A protocol revision in May 2005 changed the interim analysis to 75 percent of the data, and a sample size was slightly increased to 394 with a goal to obtain 334 evaluable subjects. Due to the interim analysis, the significance level or the maximum high-point error rate allowed for the final analysis was reduced to 0.044. At the end of the study, 352 subjects were actually enrolled, and the 334 subjects were evaluable at 6 months.

The next slide shows some of the key changes in the planned statistical analyses on the primary effectiveness endpoint. In May 2002, the study protocol was conditionally approved by FDA. The statistical methods section of the approved protocol specified that the initial GEE model would contain treatment, time, baseline pain score level, and baseline level by treatment interaction term. If the interaction term was not statistically significant, it would be removed from the model and baseline level would remain as a continuous covariate.

The study enrollment began in August 2002, and the interim analysis was conducted in April 2006. In December 2006, the revised statistical analysis plan stated that all clinically relevant baseline

factors would be screened in the primary endpoint analysis and interactions with treatment would be studied.

This slide shows the subject dispositions and the different subject populations. 177 subjects were randomized to the Oxiplex group and 175 subjects were randomized to the control group. All but one of these randomized subjects were included in the intend-to-treat, or ITT population. Subjects withdrawals and lost to follow-up represented less than 5 percent of the total enrolled subjects. Four Oxiplex and one control subjects had 6-month visit far beyond the visit window and were excluded from the completed cases population by the Sponsor.

In your Panel pack, some analyses were conducted on the FDA's modified completed cases population, which included these five far-out-of-window subjects. However, since the results were similar, we will present most analyses on the completed cases, also identified as PMA CC population here, to be consistent with the Sponsor. However, even though the Sponsor referred to these 167 Oxiplex subjects and 167 control subjects as the completed cases population in the original PMA submission, recently, they appear to call the 6 months in- window

population as the completed cases, or CC population. To avoid confusion to the Panel, we will use PMA CC as a name for these 167 Oxiplex subjects and 167 controls. But please be aware this PMA CC population may be different from the Sponsor's CC population presented earlier.

Analyses on ITT population also gave similar results. However, since the Sponsor used linear interpolation for out-of-window visits for the ITT population and the Sponsor's single-imputation model was not pre-specified, we will not focus on the ITT population in this presentation although we can provide results on the ITT population if the Panel is interested.

This slide shows the demographics on baseline characteristics of Oxiplex subjects and control subjects. The two arms are comparable in age, gender, BMI, baseline leg pain and back pain, indicating good randomization was achieved in the study.

Next, I will talk about the primary effectiveness endpoint. Again, the primary effectiveness endpoint in the pivotal study was improvement in leg pain from baseline at one, three, and 6 months post-surgery. The leg pain was measured

by the Lumbar Spine Outcomes Questionnaire and converted to 0 to 100 scale. The repeatedly measured data were analyzed with generalized estimating equation in the ITT and completed cases population. Again, here, we will focus on the completed cases population although the results on the ITT population are available upon the Panel's request.

I will first talk about the overall treatment effect, then the Sponsor's exploratory subgroup analyses and the issue of site variability.

Before getting into the details of the results, I want to talk a little about the model selection process. First, why do we need statistical models? One main reason we use statistical models in clinical trials is to adjust for potential covariate imbalance between the treatment and the control arm.

There are several ways to select covariates to be included in the statistical model in a clinical trial setting. The best way is to pre-specify all the covariates to be included in the model at the IDE stage. Automated covariate selections based on software or a combination of pre-specified covariates and automated selection are also used sometimes. Usually, the treatment by covariate interactions are pre-specified, clinically plausible, and added at the

last step after all the main effects covariates have been added to the model.

The Sponsor's model selection process was quite unusual. Instead of using the pre-specified covariates from the IDE protocol, the Sponsor screened a large number of treatment-by-covariate interaction terms very early in the process. Then, a manual backward selection process was used to select the covariates and interactions to be included in the model. This manual process was prone to biases and could not be replicated easily. The unusual model selection process produced a complex GEE model that was difficult to interpret clinically or statistically.

Here's the Sponsor's GEE model in the PMA CC population. The model included 10 covariates, site, baseline leg pain, baseline back pain, baseline functional scores, CPT, pulmonary abnormality, three neurosensory exam results, and sexual function. Five treatment-by-covariate two-way interaction were also included. Due to the presence of treatment-by-covariate interactions in the model, the overall treatment effects was difficult to characterize because the treatment effects was considered to be different for different covariate combinations.

The P-value for the main treatment effect should not be use to interpret the overall treatment effect when treatment by covariate interaction is present in the model because that P-value only corresponds to the treatment effect in a very specific subgroup.

Later, we will ask the Panel to comment on the Sponsor's GEE model.

To illustrate what interactions really mean, this slide shows a hypothetical treatment-by-gender interaction. Suppose an investigational device was applied to the treatment group and a standard device was applied to the control group and the result was stratified by gender.

The graph A, on the left, shows that the treatment effect, represented by the difference between the treatment group and the control group, was positive and consistent for males and for females, indicating positive overall treatment effect and no treatment-by-gender interaction.

Graph B, in the middle, shows that the treatment effect was positive for both males and females, but males appear to benefit more from the device than females. In this case, there may exist what we call a quantitative treatment-by-gender

interaction, but the overall treatment effect is still positive when males and females are combined.

Graph C, on the right, shows the treatment did better than control for males but the control did better than the treatment for females. In this case, there may exist a qualitative treatment-by-gender interaction, and the overall treatment effect may be close to zero when males and females are combined.

Therefore, putting treatment-by-covariate interactions into the model allows the treatment effect to be different for different subgroups of subjects. It is usually the first step towards subgroup analyses. However, these type of subgroup analyses should be considered exploratory if they are not pre-specified, which I will talk more about later.

In general, including treatment-by-covariate interactions in statistical models makes the treatment -- makes the model more complex and the overall treatment effect difficult to characterize.

In order to evaluate the overall treatment effect, FDA develops two GEE models without any treatment-by-covariate interactions. Model 1 was derived from the Sponsor's GEE model in the PMA CC population after removing the five treatment-by-

covariate two-way interactions. Model 2 included the only two covariates specified in the original IDE, visit and baseline back pain.

Now, here are the results from these two models. The adjusted overall treatment effects was 1.8 in Model 1 and 0.1 in Model 2. Both 95 percent confidence intervals included zero, indicating the treatment effects was not statistically significant in either model.

Even though the pre-specified primary analysis was GEE model-based, it always helps to look at the unadjusted analysis, especially for studies like this with balanced covariates between arms.

This table shows the simple averages of leg pain improvement from baseline for Oxiplex group and the control group. At one, three, or 6 months after surgery, both Oxiplex and control group had average leg pain improvement of around 50 points on a 0 to 100 scale. The treatment effect, which is the difference between the average Oxiplex group improvement and the average control group improvement, range from -.1 at month one to 0.9 at month six. The treatment effects was not statistically significant at any of the three time points.

The Sponsor also conducted post-hoc subgroup analysis on the primary effectiveness endpoint. As I mentioned earlier, testing the treatment-by-covariate interaction is usually one of the first steps to subgroup analysis. Listed here are 48 different treatment-by-covariate interactions screened by the Sponsor. Please note, treatment-by-baseline back pain was one of them.

Screening 48 different treatment-by-covariate interactions was analogous to conducting subgroup analysis in 48 different ways. This practice dramatically increased the chance of finding at least one subgroup with a favorable treatment effect. Please note, even though the primary endpoint was leg pain improvement, the subgroup the Sponsor focused on was associated with baseline back pain.

On the next slide, the results on the primary endpoint was stratified by baseline back pain in a way similar to the Sponsor's exploratory subgroup analyses. It appears that for subjects with baseline back pain less than 63, the control subjects had better leg pain improvement, indicate a negative treatment effect. But for subjects with baseline back pain greater than or equal to 63, the Oxiplex

subjects appear to have better leg pain improvement, although none of the treatment effect was statistically significant at any of the three time points.

Please note, the estimated treatment effect and P-value for the last row, which is month six results for subjects with baseline back pain greater than or equal to 63 may be different from what the Sponsor presented earlier. Again, this is because this table here used the PMA CC population, which included 166 -- I'm sorry -- 167 Oxiplex subjects and 167 controls at 6 months. The Sponsor might have used the 6-month in-window population, which included 145 Oxiplex subjects and 141 controls.

Please note, later on, we will ask the Panel to comment on the Sponsor's subgroup analyses.

One treatment-by-covariate interaction the Sponsor screened but did not include in the final model was treatment-by-site interaction, which evaluates whether a treatment effect is consistent across sites. Since site variability is usually of great interest to FDA, we developed two GEE models similar to the ones we used in the primary endpoint analysis except that we added the treatment-by-site interaction term in the model. In both models, the

P-value for the treatment-by-site interaction was significant, indicating site variability existed in this study.

To illustrate what this means, this graph shows the treatment effect of the primary effectiveness endpoint at 6 months by site. Here, the treatment effect is defined as the difference in leg pain improvement between Oxiplex and the control group at 6 months after surgery. A positive treatment effect indicates advantage of the Oxiplex group.

The yellow bars represent standard deviations. The numbers along the horizontal axis represent the site number. Some small sites were grouped together in the same way the Sponsor did. Out of the 19 sites or grouped sites, 10 of them appear to show a negative treatment effect, shown in red, on the left side, and 9 of them appear to show a positive treatment effect, shown in black, on the right side.

In other words, in the 10 red sites or group sites, the control group appear to have better leg pain improvement at 6 months after surgery, but in the 9 black sites, Oxiplex group appear to do better. Although by itself, none of the sites or

group sites can claim statistical significance on the treatment effect being different from zero, there does appear to be a lot of variability between sites.

Later, FDA will ask the Panel to comment on the issue of site variability.

Now, let's move on to secondary endpoints. In December 2006, the Sponsor proposed hierarchical closed testing procedure to adjust for multiple comparisons for secondary endpoints was conditionally approved. This procedure involved sequential testing of the primary and secondary endpoints until one endpoint fails the test, which is illustrated on the next slide.

Please note, this is FDA's understanding of a hierarchical closed testing procedure, which the Sponsor may not agree. First, the primary endpoint is tested. If the primary endpoint is statistically significant, then the secondary endpoints are sequentially tested at the same significance level. If the primary endpoint or any of the secondary endpoints fails the test, the procedure stops and the rest of the secondary endpoints are not tested. In the pivotal study, the primary endpoint, leg pain reduction, failed to achieve statistical significance. According to this hierarchical closed

testing procedure, the process should stop and no secondary endpoints should be tested.

However, the Sponsor still conducted secondary endpoint analysis similar to their primary endpoint analysis, which went through the same unusual model selection process that screened the same 48 different treatment-by-covariate interactions and produced similarly complex GEE models with multiple two-way treatment-by-covariate interactions. It was very difficult to characterize the overall treatment effects of the secondary endpoints from these complex models.

In addition, the Sponsor conducted post-hoc exploratory subgroup analyses similar to those conducted for the primary endpoint and presented results stratified by baseline back pain again.

To reiterate, even though there was no need to analyze the secondary endpoints when the primary endpoint failed to achieve statistical significance, in order to give the Panel a more complete picture of the study, we're showing the results of secondary endpoints here.

This graph shows the point estimates and the corresponding 95 percent confidence intervals for the unadjusted treatment effects of the primary

endpoint and all secondary endpoints at 6 months. All confidence intervals include zero, indicating no primary or secondary endpoint was significant at 6 months without implementing the hierarchical closed testing procedure. The 1 month and 3 months results were similar.

In summary, based on FDA's analyses on the PMA CC population, the overall treatment effect for the primary effectiveness endpoint was not statistically significant. The Sponsor's post-hoc subgroup analysis should be considered exploratory. In addition, site variability may exist in this study. Finally, ignoring the hierarchical closed testing procedure, none of the secondary endpoints were statistically significant at 6 months in the PMA CC population.

The results on the ITT population would reach the same conclusions, and they are available upon Panel's request.

This concludes my presentation. Next, Dr. Jiping Chen will talk about post-approval study.

DR. CHEN: Thanks, Jack. Good morning distinguished members of the Panel and members of the audience. My name is Jiping Chen, and I'm one of the epidemiologists in the Division of Post-market

Surveillance in the Office of Surveillance and Biometrics.

As the epidemiologist in the PMA review team, I'm responsible for working with the sponsor for the development of a 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 Sponsor can agree on.

Here is outline of my presentation today. First, I will discuss the general principles that were utilized when thinking about the need for and designing post-approval studies. Then I will comment on the rationales for the post-market questions that the pre-market study was not designed to answer but may be addressed in the post-approval study. Then I will summarize the latest version of the Sponsor's PAS outline and our assessment of the PAS outline. Finally, I will describe the PAS issues that we would like the Panel to discuss on the design of the post-approval study if the PMA is approved.

Before we talk about post-approval studies, we need to clarify a few things. The discussion of a

post-approval study prior to a formal recommendation on approvability of this PMA should not be interpreted to mean FDA is suggesting the Panel find the device approvable. The plan to conduct a PAS does not decrease the threshold of evidence required to find the device approvable. The pre-market 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 studies. The main objective of conducting post-approval studies is to evaluate device performance and the potential device-related problems in a broader patient population over an extended period of time after pre-market establishment of reasonable evidence of device safety and effectiveness. Post-approval studies should not be used to evaluate unresolved issues from the pre-market phase that are important to the initial establishment of device safety and effectiveness.

The reasons for conducting post-approval studies are to gather post-market information, including long-term performance of the device, data on how the device performs in the real world, in a

broader patient population that is treated by community-based physicians, as opposed to highly selected patients treated by investigators in clinical trials; evaluation of the effectiveness of training programs for use of devices; evaluation of device performance in subgroups of patients, since clinical trials to have limited numbers of patients or no patients at all in certain vulnerable subgroups 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 clinical trials. And, finally, we conduct post-approval studies to address issues and concerns that Panel members may raise based on their experiences and observations.

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

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

The second question is: What is the long-term safety and effectiveness of the device post-

market?

This table presents an overview of the Sponsor's latest PAS outline. To confirm the safety and reduction in disability days in subjects who receive Oxiplex during first-time lumbar disc surgery, the Sponsor proposed a prospective multicenter cohort study with a non-inferiority design and historical controls.

Study population consists of 210 Oxiplex PAS subjects and 145 historical control subjects, a total of 355 subjects. The Oxiplex PAS group are subjects who will be treated with Oxiplex in the PAS, while the Oxiplex-treated subjects in the pivotal study, who completed 6-month follow-up visits, will serve as the historical control group. The Sponsor proposed to follow the subjects for 6 months after surgery.

The proposed effectiveness endpoint is a mean reduction in disability days that occurs over the last 30 days of the 6-month period after surgery. As the Sponsor described earlier today, the disability days are defined as days when the subjects are completely disabled by their lower back conditions. The hypothesis of the study is that the mean reduction in disability days for the last 30

days of the 6-month period in subjects who will be receiving Oxiplex is no worse than the Oxiplex-treated subjects in the pivotal study by a margin of 2.5 days.

The proposed safety endpoints include the following: Procedure and device-related AEs, number of re-operations, and musculoskeletal and lower extremity neurological functions and will be evaluated with descriptive statistics up to 6 months post-surgery.

The Sponsor will list relevant adverse events by type and overall, with rates and the corresponding 95 percent confidence intervals.

Should the device be approved, we will continue to work with the Sponsor to develop an appropriate post-approval study. We would like to bring to your attention a few issues regarding the Sponsor's post-approval study outline. Here is our initial assessment of the Sponsor's PAS outline.

First, our assessment of the study design. Are we convinced that historical controls are the most appropriate? What about concurrent standard-of-care controls?

The Sponsor proposed in the non-inferiority study design to compare the reduction in disability

days in subjects who will be treated with Oxiplex in the PAS versus Oxiplex-treated subjects, in the pivotal study. Under this design, patient comparability important baseline factors in the two studies is required to ensure that an interpretation of safety and effectiveness data is valid. The potential difference between the PAS population and the pivotal population needs to be accounted for in the analytical phase and in data interpretation.

In addition, it is not clear whether the non-inferiority margin of 2.5 is appropriate. A non-inferiority margin of 2.5 days means that the PAS Oxiplex subjects will retain at least 67 percent of the effect of the pivotal Oxiplex subjects. The Sponsor needs to provide a clinical justification for the selected margin to see whether the margin is small enough to be clinically insignificant.

Furthermore, the Sponsor stated that the margin of 2.5 days is 3.5 times lower than the standard deviation and appears to be a reasonable estimate. The Sponsor needs to explain why they used standard deviation for the calculation. It should be noted that the same standard deviation should not be applied to both groups.

Second, the Sponsor proposed effectiveness

endpoint as a reduction in the disability days in the last 30 days of the 6-month period post-surgery. FDA is uncertain whether this is an appropriate effectiveness endpoint for the following two reasons.

First, the Sponsor did not provide a justification for not using mean changes from baseline in leg pain as a primary endpoint, which was the primary effectiveness endpoint in the pivotal study.

Second, as the statistical reviewer mentioned earlier, it is unusual to test the statistical significance of the difference in the secondary endpoints if the difference in the primary endpoint is not significant.

Third, for safety endpoints, the Sponsor stated that subjects will be followed for procedure and device-related adverse events only. In the pivotal study, there were seven AEs that were either possibly or probably related to Oxiplex as opposed to zero in the control group at 6 months post-surgery. Because the study sample size of the pivotal study is small, there are some questions about the product safety when the product is put into actual conditions of use post-market.

To be most meaningful and interpretable,

all AEs should be documented and those assessed to be procedure or device-related clearly noted and summary frequencies provided. Assessing only procedure or device-related AEs will potentially underestimate the rate of adverse events, which will limit FDA's ability to adequately interpret the device long-term safety profile. FDA believes it would be more appropriate to include all adverse events, not just those assumed to be device or procedure-related in order to detect any potential unexpected association.

Fourth, the duration of follow-up. The Sponsor proposed to follow subjects for only 6 months post-surgery. FDA is uncertain whether the 6 months of the follow-up is long enough to observe adverse events that were associated with the use of Oxiplex based on the literature and the post-market adverse events reports from a device with similar composition.

As I mentioned earlier, there were seven AEs that were either possibly or probably related to Oxiplex as opposed to zero in the control group at 6 months. And from a literature review, it is known that one of the seven AEs, intervertebral disc protrusion, may occur with scars within 12 months after the surgery. Second, an AE report analysis

conducted on another CMC-based adhesion barrier indicate that approximately 8 percent of adverse events occurred beyond 6 months post-surgery. We are seeking your input into the optimal length of follow-up to better address device long-term safety.

Finally, the Sponsor stated that up to 210 subjects will be enrolled at up to 20 clinical sites in consideration of a potential drop-off rate of 25 percent, resulting in the total sample size of 355 patients.

Based on the information provided in outline, the Sponsor is believed to have used the following assumptions in the current sample size calculation: Mean reduction in disability days of 7.67 in pivotal Oxiplex subjects and similar in PAS Oxiplex subjects, a 2.5 days non-inferiority margin, and a one-sited test with alpha of .05 and 145 pivotal Oxiplex subjects.

The Sponsor stated that 156 evaluable post-approval subjects will provide 80 percent power for the non-inferiority test. From our calculation, 154 subjects are needed.

In addition, the 25 percent drop-off rate seems high in view of the relatively short 6-months follow-up that the Sponsor proposes. We will

continue working with the Sponsor to clarify issues, including developing a better plan to minimize loss to follow-up and in specifying any measures that will be taken if the number of subjects falls below 355 during follow-up visits.

Based on the Sponsor's proposed PAS outline and our initial assessment, we will be asking the Panel during your afternoon deliberations to discuss whether the proposed PAS plan is appropriate to address device long-term safety and effectiveness and make recommendations if the Panel recommends device approval with the condition of a post-approval study.

First, the study objective and question. The Sponsor proposed to conduct a PAS to confirm device safety and reduction in disability days in subjects who received Oxiplex during first-time lumbar disc surgery. We would like to ask you to discuss whether this is an appropriate objective to be studied in a PAS. If you think the PAS question that the Sponsor intends to address is not appropriate PAS question, we will ask you to consider what PAS questions need to be studied in a post-approval study to address device long-term safety effectiveness.

Now, if we assume the PAS question proposed

by the Sponsor is a valid PAS question, then for the next couple of slides, we will look into the details of the study design and so on and ask for panel to discuss and make recommendations.

Second, the study design. The Sponsor proposed to conduct a PAS with non-inferiority design to compare the reduction in disability days in PAS Oxiplex-treated patients versus Oxiplex-treated patients in the pivotal study. We will ask you to discuss whether this is an appropriate design to address device long-term safety and effectiveness in the real world. In addition, the Sponsor proposed a non-inferiority margin of 2.5 days. We would like the Panel members to discuss if the margin is clinically relevant and what would be an acceptable difference in a post-approval study.

If the non-inferiority design is not appropriate, we will ask you to consider what study design will be appropriate to address device long-term safety and effectiveness post-market.

As to control selection, the Sponsor plans to use Oxiplex-treated subjects from the pivotal study as historical controls in the PAS. We will ask you for guidance regarding the appropriate control group for the PAS study.

Fourth, the proposed effectiveness endpoint is a reduction in disability days that occur over the last 30 days of the 6-month period after surgery. We will seek your input regarding what effectiveness endpoints should be addressed in the post-approval study.

Finally, the Sponsor proposed a 6-month follow-up. We would like to ask you whether 6 months is appropriate to address device long-term safety and effectiveness post-market. Again, if 6 months is not appropriate, we will ask you a question about the optimal duration of follow-up in the post-approval study.

Before I conclude, let me remind you that the discussion of a post-approval study prior to a formal recommendation on approvability of this PMA should not be interpreted to mean FDA is suggesting the Panel find the device approvable.

This concludes my presentation, as well as FDA presentation this morning. We welcome any question that you may have.

DR. MABREY: I'd like to thank all of the FDA speakers for their presentations.

At this point, before we begin with our presentations from the Panel members, does anyone on

the Panel have a specific question for the FDA while it's on your mind? Keep in mind that we can ask questions of the FDA later on this afternoon during our discussions.

DR. HORLOCKER: I would just like --

DR. MABREY: Dr. Horlocker?

DR. HORLOCKER: -- clarification on the presumed endpoint that the FDA chose. It was a 33 percent reduction from baseline. And so I just want clarification that if you start with the mean reduction from baseline as about 50 points on that 0 to 100-point scale, you were looking for a 33-percent further reduction to denote efficacy of this device, is that correct?

DR. LEE: When calculated the improvement from baseline between two groups, the difference should be 33 percent.

DR. HORLOCKER: So --

UNIDENTIFIED SPEAKER: That's a yes.

DR. LEE: For example --

DR. HORLOCKER: Yes, thank you.

DR. LEE: You start from 100 and it comes down to 70 for control, and that is a 30 percent decrease. And, for example, Oxiplex comes down from 100 to 50, that is a 50. So the difference should be

33.

DR. HORLOCKER: Okay. Thank you.

DR. MABREY: Yes, Dr. Blumenstein?

DR. BLUMENSTEIN: Just a quick question. The FDA did a model in which they had the treatment-by-site interaction, and I'm wondering, did that also include the site main effect?

MR. ZHOU: Yeah, the models we include are site main effect and treatment-by-site interaction, both models.

DR. MABREY: Okay. At this point, we will now begin the Panel discussion portion of the meeting. And although this portion is open to public observers, public attendees may not participate except at the specific request of the Panel.

This morning, Drs. Rao and Evans will help focus our deliberations by briefly commenting on the clinical and statistical aspects of this device. Following their comments, the Panel can ask questions of the Sponsor and FDA that may require preparation during the lunch break. The Panel will resume deliberations following lunch.

Dr. Rao will now give us his remarks. Dr. Rao?

DR. RAO: Good morning. I've been asked to

provide some clinical perspective this morning to help kick off the deliberations of the Panel.

The underlying premise of the device is an unmet need with a 40 percent residual or recurrent pain following lumbar laminotomy and discectomy. This results in the need for additional treatment, medication, and cost.

My review of the literature suggests that this is generally in line with what's been quoted in the literature. A 10-year follow-up study of patients undergoing discectomy, 72 of 131 patients were followed for more than 10 years, and 12.7 percent of these patients had frequent mild or occasional low back pain and 9.5 percent of these patients had occasional severe leg pain at the 10-year follow-up mark.

Revision surgery was necessarily in 12.5 percent of these patients, and, in general, the need for revision occurred from recurrent disc herniation of the same level, recurrent disc herniation at a different level, and in 1 out of the 9 patients from leg pain caused by excessive scarring. It's interesting that in this series, all patients had epidural fat graft placement after the discectomy.

There's another U.S. study looking at 10-

year outcomes on 400 or 477 surviving patients, where they compared operative and non-operative care and, in general, found that 69 percent of the operative group improved compared to 61 percent of the non-operative group. The work and disability status was similar in both groups.

In patients following a laminectomy, as opposed to just a laminotomy, here is another long-term follow-up study, a retrospective study, and 37 of 151 patients were available at the 10-year follow-up mark. Poor results were found in 22 percent of the patients, and the authors attributed the poor results to recurrent disc herniation or disc herniation and also facet joint pain.

There was a symposium carried out of the North American Spine Society meeting in 2003 on what causes poor results after back surgery, what are the causes of failed back surgery. The participants in the symposium felt that in 90 percent of patients, failed back surgery had a structural etiology and foraminal stenosis was the leading culprit in most of these cases. Discogenic pain resulting in back pain was in 20 to 22 percent of patients, neuropathic pain from a battered nerve root from excessive trauma during surgery, recurrent disc herniations,

instability, facet pain, or sacroiliac joint pain were the other causes of residual symptoms following back surgery.

Besides structural causes, inappropriate patient selection and psychological factors were also felt to contribute to a less than optimal outcome following back surgery.

The Sponsors state that the device is intended to coat and protect neural tissue and thereby significantly reduce nerve root-related postoperative pain and related symptoms during lumbar disc surgery. In the absence of a clear basis of efficacy in that pivotal study, I'm going to assume that they presume that this basis of efficacy is the reduction of peridural fibrosis, which they reported in their pilot study.

Peridural fibrosis can occur following laminotomy and discectomy at two sites, either at he laminotomy site, dorsal to the nerve root structures and dural sac, or ventrally, directly over the annulus. The exact cause of this peridural fibrosis is unknown. It's presumed that fibroblasts migrate from the surrounding chromatized tissue and fill in the area. There may be an individual predisposition to greater scar formation, and the extent of surgical

trauma may result in greater scar formation.

There have been many prior animal studies looking at prior -- at peridural fibrosis following fat graft placement, gelatin foam, hyaluronate, silastic sheets, polylactic acid foam, Dacron, and Adcon-L gel. And all of these studies have found reduction in postoperative scarring following all of these interposition membranes in animal studies.

In theory, this peridural fibrosis may result in leg pain, and the leg pain is presumed to be from tethering of the nerve root, or the dorsal root ganglion or from root ischemia. There is no physiologic association between the formation of this peridural scar and back pain or no clear physiologic association.

There have been some studies that have been done that have shown that in spite of peridural fibrosis, patients can have good results. There have also been some other studies done that have shown that there is no difference in the amount of scar formation in patients with or without symptoms.

There have been some clinical studies done with the use of these interposition barriers. McKay and others did a study on 156 patients following single-level lumbar laminectomy/discectomy, and they

found that the clinical outcome did not vary based on whether an interposition membrane was used, whether fat graft was used, or gelatin foam sponge was used.

Jacobsen and others did another study where they found that patients with free fat graft placement after surgery had better outcome than patients with gelatin sponge placement.

There have been a number of studies carried out with Adcon-L gel, most or all of which was supported by the manufacturer and had good results. There are some studies with poor results and some studies that report intraoperative hypotension following the use of Adcon gel. Product was subsequently withdrawn.

In the Sponsor's application, 352 patients were enrolled and 334 completed the questionnaire. Primary and secondary safety variables were looked at and there was no statistically significant difference between the control and study groups. Primary and secondary effectiveness variables were looked at, and using FDA data, there was no statistically significant difference between the control and study groups with a univariate analysis.

The Sponsor focused on a multivariate analysis with the rationale that pain associated with

the lumbar spine is a complex multifactorial phenomenon, and a univariate analysis was not felt to be adequate.

Using this multivariate analysis, they reported that in patients, or in the subgroup that had severe back pain, there was statistically significant improvement in leg pain. And, similarly, in the subgroup that had severe back pain, there was statistically significant improvement in the back pain, as well as patient satisfaction. It was unclear to my review, based on the submitted application, as well as this morning's presentation, whether the improvement in disability days was based on a univariate or multivariate analysis.

As far as the methodology, patients -- some of the inclusion criteria were patients undergoing laminotomy/discectomy for a single-level disc herniation at L4-5 or L5-S1 with radiculopathy. Patients were selected and drafted into the study following two weeks of non-operative treatment or earlier if the pain was impractical.

The exclusion criteria used were if patients had a myelogram or lumbar puncture, if they had foraminal stenosis, and there were a group of intraoperative inclusions, including what was

reported as dural entry and multiple others listed, epidural fat placement, or an intraoperative determination that a hemostatic agent must remain at the surgery site.

Randomization between control and the study group was carried out by the surgeon after hemostasis had been achieved and the surgeon was ready to close the site.

Some of the issues that occurred to me as I was reviewing the application was why only two weeks of non-operative treatment? Most patients will do well after -- most studies have carried out four to six weeks minimum of non-operative care. How did the Sponsors define foraminal stenosis? Were objective criteria used? And would the Sponsors recommend avoidance of this device in patients with foraminal stenosis.

I also noted that the control group had one CSF leakage, as reported, and one reported dural tear. And the control group had 3 of 115 patients where the hemostatic agent was left in place. This is a deviation from the listed intraoperative exclusions.

In addition, I noted a trend but not a statistical significance towards greater surgical

times, prolonged surgery, blood loss, and the blood loss range in the control group, as opposed to the Oxiplex group. And some of the questions as to whether the control group surgery was more complicated and potentially resulting in greater postoperative pain could have been obviated by an alternative randomization process, as opposed to the intraoperative randomization process. Thank you.

DR. MABREY: Does anyone on the Panel have a question for Dr. Rao?

(No response.)

DR. MABREY: At this point, we'll hear from Dr. Evans.

DR. EVANS: All right. I'm old, so I need to stay seated, if that's all right with everybody. And I don't want to turn into lecture mode.

Let me first thank the folks at FzioMed and the FDA reviewers for their diligence and hard work. I recognize the complexity of the issues, and I appreciate your efforts to try to understand the data.

I've been asked by Dr. Jean to very briefly summarize what I see as the key statistical issues, with the intent of identifying these issues, but not to comment on them specifically at this time. So I'd

like to make -- I'd like to talk about three topics very, very briefly.

The first is the composite nature of the endpoint, and there's been -- it's been alluded to during the course of the morning, but I'd like to make -- with my knowledge of composite endpoints, I'd like to make a couple of points about the composite nature of the endpoint.

I would like to make a couple of comments about the unadjusted analyses versus model-based analyses. And all of this leading into what I see as the largest statistical issue involved in this submission, and that is the interpretation of subgroup analyses.

So the first point I would like to make or issue I would like to bring up is the composite nature of the endpoint. And composite leg pain from the LSOQ was a combination of, as we saw in a presentation this morning, that it was a composite of asking the patient, "How much do you hurt now? What was your average pain during the day? What was your pain when it hurt the most, it hurt the least, the end of the day, when you're waking up," et cetera. And the advantage of a composite endpoint is that you can perhaps get a more complete characterization of

the effect and possibly more power and avoids a multiplicity issue if you can formulate such composites.

One disadvantage of a composite endpoint that we really haven't heard about this morning is that it can be difficult to interpret if the components of the composite vary in importance and if the treatment effects vary across those components as well. And so you can imagine a scenario where if you decide that the most severe pain that you have during the day is more important than the other pain levels and one treatment arm performs better with respect to the composite but performs worse with respect to the most severe pain, then we have difficulty in interpreting what that means. And so some discussion perhaps about the relative importance of these different types of pain may help in trying to identify or trying to interpret the results.

The other disadvantage of composite endpoints is that some argue that you can gain more power because you ask more questions and get more data. But you can also lose power. If there's effects in some of the components but not the others, you essentially dilute the effect. And it's something to be aware of when trying to interpret

these data.

So the primary endpoint involved in the key trial here was composite leg pain. And it's a very -- composite leg pain or any pain endpoint is a very, very complex endpoint, as pointed out by FzioMed. Pain is very subjective. It's highly variable and is subject to "a placebo effect," in the sense that there are now imaging studies that suggest even the expectation of pain relief can cause not only psychological changes but physical changes in the brain. And so it is a very, very complex endpoint, which leads into possibly thinking about whether we should do multi, sort of multivariable modeling approaches.

Now, despite complexities of endpoints, valid analyses of randomized clinical trials do not require covariate adjustment. The randomization is a very, very powerful tool, and from a statistical standpoint, randomization gives you valid treatment comparisons. And, thus, simple, unadjusted analyses are often considered primary.

However, model-based analyses can be informative, and certainly things like analyses of subgroups and subgroup questions are motivated by very important and very practical questions. Does

the treatment effect vary across patients that have different characteristics? And so this is an important question to consider.

So the issues behind subgroup analyses, there's three significant statistical concerns with subgroup analyses to be aware of. The first is that any time you chop your data up into subgroups, you can have smaller sample sizes within those subgroups. And smaller sample sizes means greater variation and less power to see certain things, and so with greater variation, more uncertainty.

The second of these key issues to be aware of is multiplicity. And every time you look at a new subgroup, there is a possibility of a false positive, and so there's an increased risk of a "false positive" or what we call Type 1 error in clinical trials. And so we have to be aware of the potential for a false positive result. Now, you can try to manage the multiplicity problem by pre-specification of what subgroups you're going to look at and what you're going to examine and with appropriate multiple testing procedures implemented in that.

And the third issue behind subgroup analyses is the analysis of subgroups is done through modeling. And modeling is subject to subjective

decision-making, modeling may generate data-driven hypotheses, and that modeling may need validation, assessment of its assumptions, instability, and sensitivity analyses.

And, thus, given those sort of three statistical concerns with subgroup analyses, they are conducted and interpreted with great caution. And, therefore, subgroup analyses in clinical trials has historically and generally been used to inform future research and viewed as either exploratory or hypothesis-generating or used to assess the consistency or the robustness of a result of a treatment effect across varying patients with varying characteristics and is generally used less so to make definitive conclusions within subgroups.

So in terms of the identification of subgroups -- and I think some of the issues that we'll need to discuss today is if you're looking to make confirmatory statements and confirmatory analyses, you base subgroup identification on data that are collected prior to randomization. If you're using data post-randomization, you can run into real issues. But, ideally, you try to pre-specify the subgroups very selectively and very specifically, hopefully, with some biological justification of why

those subgroups are selected.

And then subgroups that are sort of suggested by the data, or data-driven subgroups, are generally considered exploratory and require some sort of confirmation. And then when we go on to assess subgroup differences, as was done in these analyses, we assess subgroups using tests for interaction, which basically addresses the question of whether treatment effects vary across these different subgroups. So a common error is to actually compare P-values within subgroups, but that's generally not an appropriate thing to do.

So the key issues, as I see them, is the evaluation of the use of model development and subgroup analyses versus sort of the unadjusted, or analyses, and whether the control of false positive error rate has been compromised with the use of such modeling procedures. And so that's what I see as the key issues.

DR. MABREY: Thank you, Dr. Evans. Does anyone have any specific questions for Dr. Evans before I open up the discussion?

(No response.)

DR. MABREY: At this point, I would now open up the floor to the other Panel members for

questions to either the Sponsor or to the FDA. And this is also an excellent opportunity to ask the Sponsor and FDA more detailed questions that may require a more extended answer that they can respond to in the afternoon. It's known as the lunch rush.

I'll start with Dr. Blumenstein.

DR. BLUMENSTEIN: I believe it was the Sponsor, although it's very hard to keep track of all this stuff, mentioned an O'Brien analysis, and I would like further details on what that analysis consisted of.

MR. KRELLE: Thank you for your question. I think a larger explanation of the O'Brien would give the Panel a better understanding of that, so we prefer to do that with some support from presentation materials this afternoon.

DR. BLUMENSTEIN: That's what I thought.

DR. MABREY: Great. Yes, Dr. Sang?

DR. SANG: This is for the Sponsor. I wonder if you could comment -- I may have missed it in the inclusion and exclusion criteria, but I wonder if you could comment on the treatment that subjects obtained prior to surgery, during the two-week period or prior to that, if that's available, from pharmacological management, including tricyclic

antidepressants, anticonvulsants, SNRIs, epidural steroid injections, selective nerve root blocks, any of those, and, certainly, randomization is a powerful tool if the sample size is high enough. So I just wonder if that has been something that was looked at.

My second question -- go ahead.

MR. KRELLE: Yeah, I think you know what the answer is going to be. That seems like an extensive list, and we'll get that to you this afternoon.

DR. SANG: Thank you. My second question has to do with opiate requirements or any analgesic requirements post-op. Did you record the opiate requirements either immediately post-op or at the one-month, 3-month, 6-month windows?

MR. KRELLE: I think the answer to your simple question is, yes, we did. And if you would like to see some examples of those, we'll have those ready this afternoon as well.

DR. MABREY: Yes, Doctor?

DR. HORLOCKER: Terese Horlocker. This is a question for the Sponsor also. You had exclusion criteria as a dural rent or a previous lumbar puncture. What kind of safety data do you have, as far as if this device is injected intrathecally? Is

there any evidence of arachnoiditis? I didn't see any of those tests done in the pre-clinical studies.

DR. DiZEREGA: Thank you for your question. We'll be happy to share with you our experience with that after lunch.

DR. HORLOCKER: Thank you.

DR. MABREY: Dr. Hanley?

DR. HANLEY: You know, this is the most complex discussion of a simple clinical problem I've ever heard. Lumbar disc herniation is the easiest thing we deal with as spine surgeons. I've always thought it was pretty straightforward until today.

One of the first questions we ask patients in the office is what hurts more, your back or your leg, as we generally believe those who have a predominant back pain component will not do nearly as well with the surgery. Now, maybe I missed this, but I didn't see a percentage relationship between back pain and leg pain in these people with so-called severe back pain.

So my question is, do we have any information on this? Did these patients have back pain that exceeded their leg pain or did they just have severe back pain and severe leg pain, because I think that makes a difference.

The big issue is failure of the surgery from back pain, not usually leg pain. But maybe we could have some comments on those individual patients who were put into that category of 63 or greater who -- or proceed to have bad back pain.

MR. KRELLE: Yes, we can get that information. Just so that I get the question correctly, I think you're looking at perhaps the number of patients in the study who had back and leg pain?

DR. HANLEY: I presume everybody had leg pain?

MR. KRELLE: Yes.

DR. HANLEY: But did we have a group of patients who had more back pain than leg pain?

MR. KRELLE: Yeah, we do have that data, and that would benefit from a slide, too.

DR. MABREY: Dr. Goodman?

DR. GOODMAN: Thank you. I have a number of questions that you'll probably want to answer after lunch.

The first is I realize that the Sponsor has already stated that the two main components of the gel are well-known. Is there information that they can give us as to how long the combination is around

how it is metabolized, and why they chose this particular formulation out of many other polymers that could be available? I think that would be important information to know.

The second question pertains to some questions that have been asked by others. Some of the exclusions were people who got epidural steroids or epidural fat that was placed at surgery, and the question I think a few of the panel members have brought up is were these the more serious cases? So is there a systemic bias, because the patients who received epidural fat or local steroids or even epidural steroids, could those have been the more serious cases, and, therefore, were those excluded because they had more serious back pain and leg pain?

Another question pertains to the control group, and I think that this was asked by Dr. Sang. The control group, as I understand it, received no additional local treatment, and I'm wondering why the Sponsor chose this as a control. Why didn't they have a saline injection or another polymer injection that they thought would not have this effect rather than have no treatment at all?

The site variability issue, I think -- did you want to answer all these now?

DR. DiZEREGA: Actually, Dr. Goodman, we're really enjoying your questions, but I think I can take one out of the hopper for this afternoon.

DR. GOODMAN: Sure, okay.

DR. DiZEREGA: That's the issue of the control. We did speak extensively with the members of our scientific advisory board, all of whom performed spine surgery, as well as potential clinical trial sites, and we were told uniformly from an ethical point of view, the only acceptable control would be the standard of care, which, in fact, was addition of nothing. That was from an -- point of view, from an ethical point of view, the control that was chosen. And I thought -- I assume that that's the best way to go. That's what people do, and so we compared our product against the standard of care.

DR. GOODMAN: So, in other words, it was more of a historical control? Nothing else was done for the majority of patients, so it was a control based on the standard of care, historically?

DR. DiZEREGA: Actually, it wasn't historical. It was proactive. So these patients were randomized on a proactive basis.

DR. GOODMAN: Right.

DR. DiZEREGA: It wasn't historical

information.

DR. GOODMAN: I think we're saying the same thing but different words.

I think the issue about site variability must be addressed. I think at least for me and perhaps a number of the other Panel members, some of the graphs that we saw were very provocative with regards to site variability, and I'm hoping that the Sponsor will explain this in detail in the afternoon.

I suppose my final question pertains to a philosophical question. If the substance is placed around an exiting nerve, how does the Sponsor postulate that it decreases back pain? If you could answer that question, it would be very informative.

Thank you.

DR. MABREY: Dr. McCormick?

DR. McCORMICK: Thank you, Mr. Chairman. I have a couple of questions, and I'm happy to wait until this afternoon.

First, I want to echo Dr. Hanley's observations as a full-time practicing spinal surgeon. I'm a little puzzled by the patient profile here, in the sense that back pain was such a predominant compliant. On average, the back pain scores are only 8 points less than the leg pain

scores. I think one was 67 and the other was 59. And that either raises the issue regarding the sensitivity or the discriminating value of the measurement tool or the patient population itself, because I think patients who are operated on for a herniated lumbar disc almost exclusively and certainly predominantly have leg pain in excess of any degree of back pain. So I would like that addressed, if you could, in a little bit of detail this afternoon.

The other issue with respect to the scoring of the instrument, listed on Page 39, under the statistical analysis, the way the instrument is said to be scored, to my calculation, that gives a range of 20 to 120, but it's listed as 0 to 100. So if you could just tell me how you -- just how you did that, because it doesn't synch up in what was given out.

And the final issues are, you know, as we look at these statistical associations, we try to -- they are certainly there in some of the analysis, and we have to look at, you know, is it statistically significant and is that statistical significance a deterministic one. And for that we look at the associations, and we try to determine biologic plausibility. Is there a reason why one would

causally relate to the other?

Recurrent disc herniation, for example, both the clinical presenters this morning inferred that this seemed to be a direct result, even though it didn't reach statistical significance, it certainly approached it. Six patients in the control group underwent re-operation; only one in the treatment group. And I'd like to know what the biologic mechanism for that increased rate of recurrent disc herniations is, if, in fact, that's what they -- if I interpreted what they were saying. You know, I'd like to understand that a little bit better.

And one of the comments made this morning, the surgeons were not blinded to the treatment, and so the surgeons explicitly, in this manual, as I see it, surgeons determined whether patients went back to the operating room for that recurrent disc. So non-blinded surgeons were making those decisions, and I think that's just a correction -- if I'm wrong, please correct me.

And then the final issue is in terms of statistics -- statistical significance and clinical relevance. And I think someone else brought it up this morning that for this study, for this

instrument, the LSOQ, the MCID was stated to be either 20 percent or a 33 percent reduction. And on none of the primary analyses nor any of the secondary analysis, either for the whole unadjusted or using the models, demonstrated what would have been by the FDA's and I assume the Sponsor's own acknowledgment of what would considered an MCID. And I think if there is some -- there was a lot of terms thrown around this morning about clinically important, clinically relevant, clinically significant, but none of them as they relate to the MCID, which is defined as the minimal difference that a patient perceives as beneficial. Thank you.

DR. MABREY: Thank you. Dr. Blumenstein?

DR. BLUMENSTEIN: Just a few more items probably for this afternoon. I'm interested in some analyses with respect to whether the missingness of data that led to the reduction from the ITT to the CC analysis sets -- I think I've got those terms right -- whether the patients, for example, that didn't get their 6-month follow-up on time or didn't have it at all, whether there is something informative about the missingness of the data.

I'm also interested in there was something mentioned about a correlation of 0.55 between the leg

pain and back pain. I think that's what was said. Another interesting correlation to me would be the correlation between baseline and follow-up for, say, leg pain.

I'm also interested in whether the -- how many cases remain if all cases with any one of the covariates being missing are deleted. In other words, what is the subset of patients who have complete data on all covariates? What's the size of that subset.

And, finally, I would like to know a little bit more about the randomization mechanism. Was it blocked within clinical site? What was the block size, and so forth?

DR. MABREY: Thank you. Yes, Dr. Sang?

DR. SANG: Just wanted to briefly elaborate on the question I asked earlier, which was about preoperative management or management of subjects prior to their enrolling. Actually, if you have the data -- you may not -- here for the different sites, particularly -- I cannot read this, but it looks as though the two sites on either end of the extreme on this site variability slide from the FDA, it looks as though 24 subjects enrolled at the site with a negative effect, treatment effect, and 33 enrolled at

the site with the largest treatment response.

I'd be interested to know whether or not -- what management was chosen prior to these subjects enrolling, but I'd also be interested -- I don't want to say again that I have an interest in nerve root -- whether or not these subjects had selective nerve root blocks. I don't actually -- I'm more interested in the diagnostic approach of using these blocks, actually, than the deposition of steroid at that level. So if you have that, that would be very useful. Thank you.

MR. ZHOU: I just have one quick clarification on the slide we showed there, the graph. The number there represents site numbers, not number of patients enrolled in the site.

DR. SANG: I'm sorry about that, then.

DR. MABREY: Thank you. Ms. George, do you have questions for FDA or the Sponsor?

MS. GEORGE: Definitely have one for the FDA. Excuse me. I'm -- allergies. There seems to be a lot of focus and question on the statistical analysis plan, which I thought, based on my understanding, that the endpoints were all agreed to and the plan was all agreed to. So a lot of the questions seem to be about validity of the plan. And

that seems to be a little bit unclear to me as to why the validity of the plan would be questioned and the focus would be so much on presenting the data in the univariate format when it seems that it was already discussed and agreed upon for the multivariate methodology and the multiple endpoints.

DR. MABREY: Does the FDA want to address that now or later?

MR. ZHOU: I think we'll take a quick -- address it later, but in terms of the initial responses, the supplement proposing an alternative to the originally proposed statistical plan was approved by FDA. The issue of trying to understand what the data meant is why we looked at the information as originally proposed in the original IDE and, subsequently, as proposed in the supplement that was approved in '06.

DR. MABREY: Thank you. Ms. George?

MS. GEORGE: I think that's it for right now.

DR. MABREY: Ms. Whittington?

MS. WHITTINGTON: The only thing that I would like to ask again -- and I think one of my -- the other Panel members, several of the questions I had they've already put on the table -- was the

number of failed treatments and the types of failed treatments these people had before they had surgery. There is not any discussion about that, and you hit on some of that as well, you know, what was done before you tried surgery. It should not have been certainly a first approach, and I'm sure that it wasn't, but that wasn't included, that I could tell.

MR. KRELLE: Thank you. We'll include that in our afternoon session, too. Thanks.

DR. MABREY: All right. And Dr. Rao?

DR. RAO: Just a quick -- some of my questions were in my presentation. Just a quick follow-up on what Dr. Blumenstein was talking about. Just to play devil's advocate, if the correlation between back pain and leg pain preoperatively was 0.55, which I presume is poor correlation -- is that not poor correlation?

UNIDENTIFIED SPEAKER: No.

DR. RAO: Okay. Then, I'll retract. What would a good correlation number be? Should it be higher or lower? Could you answer that question?

UNIDENTIFIED SPEAKER: Higher.

DR. RAO: Higher? So 1 would be an ideal correlation?

DR. CHIACCHIERINI: Yes. This is

Dr. Chiacchierini again. That is correct. Zero would be no correlation.

DR. RAO: Correct.

DR. CHIACCHIERINI: One would be perfect correlation, and for clinical trials, a value of .55 is pretty good. That's pretty good correlation. Poor correlation would probably be in the vicinity of less than .3.

DR. RAO: Generally, when I submit papers for publication, if I have correlation of .55, the reviewers seem to reject it, in terms of -- variability --

(Laughter.)

DR. RAO: So I presumed .55 was relatively low. And if the correlation between back and leg pain, preoperatively, is low, just to play devil's advocate, why should relief of leg pain in a subgroup with severe back pain be clinically relevant? That would be just a hypothetical question.

The second question I have is -- Dr. Rhyne talked about this briefly. Is there any intrinsic known analgesic effect from the device itself? Does the device contribute to reduction in inflammatory markers, cytokines, or anything else, locally?

DR. DiZEREGA: Thank you for your question,

Dr. Rao. I think we can respond to the question regarding an intrinsic effect. All the effects that we are aware of with these biomaterials, polyethylene oxide and carboxymethylcellulose, our device, no pharmaceutical effects that we're aware of. And so I think the answer to your question is no. We would not expect in a de novo situation any analgesic effect, per se, from PEO or CMC.

DR. RAO: The third question is, in part, from my poor knowledge of statistics and is more a philosophical or a global, or like my chairman likes to say, a 30,000-foot view question, and it's addressed to the statisticians, both on the Sponsor's and the FDA side.

I think the theory behind this multivariate analysis is that back pain is complicated and that there are many factors that contribute to back and leg pain in patients with spine problems. And I think that assumption is fairly accurate for the most part. When we design a statistical study, if we're looking at relieve of leg pain as a primary effectiveness variable and if we want to impute a number of covariates that may effect this primary leg pain, we have to look, from my ignorant background, we have to look at the potential causes that may be

contributing to this leg pain.

So, for example, we may look at disc herniation. We may look at foraminal stenosis. We may look at other factors, instability, other issues like that. And that may give us some reason to understand what are the different factors contributing to the relief or lack of relief of the leg pain. But my interpretation of the Sponsor's application is that they have looked at a number of symptoms or findings in the patient group as opposed to underlying physiological pathologic conditions.

So a statistical design of that type, to my understanding, would be more geared or designed to produce a group of patients in which this device may work out, as opposed to understanding whether this device actually helps with a particular condition. So the statistical design, if you'll excuse my poor explanation of what I'm trying to get across, is designed to determine a small subgroup or a subgroup of patients in which this device may work better. So that's my understanding, and I'd just like you to give me your thoughts, maybe both from the FDA, as well as from the Sponsor's side, as to whether that thought process is, in general, accurate.

DR. MABREY: And, Dr. Evans, any final

questions? Any final questions before we go to lunch?

DR. EVANS: No, I guess I don't have any final questions.

DR. MABREY: Go ahead.

DR. SANG: I do, to elaborate on Dr. Rao's comments because -- comment -- which is, in fact, exactly why I asked about selective nerve root blocks, what efforts were taken to try to identify the different subgroups of enrolled subjects and whether or not -- I mean, you certainly may be able to infer a number of different additional questions, questions that have been raised over many years in the pre-clinical studies.

I mean, I really admire this company for choosing outcomes that are challenging in the world of analgesic clinical trials. In the world of analgesic trials, using pharmacological therapies, it's very hard in low back or radiculopathy to identify the appropriate endpoints. You have really looked to identify these endpoints. It's unfortunate that the primary endpoint, which was a composite endpoint, did not turn out to be positive. But I think it was really a valiant effort to try to at least take this look.

I think that the question of leg pain versus back pain is the critical question because in the presence of mechanical back pain, we have a host of potential problems, even with radiculopathy, identifiable radiculitis associated with a herniated nucleus pulposus. We know from animal studies over several years that -- and you know this -- that you can see apoptosis at the level of the DRG. There is, you know, there is an acute inflammatory response associated with cytokines and chemokines, and everything you've already described.

So I think that this was really a valiant effort. I just wonder if you could, when you look at your subgroup analyses, perhaps identify other subgroups based on potentially what I would interpret as mechanisms, but certainly based on simple history and physical maybe diagnostic procedures, whether or not the subgroup of truly radiculopathy subjects may have had a better response.

DR. MABREY: Dr. Evans?

DR. EVANS: I guess I will make a couple of comments based on other questions and have one additional question myself.

So there was talk about some correlation and how do you interpret .5 correlation, and I'm not

sure I understand the reason for examining this correlation exactly, but just to give you a very simplistic way of -- a correlation of .5 with some assumptions essentially says that knowledge about this endpoint would explain about 25 percent of the variation in another endpoint, and so that's sort of a quick way to think of how the two are related.

You also had a question about the importance of other covariates and other things that could potentially affect outcome, and the issue with randomization is that there's an expectation of balance with that, particularly in large trials. And that's why you get valid sort of inference in randomized studies.

But to get more at your point about -- your question about that, I would like to ask a question, as I do think the key issue here is the interpretation of these subgroup analyses and whether there's a control for -- whether there has been a threat to the control of a false positive error rate.

And so the question is this. The Sponsor, in their analysis, examined I believe they said 40 covariates and their related interactions for potential inclusion into the model. And so it's the natural concern that there could be subgroups for

which treatment effects could vary.

Now, there's been allusion to some control of error rates by looking at the interaction first. And by looking at the interaction first, it may negate you from looking at subgroups. If there is no significant interaction, you don't look at subgroups and therefore sort of avoid potential false positive error rates by looking at subgroups, by examining interactions first.

However, I guess my question is, how do you control -- if you're going to look at 48 possible interactions, and you're looking at those interactions at a .15 level, I believe, how do you control a multiplicity issue of every interaction you examine has a possibility of becoming significant, even erroneously, even false positive significance? And so every interaction you examine has a possibility of a false significance. And by doing that, and if an interaction becomes significant, then you go looking at subgroups. And so my question is I don't understand how error rates are controlled if you're looking at lots and lots of interactions because every interaction you look at, there's a possibility you could find something that isn't there.

And so that's sort of independent of the issue of looking at the interaction first and then subgroups, but every interaction you look at, there is a possibility of a false inclusion. And if you're judging effects based on the finding of an interaction and you've looked at 48 interactions, then there's concern for false positive error rates, sort of losing control of those error rates.

And so that, I think, from a statistical standpoint, as I see it, is really the important issue to address. And so that's sort of my final question.

DR. MABREY: Dr. Blumenstein, one last comment before we break for lunch.

DR. BLUMENSTEIN: I'm going to pile on here. I, too, worship at the altar of randomization. And this is central to the issues that we're considering here because it's really -- it's control of the alpha or the false positive probability is the big issue.

And so just to provide another perspective on what was just said, let's suppose that we could generate thousands of data sets that would resemble the control group from this study. And we would also generate that same, just randomly generate that, for

the intervention group, for the investigational group. So what we have is a bunch of data for which there is no effect at all from the investigational intervention. And we subjected those thousands of data sets to the kind of analysis that was done here, the modeling, the statistical modeling, the screening of interactions, and so forth.

The issue is what percent of the time would you come out and find something significant even though the data has absolutely no treatment effect in it at all. And in my own intuition, my mathematics, and everything else points to the fact that this is going to be -- that you're going to find a great deal more significant findings from that kind of a thought exercise or even if you did it actually. It's going to be a lot greater than the 5 percent that is what's normally thought of as a controlled type-one error probability or a false positive probability.

Another thing that is a little bit confusing here is that the Sponsor uses the term multivariate to mean two different things. There is multivariate in the sense of multiple outcomes and the relationships between multiple outcomes. And, also, the term multivariate is used to describe what are really statistical modeling procedures, where

there is one outcome and multiple explanatory variables.

The O'Brien Test that was mentioned earlier is more of a true multivariate, if it's the one I'm thinking of, because it is modeling multiple outcomes simultaneously rather than trying to model the prediction of a single outcome by multiple predictor variables. So we need to be careful about what we're talking about when we talk about multivariate analysis.

DR. MABREY: Thank you. It's now 12:00. We'll break for lunch. We'll reconvene again in this room in 45 minutes, at 12:45 p.m.

Please take any personal belongings you may want with you at this time. The ballroom will be secured by the FDA staff during the lunch break. You will not be allowed back into the room until we reconvene.

Panel members, please remember that there should be no discussion of the PMA during lunch amongst yourselves or with any member of the audience. Thank you. And I'll see you at 12:45.

(Whereupon, at 12:00 p.m., a luncheon recess was taken.)

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

(12:57 p.m.)

DR. MABREY: Thank you for showing up almost on time. And I know that applies to some of our Panel as well, but it's almost 1:00, and I would like to call the meeting back to order to resume the panel discussion. Is the Sponsor prepared to respond to Panel questions from this morning?

MR. KRELLE: Yes, we are indeed, and we're going -- we're going to put on statistics first. Then we're going to talk about some mode of action questions and general questions -- biological -- and then -- questions that were raised (mic turned off).

DR. MABREY: Yeah, if we could have a little help with the microphone there.

DR. CHIACCHIERINI: Thank you, Mr. Chairman. I'm going to try to do the simpler questions first, and then move on to the more complex. The first question that I will answer is with Dr. McCormick's issue about the different populations that had back pain over 63. And I believe he mentioned 101 and 92 in one slide and 78 and 78 in the other.

The simple explanation for this is, this deals with two different populations.

DR. McCORMICK: I'm sorry?

DR. CHIACCHIERINI: It deals with two different populations. The 101 and 92 is the split that occurs in the intention-to-treat population. The 78 and 78 is the split that occurs for patients who have back pain, baseline back pain over 63 in the completed-cases-within-window population.

DR. McCORMICK: So that's a difference of 50 patients --

DR. CHIACCHIERINI: Yes.

DR. McCORMICK: Almost 50 patients?

DR. CHIACCHIERINI: Yes.

DR. McCORMICK: Okay.

DR. CHIACCHIERINI: Remember, there were 48 patients who had visits outside-of-window plus 5 more who had it beyond one year.

DR. McCORMICK: So in your histograms that show the *n* of 78, that show this difference in outcomes in the severe patients, their just limited to the smaller population, the 78 --

DR. CHIACCHIERINI: It is limited --

DR. McCORMICK: About 50 patients are excluded from that?

DR. CHIACCHIERINI: That is correct --

DR. McCORMICK: Fifty patients with severe

pain are excluded because they were outside the window, the --

DR. CHIACCHIERINI: They were well outside the window.

DR. McCORMICK: Okay.

DR. CHIACCHIERINI: Most patients were well outside the window. The one patient had a visit very close to 52 weeks.

DR. McCORMICK: Okay. Thank you.

DR. CHIACCHIERINI: Okay. The next question I'll respond to involves the O'Brien's Test. Dr. Blumenstein is absolutely correct. There was a mixture, an inadvertent mixture of terms and terminology. The multivariate analysis does, in fact, analyze several outcomes, and O'Brien's Test was the only true multivariate analysis that was done.

And what O'Brien's Test does -- it's based on a mid-1980s paper on biometrics by O'Brien -- is that every patient's improvement was ranked across the entire population of patients. The ranks for all of the seven endpoints were then summed, and then the sum of the ranks were tested by a T-test. Before the T-test was done, a test of the normality, the consistency with normality, and, remarkably, the sums

had an excellent consistency with normality and the T-test was consistent.

So the value that was provided for O'Brien's Test was across all seven variables. It was done on the within-window population across all treatment -- all baseline back pains. There was no subgrouping. And so that analysis is a global analysis based on the within-window population.

To carry on --

DR. BLUMENSTEIN: Remind me what the P-value was.

DR. CHIACCHIERINI: The P-value was .0496. I need to explain, however, that both leg weakness and satisfaction had a very limited scoring range. As you know, rank tests -- this is a rank test -- rank tests are very sensitive to ties. Both of those variables had substantial ties, and when I remove one or the other of those variables, the P-value drops very nicely to .03 and down to .02.

To discuss Dr. Rao's question about the .55 correlation, what does it mean, we did, in fact, compute the number of patients who had both leg pain greater than 63 at baseline and who had greater than 63 back pain at baseline. And by some quirk of circumstance of fate, that turns out to be 55

percent. So I think the relationship of 55, while interesting, I think that's certainly just a circumstantial relationship. But 55 percent of the patients were in that upper box of greater than 63 leg pain/greater than 63 back pain.

The sample size, Dr. Blumenstein, your question of sample size, the sample size justification was provided. It is well-documented and is that which was given in the FDA presentation. That is the exact reason that the sample size was the number that was provided and that, since it was unknown -- while it was strongly suspected by the Sponsor that there would be a multivariable analysis, it was not precisely known how that would occur. And so the ability to use that in any sample size computation at the beginning of the study was extremely limited.

To discuss your question about what was involved in the SAP, the SAP, as is the common occurrence in almost all of the clinical trials with which I have been involved, the statistical section of the protocol is usually very rudimentary and doesn't provide enough information to allow an adequate evaluation. And so prior to database lock, there is a drafting of a detailed statistical

analysis plan that tells precisely what will be done and when it will be done.

In that plan that was approved by FDA and which is the primary analysis method, as we found from Mr. Melkerson, the primary analysis was to be an ITT multivariable analysis of leg pain. The patient covariates were pre-specified. The list was much longer than the Sponsor had anticipated simply because of the insistence on it by FDA of including all clinically relevant covariates.

The method of imputation was pre-specified in the SAP. The method of model screening and selection was totally pre-specified in the SAP. And, in fact, we went on in great detail to repeat the language for each variable that we analyzed.

However, the investigation of the interaction, we said in the SAP that we would study it. We did not pre-specify the mechanism of that study.

Dr. Rao, the improvement in disability days was done globally, across the entire population, with the limitation that it was done within the within-window completed cases population. There was no subgrouping other than making sure that the subject had a within-window visit at 6 months.

Now, Dr. Evans, your question about the interpretation of the LSOQ. In conferring with the clinicians at lunch time, there are six questions for leg pain and six questions for back pain, and while for any individual patient one of those questions may be more important than the other five or possibly two or more, it is impossible to tell for any patient, as you well know with a pain instrument, that any one question will be globally important to all patients and so this is a validated construct and its validation is what we have to rely on for its validity.

Your comment on the fact that the unadjusted analysis should have merit because there should be balance among the baseline characteristics is valid only in part. While it is true that randomization does provide not a guarantee but a strong amount of balance between known and unknown baseline covariates, it cannot possibly balance on interactions with treatment because that is determined after the study is done and not at the time of randomization.

The subgroup analyses resulted from a valid statistical pre-specified interaction. It is the P-value for that interaction that is the thing that

leads us to do the so-called post-hoc testing to determine and to interpret the interaction. So while our discussion of P-values, and so forth, is important, the global interaction P-value is the thing one should not lose site of, and that was .0113 for leg and .0007 for back.

And one of the analyses we did not present to you was that for patient satisfaction because -- just to give you a little bit more information, we did, in fact, employ the hierarchical principle. We went to leg weakness and leg weakness did not show significance. It did show significance in an interaction. However, there was an inconsistency between the ITT analysis for leg weakness and the CC analysis for leg weakness. And so we dismissed the ITT result and stopped testing. However, we did do post-hoc testing of the endpoints, and patient satisfaction also exhibited a baseline by back pain interaction.

The number of patients that were missing at least covariate at the -- for the GEE analysis for the final model was only 10 patients. And that's pretty remarkable. That is, those 10 patients were missing at least one of the covariates and were not included in the final model. And that's 10 patients

out of 351.

Now, for the really difficult question. How do we account for multiplicity? I don't think anyone knows. And the reason no one knows is because you cannot do a simplistic computation as though all of these covariates and all of these interactions are independent of each other because they are not. I was careful to point out that the elements of the LSOQ are correlated, and they're not correlated with a low number of .1 or .2 so we could dismiss it. It's something that we have to consider.

On the other hand, computing that -- trying to compute the possibility of an alpha inflation with a complex procedure like this would be extremely difficult from that basis alone.

On the other hand, we have discussion from Dr. Pocock's article in *Statistics in Medicine* on the value of interactions in clinical trials. And in that discussion, Dr. Pocock says that interactions are actually underpowered. So, on the one hand, you may have some alpha inflation indeterminate, and on the other hand, you may have a power issue with regard to interaction. So the only thing I can tell you is we can provide no further elucidation on either of those issues.

I think that's the predominant number of statistical questions. Are there others?

DR. MABREY: Panel? Yes? Dr. Blumenstein?

DR. BLUMENSTEIN: Did you have some data on the correlation, say, between baseline leg pain and one of the follow-up times?

DR. CHIACCHIERINI: We did not have opportunity to compute that. I apologize. I can tell you that the correlation would not be as high as the correlation at the beginning of the study because so many of the patients dropped to a level.

If you looked at the responses over time, the improvements over time, the improvements over time at one month, there was nearly a 50-point drop in the improvement -- a 50-point drop in the pain score at leg and back. And that improved slightly over time, but it improved differentially between the treated and controls group, as you can see from my slide.

DR. BLUMENSTEIN: Actually, what I thought you might come back with was the estimate from the GEE modeling of the covariates matrix between the -- I assume that you did compound symmetry or exchangeable covariate --

DR. CHIACCHIERINI: We have that, but it's

deep -- we didn't have sufficient time to investigate it.

DR. BLUMENSTEIN: Okay.

DR. MABREY: Thank you.

DR. RAO: Excuse me, I just have a quick question.

DR. MABREY: Oh, I'm sorry, Dr. Rao?

DR. RAO: For Dr. Chiacchierini, again. I'm sorry. Just a follow-up on a statistical question. I think your point on the rule of interactions is very valid, and, in some cases, the rule of interactions may be underestimated or lower than it should be. But I think my point earlier was the covariates we choose will affect our statistical interpretation of the study. So I'll try to put it somewhat simpler.

Suppose we're trying to look at the effects of lunch on my system. And if we're looking at a pasta dish or a salad dish, and if we look at factors like how hungry I was before lunch, how satiated I feel after lunch, how many times I belch after lunch, things like that; whereas if we look at factors like age, height, lactose intolerance, genetic abnormalities, things like that, the statistical results of any computation on this set of data will

vary based on what we throw into the statistical program.

So if we look at things like how well I feel satiated after lunch, how hungry I was before lunch, when did I eat my last meal, and if we get statistical results from that set of data, we're going to be able to attribute one set of conclusions from that set of data; whereas results from maybe genetic abnormalities, lactose intolerance, will give us some scientific results as to causation, physiologic basis of efficacy, and things like that.

So in a study like this, if we're looking presumptively at some barrier effect around the dural sac and the nerve root and relief of leg pain, which I presume you would agree with?

DR. CHIACCHIERINI: Yes.

DR. RAO: If we're looking at relief of leg pain, maybe instead of looking at baseline back pain, sexual function, GI intolerance, things like that, if we looked at the presence of concurrent foraminal stenosis, foraminal height, that might give us some more valid statistical conclusions that we could use or put our fingers, grasp a little tighter as to which particular we should use this in; whereas the data we've thrown into the statistics right now is

the presence of concurrent back pain and the other things like that.

So the results and the conclusions we draw from the statistical data now is in which type of symptomatology or clinical findings, preoperatively, this product is most efficacious in. It doesn't tell us why this product is efficacious. That was my point, and I was just wondering if you have a response as to how the statistics may apply to this device or to the physiologic basis of this --

DR. CHIACCHIERINI: Dr. Rao, you know a lot more statistics than you're letting on.

(Laughter.)

DR. CHIACCHIERINI: You're absolutely correct. The model one comes out with depends upon what is considered at the beginning and throughout the modeling process. We did not do the measurement of the spaces, and so on, at surgery. You know, maybe at another time, we would propose to do that, but that was not done in this trial. And so I couldn't use that information in the modeling process. Had we done so, we would have done so.

But what we do know is that the responses to pain whether it be back, whether it be leg, are correlated. And I don't know if it's a psychological

phenomenon. I don't know the rationale, but when people feel better, they will fill out the questionnaire and have less pain in both aspects. So while we could not address all of those issues, we did the analysis that provided information on the variables for which we had measurements. And that's all I can say.

DR. DiZEREGA: As you can tell, Dr. Mabrey, I've been trying to organize this into some logical sequence, and I'm approaching brownian motion, so I thought I would try to address a number of the Panel's questions in groups and then the questions that are left should be two types. One, the interesting pre-clinical questions that Dr. Sang asked and relating to pre-treatment, and another speaker will address those, whereas I'll try to address things relating to some of the bits and pieces that were asked.

And then something that I think we've all been thinking about this morning and, frankly, we've been thinking about for many months, and that's the mechanism of action, or, as Dr. Goodman said, give us a plausible biological reason about these two compartments, leg and back, and how they're interacting, as Dr. Hanley alluded to.

So, as Dr. Chiacchierini said, I'll leave the hardest one until the last, but I think we have some information that will be of some interest to you.

The first question, and I believe Dr. Horlocker asked it, was about intrathecal studies. If this material was used in a patient that had some kind of a durotomy, irrespective of the size, what would be the consequence to the individual? We did two studies to actually address that specifically.

In one study, we administered 1 milliliter of Oxiplex directly into the thecal compartment of primates, and we measured a number of things, including intrathecal pressure, general well-being, and the usual types of things. And we found absolutely no alteration whatsoever in any of the parameters.

And then, of course, you saw the dural nick study histology data, and I'd like to reemphasize that dural repair occurred normally. There was no evidence of an inflammatory response, obviously, on the one histological slide that you saw, but if you saw them all, it simply -- there was no inflammatory response, and the animals ambulated just fine. They

seemed perfectly well.

I think one of the more interesting observations that address your question -- and this isn't something that the Sponsor did. This was something that, in fact, has now been published. A European study by an anesthesiologist doing epidural anesthetics professionally took some Oxiplex, add an opiate to it, and administered it as a spinal anesthetic. And what was reported is that the duration of the anesthesia was, in fact, prolonged. That is, overall, less opiate was administered throughout the surgical procedure, and, in addition, the patients did very, very well, to the point where the physicians that actually did this went on to publish it.

So I think from the standpoint of view of intrathecal administration, that seems to be in pretty good shape from a safety consideration as to the other things -- from a safety consideration.

Now, what I'd like to address next, if I could just show this slide, is the issue of the different numbers in the different populations. And, as Dr. Lee said, and we certainly agree with him, we want the Panel to be entirely clear about which populations were referred to today and why we think

the most important population and, really, the population that we should be considering, is the 286 patients, which is the per-protocol window.

And what I'm referring to specifically are the patients that had their 6-month LSOQ forms filled out within a 22 to 28-week window that was specified by the protocol. And those were the 286 patients that we spent our morning talking about and we discussed the efficacy broadly.

UNIDENTIFIED MALE SPEAKER: Standby one minute --

DR. DiZEREGA: Well, in that case, I'll have a drink of water.

And the concept that I'm going to get to is the time-sensitive nature of measuring pain outcomes and particularly postoperative pain outcomes. There is ample information in the literature which has made it very clear that the ability to reproduce -- measure pain following surgery as a function of time can be quite challenging depending on the type of surgery.

In the case of decompression surgery, in 2004 or early 2005, there was an international consortium that met to make some recommendations on when this kind of information should be collected.

When should postoperative pain information be collected following decompression surgery? When is it the most predictive of longer term outcomes, and when should it not be collected? And this was published in the *European Spine Journal* by Mannion and Elfring. And they were very clear that the ability of patients to recall current pain at the time they're asked is indeed very accurate, but that information changed as a function of time following decompression surgery. And they made a very strong statement in that consensus that 6 months was the appropriate time.

And so we used that information to define our protocol window. And that per-protocol window we've referred to as the 286 patients. And so if you look at the overall flow of the patients, overall, there were 352 that were randomized in the intent-to-treat population. You can see that distribution was balanced. These are the patients that came out because their 6-month LSOQs were collected beyond the protocol window of 22 to 28 weeks. And Mr. Zhou talked about those patients earlier, and Dr. Lee talked about some of the efficacy determinations if these patients are included in the population.

This is the population that we think is

appropriate for the reasons that I've said, the time-sensitive nature of pain collection after decompressive surgery. And as you can see, as you know, those patient populations are balanced. And so that's sort of the way patients overall are broken up.

With that in mind, I'd like to talk about the variability that was discussed earlier this morning. I thought there were some very good questions about that. And that's Slide CZ-8, CZ-8.

The point was made earlier this morning that there was site variability, and, indeed, Dr. Chiacchierini showed that with his site effect in his multivariate analysis. You saw that very clearly. And Mr. Zhou showed the same thing.

We're clearly of the opinion that that site effect, although a main treatment effect, is really not important when you're asking the question about the differences between Oxiplex and surgery alone because there is no site-by-treatment interaction. That P-value was 0.6.

All right. That's a statistical argument. What's it look like from the standpoint of view of the results of the individual sites if we plot them site by site as we've shown for you here? The

population is the within -- per-protocol population, the 22 to 26 weeks of the group that had the severe baseline back pain, the ones that we think clearly benefited from this product. The unadjusted change from baseline is shown for you on the vertical axis, and this happens to be leg pain. Oxiplex is shown for you in the blue, and control is shown for you in the orange.

And as you scan your eye across this -- and make no mistake about it. The ends are very different. If you did the different ends in the populations, obviously, they're quite different. But as you go across the slide and you ask yourself the question how much site variability was there, I think you would come away -- at least I would come away by saying that, essentially, in every instance, the blues are above the oranges. Obviously, there is some ties, as you see here. But, quite clearly, going across the sites, the sites show that where there was a difference Oxiplex patients fared better on individual sites than control in this particular population.

Now, there were some questions about effectiveness, and I'd like to show this slide, please. And I think there were a lot of questions

about effectiveness. I think it's clear that the Sponsor's claims for effectiveness, in terms of the LSOQ, are focused principally on the subgroup. Our observations about clinical safety are, of course, across the ITT population. And as we've shown this morning, there are, in fact, some very important differences in clinical outcomes in the entire population, one of which, based on Dr. McCormick's question, I'll get back to very specifically.

But I would like to remind the Panel and then go on to talk about what Dr. Goodman asked me about the plausible reason why this would be the case, and at least we'll give you what our thinking is right now.

On the left-hand side are the results of the leg pain improvement in the subgroup that had severe baseline back pain. This is the ITT and this is the CC. And, of course, this is the back pain improvement. I don't think you saw this slide, but, obviously, the curves are very, very similar. The numbers are different because of the different populations. But at least to my eye, looking across this at 6 months, it's very clear that there was a difference -- obviously, statistics speak for themselves, but, quite clearly, there was a

difference in these populations.

Now, if I could just show this slide, please. Dr. Horlocker asked I thought a very specific question. I'd like to address it in a very specific way. And it related to the whole issue of success of the study. Dr. Lee brought that up earlier in the criteria that were available. And let me first make the point that the -- and Dr. McCormick talked about the minimum effective difference.

The success criteria that both Sponsor and FDA talked about were Sponsor's success criteria. The 20 points and the 33 percent difference, those were recommendations or advisories by the Food and Drug Administration. So those are what we're actually addressing. They're not the Sponsor's success criteria.

I think, having said that, if we focus for a moment on the controls, and we're now talking about how much pain was left after surgery, we refer to this as the residual pain. So this is the patient's had their surgery. Whatever very large reduction in pain occurred as a result of the surgery, which approached 70 percent in the study in general, that's very difficult to show a difference when the magnitude of the control is so large. When the

control changes 70 percent, it's very difficult to show a difference.

So we think the more clinically important contribution of this product is reducing the amount of pain that's left, the residual pain. This is the kind of pain the patient would go home with following surgery. And with the use of Oxiplex, that residual pain is reduced by the percentages that you see here. And these are, of course, the P-values.

So we thought it was very important to bring to Panel's attention that, number one, this is, in fact, the residual pain. This is not starting from baseline, so we're clear about that. And, number two, that this 35 percent is above the 33 percent, and this we think is the most robust difference that we saw relating to leg pain reduction. And, as you've heard, we think it's quite clinically significant, and I might say it's somewhat -- I think it's quite impressive. As Dr. Rao alluded to earlier, with all the other things that have been tried in this area, this is far and away the most impressive result that's been published. I think it's remarkable.

But, to be sure, this is the amount of pain that the patients would otherwise go home with. This

is not the preoperative baseline pain. This is the residual pain.

The other point to make with this slide is the one about the 20-point reduction. If you look at the baseline -- excuse me -- if you look at the control -- and I have actually forgotten what that number is, but it's almost 20 -- 21? Thank you, John.

It's fundamentally impossible in this kind of a clinical trial to eliminate pain. Had we eliminated pain, then there would have been a 20-point reduction from -- compared to the control. I don't think that's a tenable hypothesis. So this is the information that we think does make clinical sense and, obviously, showing the benefit to the patients.

Now, if I could just have the slides off for a second, I'd like to address the issue of mechanism of action, and I thought although you all asked us this question, I thought Dr. Goodman put it in a very nice context: some sort of a plausible biological hypothesis as to why we see the results that we see.

And let me first start off by saying we don't know. So let's just put that on the table. We

do not have evidence from this clinical trial that would directly address in any kind of scientific way the mechanism of action that we would like to talk with you about and share with you at least our hypotheses and the information, much of which you've spoken about today. But I would like to bring to Panel's attention an additional animal study that was not performed by ourselves, but it was performed at University of California in Santa Barbara.

And so, Jan, if you could project the slide that starts out with polysaccharides, that would be helpful. Well, here is our current view. As you know and as Dr. Sang and a couple of you have mentioned, in a situation where there is a compression on the nerve root, there is an inflammatory situation. There is also an enhanced sensitivity to pain because of that inflammatory situation. Quite clearly, by reducing the compression on the nerve root, by removing the herniated material or whatever is compressing on the nerve root, there will be reduction in leg pain. As Dr. Hanley said, that's a very straightforward concept.

But what about locally in the back? Well, the back is a very complex concept because the

etiologies of back pain we could spend the rest of the afternoon talking about. But I would like to say that there is a homology in the interaction of cytokines in producing pain during the preoperative condition as a result of the compression on the nerve root that innervates the sciatic nerve. And so by removing that compression, you would expect a reduction in leg pain, and, of course, we saw a reduction in leg pain.

Now, what about the back pain? Why would that be reduced in a situation where patients have a lot of back pain? Why would they have a lot of back pain? We're hypothesizing that the severity of the inflammatory response that is occurring as a result of the compression on the nerve root is also affecting the sensory components of the back. And following surgery, in fact, there is an outpouring, we hypothesize, of cytokines that would further produce pain in the back that we believe our product is interacting with.

Now, I'll speak more about our product in just a moment, but I would like to share with you this recent publication that came from San Diego, the group at San Diego Healthcare, as well as UCSD. Garfin (ph.) was involved with this. I probably

remember this panel. Cole Kim (ph.) was involved with this study.

And, basically, what they found is if they took rodents and they performed laminectomies that these animals often had pain and they had heightened sensitivity to pain as a result of those surgical procedures. And so the question was can they affect that, can they block that? So what they did was they performed a laminectomy at L5 and 6 and then they treated that laminectomy with hyaluronic acid. Now, this is not Oxiplex, but it is a polysaccharide, which does coat the tissues and remains in place for a period of time postoperatively.

And then they followed the animals postoperatively by measuring inflammatory mediators, IL 1, IL 6. They did that histochemically. And then they evaluated monocytes in the epidural space using, once again, an antibody for monocytes. And they found very clearly that HA reduced in the epidural space and around the nerve root. So now I'm bringing into consideration not just the nerve root but the epidural space. Utilization of a polysaccharide placed in that area reduced acutely the inflammatory mediators, reduced white-celled infiltration, and then, of course, both inflammatory mediators and

white-celled infiltrates were essentially normalized. In the case of the inflammatory mediators, not detectable seven days postoperatively.

So their conclusion was that pain reduction by polysaccharides, treatment after laminectomy and disc injury in a rat model resulted from reduction of the cytokines and inflammatory white-celled infiltrates that would otherwise occur around the nerve root and the epidural space.

Now, as I said, we have no human data that this occurs in patients. It's obviously a very difficult thing to test in humans, but we believe this is a plausible hypothesis.

So to go back to the original question, we believe there is interaction in these patients with severe back pain, that the progression on the nerve root is producing an inflammatory environment that is affecting the perception of pain not only in the sciatic nerve, but also in the lower back. And when that compression is removed in the control population, there is the usual reduction of sciatic pain in a large number. But some of them continue to have leg pain perhaps because of fibrosis, but also because of the memory that occurs in the sciatic nerve as a result of having been compressed for a

long time.

And in the lower back, the utilization of a mechanical barrier which coats and adheres to the tissues and separates them would provide at least some measure of protection against inflammatory mediators that occur as a result of surgery as well as outpouring from the annulotomy site itself.

So that's what we think is the situation. That's our plausible hypothesis that would begin to put these pieces together. And the information is in bits and pieces. I think this is the most direct information to address it.

Dr. McCormick asked a very interesting question relating to the whole question of the re-ops. And it's interesting, Dr. McCormick, when we began this study five years ago, we were not anticipating a reduction in re-operation rates. We were anticipating the usual re-operation rates that Dr. Rao talked about a little bit earlier today. But, in fact, that was not what we found.

We found, as we said earlier today, there was one patient that received Oxiplex that had a re-operation. This is within 3 months postoperatively, and there were six in the controls. And we began to look very carefully at that.

First, let's talk about the distribution of those surgical procedures. They occurred at six different clinical trial sites. It wasn't one site. Of the seven re-ops, they occurred at six different sites. So, therefore, one site had two, and that was not the Oxiplex patient. They Oxiplex patient was re-operated on as a single re-operation at a site. So that phenomenon was spread broadly across the population. And then you asked a very challenging question. Well, why might this be? What on earth could you think of, something that might reduce fibrosis, that might reduce the inflammatory process, why would it reduce re-operation rates?

So we began to look at this question, and we performed a bone healing study to see if there was any affect of Oxiplex on bone healing. Now, the bone that we used was the rat tibia. So this is not a human and this is not a vertebral body. This is the rat tibia. We used this model because it's an industry standard model that's used to evaluate active bone healing agents.

And we measured osteoid activity within the tibia, fairly standard assay, and you can see that at a 14-day time period, there was very clearly a difference. The Oxiplex-treated animals had over

twice the amount of osteoid activity -- that means active bone healing -- than did the controls.

So I think there's at least reason to speculate that Oxiplex may facilitate the repair process that occurs after these surgical procedures, and, if so, as I said to Dr. Goodman earlier today, it's not a pharmaceutical action. We're speculating that it provides a scaffold to support extracellular matrix and cellular ingrowth. And studies are ongoing to address that.

So, if this is the case in the area in the epidural space, you, in fact, would predict that there would be a relative reduction in re-operation rates if, in fact, there is an active healing component or a facilitation of healing that goes on following these types of surgical procedures.

I'd like to talk a bit about the choice of materials that Dr. Goodman asked. And his question was why did you choose polyethylene oxide and why did you choose carboxymethylcellulose. And, as you can imagine, when we embarked on this process now over 12 years ago, it was a process of trying to find something that made sense. And as Dr. Sang said, this was a very challenging clinical project because of the success rates and the difficulties in

measuring pain.

And now comes the difficulty of choosing the biological -- a biomaterial that you can place in a very actively healing site that has bleeding and is a dynamic site and you're right on top of a nerve root. And we screened a lot of materials. And we ended up with carboxymethylcellulose and polyethylene oxide for the following general reasons.

Carboxymethylcellulose, as we said earlier, does not produce an inflammatory response, but it does provide muco-adherence. One thing about this material, when you place it on the nerve root, you place it under the nerve, as Dr. Rhyne showed, in the epidural space. There is excellent adherence to the surgical site. It does not move, which we think is something that's very important for a lot of reasons, including safety.

However, the addition of polyethylene oxide we found also to be very important because polyethylene oxide prevents the protein deposition in the surgical area. And we're now talking broadly about proteins. The types of proteins that Dr. Rao referred to, all of those, if you placed them in the presence of polyethylene oxide, there are interactions. And so as we look into these areas

histologically after adding PEO and CMC, we simply don't see the proteins, we don't see the cellular infiltrates. That's why we went forward, of course, first, with our safety studies, and that's why we chose those two materials.

We added calcium chloride because it facilitated the increase in viscosity. And we chose a viscosity that made sense from a surgical application point of view so it's easy to use. It's not too runny. It's not too difficult through the catheter. So those are the -- that's why we chose those materials, and they've worked very, very well. As you know -- excuse me -- as we said, there are over 100,000 patients who have been treated with this material worldwide. We've gotten excellent reports. There have been publications, as we said earlier, that occurred completely outside of the company, and a lot of excitement in a number of areas.

The last thing I think I have on my list, although I -- oh, no -- the two things -- excuse me -- the next to last. Dr. Hanley asked how many -- what percent of the patients had more back pain than leg pain. And this is an interesting question, and we've tried to work with these numbers to facilitate a physiological understanding, and that

work is still going on. But the number is 87 patients had greater back pain than leg pain out of a total population of 351. Keep in mind that a lot of patients had a lot of pain in both groups, in both measurements. But in terms of the absolute difference, whether it was a difference in the case of back pain, it would be 87.

In looking at the response to those patients, it's, as I think Dr. McCormick was alluding to, in terms of response occurs, the higher the pain, the more reduction. And so, in terms of the Oxiplex or the surgical patients, if they had -- if they were in the group that had higher back pain levels, the back pain reduced more than the leg pain. But I think that may well be a phenomenology. The measurement -- I don't know that we can go beyond that. At least I can go beyond that in terms of talking about a physiological response. But those are the numbers.

And then Dr. McCormick asked about -- and I believe you said you heard us in the morning talking about 0 to 100-point score and you read something about 20 to 120-points, something like that. We tried to find that over our extensive lunch break.

DR. McCORMICK: Page 39.