypothesis --

DR. WANG: Yes.

DR. MAYOR: -- for this PAS that might be worth setting up.

DR. WANG: Yes. I think for the major -- I think for the PAS study, we have to clarify what is the major primary outcome -- what is the primary question you want to look

at.

For the primary question, if I can only set up one hypothesis driven, I would like to set a hypothesis driven for the primary question to look at the PAS study. Does that answer your question?

DR. MAYOR: Yes, thank you.

CHAIR KIRKPATRICK: If I may just supplement that. My own understanding, can you give me an example of a specific hypothesis that you would apply to this post-approval study?

DR. WANG: If our primary concern for the post-approval study is look at -- just using the sponsor's primary outcome, look at the revision rate for the STAR ankle. And by using hypothesis driven, we are basically looking at whether the rate of a revision or removal among STAR ankle patients is non-inferior to a part of the control group. That would be a good hypothesis-driven question to ask in the post-approval study.

So also another question we can think about is even among the STAR ankle patients, for example, you are interested in looking at patients who are younger than 50 with post-traumatic arthritis, whether they are doing the same as other patients, then we can also state a hypothesis.

CHAIR KIRKPATRICK: You'll have to forgive me. I'm from the South. So I need it a little bit more concrete. And I didn't have as much honors participation in statistics.

My understanding is the hypothesis needs to be a question.

DR. WANG: Yes.

CHAIR KIRKPATRICK: So would an example of that be that at five years post-approval, the revision rate is no different or less than what it was in the first two years of the pivotal study.

DR. WANG: Yes. Right.

CHAIR KIRKPATRICK: Is that an example of a hypothesis?

DR. WANG: That is an example of hypothesis driven, right.

CHAIR KIRKPATRICK: Okay. Thank you.

DR. WANG: That's right. But again, we will gather some comments from the sponsor whether five year, eight, or ten year, which way is it going to be and what is the rationale for that.

CHAIR KIRKPATRICK: I understand that. I just picked that number out of the air. So, thank you.

DR. WANG: Yes, thank you.

CHAIR KIRKPATRICK: Now, Dr. Mayor, you had more comments?

DR. MAYOR: I have one more question which is very simple and just represents my own lack of insight, addressed to Mr. Zhou, his Slide 11 included a column in which the term LOCF was included. I don't know what LOCF is.

CHAIR KIRKPATRICK: Mr. Zhou?

MR. ZHOU: Sorry, that LOCF represents last observation carried forward. So it is a technique, it is a very common technique used for imputing missing data although it tends to de-emphasize the lack of uncertainty. It is one way to impute the missing data.

DR. MAYOR: Just one final plea to a general audience and that is I'm constantly disturbed by the assertion that there was no statistical significance in the analysis of the results. I sort of, in my own mind, reword that assertion to say that the study failed to achieve statistical significance given the amount of data available.

It may seem like an inconsequential distinction but I think it is very real. And we had an example in our own area. I am in a hotbed of outcome studies and our residents and many of the staff members plumb the Medicare database for information.

When you deal with a variable and